DEVELOPMENT OF DISPUTE PREDICTION AND RESOLUTION METHOD SELECTION MODELS FOR CONSTRUCTION DISPUTES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MURAT AYHAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING

DECEMBER 2019

Approval of the thesis:

DEVELOPMENT OF DISPUTE PREDICTION AND RESOLUTION METHOD SELECTION MODELS FOR CONSTRUCTION DISPUTES

submitted by **MURAT AYHAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Ahmet Türer
Head of Department, **Civil Engineering** _______________

Prof. Dr. Mustafa Talat Birgönül
Supervisor, **Civil Engineering, METU** _______________

Prof. Dr. İrem Dikmen Toker
Co-Supervisor, **Civil Engineering, METU** _______________

Examining Committee Members:

Prof. Dr. Rıfat Sönmez
Civil Engineering, METU _______________

Prof. Dr. Mustafa Talat Birgönül
Civil Engineering, METU _______________

Prof. Dr. Ali Murat Tanyer
Department of Architecture, METU _______________

Assist. Prof. Dr. Gözde Bilgin
Civil Engineering, Başkent University _______________

Assist. Prof. Dr. Saman Aminbakhsh
Civil Engineering, Atılım University _______________

Date: 17.12.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Murat Ayhan

Signature:

ABSTRACT

DEVELOPMENT OF DISPUTE PREDICTION AND RESOLUTION METHOD SELECTION MODELS FOR CONSTRUCTION DISPUTES

Ayhan, Murat
Doctor of Philosophy, Civil Engineering
Supervisor: Prof. Dr. Mustafa Talat Birgönül
Co-Supervisor: Prof. Dr. İrem Dikmen Toker

December 2019, 376 pages

Construction industry is overwhelmed by increasing number and severity of disputes proving that current practices are insufficient in avoidance. This research argues that in order to forestall and mitigate construction disputes, prediction models should be developed by utilizing machine learning algorithms. The research suggests developing three distinct models; (1) dispute occurrence prediction model, (2) potential compensation prediction model, and (3) resolution method selection model. For this reason, an extensive literature review is conducted to identify input variables that impact dispute occurrence, compensation, and resolution method selection. Findings of the literature review is used to develop a conceptual model that involves attributes related to project, parties, dispute, and resolution method characteristics along with attributes related to changes, delays, and knowledge on resolution methods. Then, a questionnaire is designed based on the conceptual model to collect empirical data from decision-making authorities. Chi-Square tests of association is performed on collected datasets to reveal the significance of associations between inputs and outputs. Insignificant attributes are eliminated and finalized prediction models are developed. These models are tested by using alternative single and ensemble machine learning algorithms to obtain the best classifiers. 10-fold cross-validation results with ten repeats showed that dispute occurrence prediction model achieved 91.11% average prediction accuracy while potential compensation prediction model achieved 80.61% average accuracy and resolution method selection model has 89.44% average classification accuracy. These promising results show that proposed models can be beneficial for management personnel by supporting the decision-making process in future disputes based on data from past disputes.

Keywords: Dispute Prediction, Resolution Method Selection, Machine Learning, Data Classification

ÖZ

İNŞAAT PROJELERİNDE UYUŞMAZLIK TAHMİNİ VE ÇÖZÜM YÖNTEMİ SEÇİMİ MODELLERİNİN GELİŞTİRİLMESİ

Ayhan, Murat
Doktora, İnşaat Mühendisliği
Tez Danışmanı: Prof. Dr. Mustafa Talat Birgönül
Ortak Tez Danışmanı: Prof. Dr. İrem Dikmen Toker

Aralık 2019, 376 sayfa

İnşaat sektöründe uyuşmazlıkların sayısı ve şiddetindeki artış, sektörü ciddi sıkıntılara sokarken mevcut uygulamaların uyuşmazlıkların önüne geçmekte yetersiz kaldığını da ortaya koymaktadır. Bu araştırmada inşaat projelerindeki uyuşmazlıkların önüne geçilmesi için makine öğrenmesi tekniklerine dayanan tahmin modellerinin geliştirilmesi önerilmektedir. Buna göre üç farklı model önerilmiştir: (1) uyuşmazlık oluşumu tahmin modeli, (2) potansiyel tazminat tahmin modeli ve (3) uyuşmazlık çözüm yöntemi seçim modeli. Bu nedenle, uyuşmazlık oluşumunu, tazminatları ve çözüm yöntemlerini etkileyen değişkenlerin belirlenebilmesi için kapsamlı bir literatür taraması yapılmıştır. Literatür taramasından elde edilen bulgular kavramsal bir model geliştirilmesinde kullanılmıştır. Geliştirilen kavramsal model proje, taraflar, değişiklikler, gecikmeler, uyuşmazlıklar, çözüm yöntemleri ve çözüm yöntemi bilgi seviyeleri hakkında değişkenler içermektedir. Tahmin modellerinin geçmiş inşaat projelerinden alınan verilere dayalı olması için kavramsal modele dayanan bir anket hazırlanarak karar vericilere uygulanmıştır. Toplanan veri setleri üzerinde Ki-Kare testleri uygulanarak model girdileri ile çıktıları arasındaki ilişkiler test edilmiştir. Çıktılar üzerindeki etkisi istatistiksel olarak anlamlı olmayan değişkenler modelden elenerek tahmin modellerinin son haline ulaşılmıştır. Bu modeller, en iyi sınıflandırıcıyı tespit etmek için alternatif tekli ve grup (çoklu) makine öğrenmesi algoritmaları ile test edilmiştir. On kez tekrarlanan 10 katlı çapraz doğrulama sonuçlarına göre uyuşmazlık oluşumu tahmin modeli %91.11 ortalama tahmin doğruluğu yakalamıştır. Bu oran, potansiyel tazminat tahmin modeli için %80.61 ve uyuşmazlık çözüm yöntemi seçim modeli için %89.44 olmuştur. Bu sonuçların ışığında, araştırmada önerilen modellerin gelecekte çıkacak uyuşmazlıkların karar verme süreçlerinde karar vericileri destekleyebileceği görülmektedir.

Anahtar Kelimeler: Uyuşmazlık Tahmini, Uyuşmazlık Çözüm Yöntemi Seçimi, Makine Öğrenmesi, Veri Sınıflandırması

To my beloved wife Cansu CİNDORUK AYHAN…

ACKNOWLEDGEMENTS

This research would not have been possible without support and encouragement of many special people. Although it is not possible to thank every single person, I would like to use this opportunity to show my gratitude to those who shared this journey with me.

I would like to thank my advisor Prof. Dr. Mustafa Talat Birgönül and co-advisor Prof. Dr. İrem Dikmen Toker for their insightful comments, invaluable guidance, endless support and patience. It has been a privilege to have the opportunity to conduct this research in METU under their supervision.

It is difficult to express my gratitude to my wife, Cansu Cindoruk Ayhan. It was her unconditional love, patience and support that made this research possible. It is a privilege to be in her life.

I owe my deepest gratitude to my family. To my mother, Hediye Ayhan, for her endless love and blessings. To my father, Prof. Dr. Rıza Ayhan, who is the reason of my academic studies. To my sister, Hatice Ebru Töremiş, who is my idol and beacon. To my brothers, Zeynel Burak Ayhan, Erkutalp Ayhan and Töre Töremiş, who were there for me whenever needed. To my sister, Elina Ayhan, for her kind and caring support. To Toprak Töremiş, for his endless energy. I would also like to thank my wife's family, Prof. Dr. Mehmet Cindoruk, Oya Cindoruk and Can Cindoruk. A more supportive and considerate family I could not have asked for. I am also thankful for two lovely ladies, Maya and Limon, whose presence is the primary reason to enjoy life.

Finally yet importantly, I would like to thank my dearest friends, Dr. Oya Memlük Çobanoğlu, Baran Çobanoğlu, Pınar Berberoğlu, Cihan Şimşek, Miran Dzabic, Onur Can Sert, Duygu Akbıyık, Çağrı Akkaya, Dr. Saman Aminbakhsh, Handan Gündoğan and Özgür Dedekargınoğlu for their invaluable friendship.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AAA	American Arbitration Association
ADR	Alternative Dispute Resolution
AHP	Analytical Hierarchy Process
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
ASCE	American Society of Civil Engineers
AUROC	Area Under ROC Curve
BDT	Boosted Decision Trees
CART	Classification and Regression Trees
CBR	Case-Based Reasoning
CHAID	Exhaustive Chi-Squared Automatic Interaction Detection
DA	Discriminant Analysis
DAB	Dispute Adjudication Board
DL	Decision List Algorithm
DPI	Dispute Potential Index
DRB	Dispute Review Boards
ECC	Exhaustive Correction Code
FIDIC	Federation of Consulting Engineers

FmGA	Fast and Messy Genetic Algorithm
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
HKIAC	Hong Kong International Arbitration Center
ICSID	International Center for Settlement of Investment Disputes
IPM	Integrated Prediction Model
J48	C4.5 Decision Trees in WEKA Workbench
KNN	K-Nearest Neighbor
LR	Logistic Regression
MAP	Maximum a Posteriori
MAUT	Multi-Attribute Utility Technique
MCDM	Multi-Criteria Decision-Making Method
MDA	Multi Discriminant Analysis
ML	Machine Learning
MLP	Multilayer Perceptron
NADRAC	National Alternative Dispute Resolution Advisory Council
OGC	Office of Government Commerce in London
OVA	One-Vs-All
OVO	One-Vs-One
PART	Projective Adaptive Resonance Theory
PMBoK	Project Management Body of Knowledge

PMI	Project Management Institute
PPP	Public-Private Partnership
PSO	Particle Swarm Optimization
QUEST	Quick-Unbiased-Efficient Statistical Trees
RBF	Radial Basis Function
RCC	Random Correction Code
ROC	Receiver Operating Characteristic
SEA	Senior Executive Appraisal
SEM	Structural Equation Modeling
SVM	Support Vector Machines
TAN	Tree Augmented Naïve Bayes
TN	True Negative
TP	True Positive
TPCC	Taiwan Public Construction Commission
TPPA	Turkish Public Procurement Authority
UPM	Universal Prediction Model
WEKA	Waikato Environment for Knowledge Analysis

LIST OF SYMBOLS

SYMBOLS	
AUROC	Area under ROC curve
b	Bias term
C	Penalty parameter in SVM
d	Polynomial kernel function degree
E	Observed event E
FN Rate	False negative rate
FP Rate	False positive rate
Gain (X, A)	Information gain
h	Hypothesis h
H	Hyperplane
Info (P_i)	Entropy of P
K (x_i, x_j)	Kernel function
L_D	Dual Lagrangian problem
L_P	Primal Lagrangian problem
P_i	Probability of an instance belonging to a certain class i
P (A)	Observed agreement between actual and predicted values
P (E)	Expected agreement (probability of agreement by chance)
TN Rate	True negative rate
TP Rate	True positive rate

w	Weight
α_i	Lagrangian multiplier i
ξ_i	Slack variable
μ_i	Lagrangian multiplier enforcing the positivity of ξ_i
Φ	Function for non-linear feature space mapping
σ	Spread parameter in Gaussian RBF kernel functions
γ	Gamma spread parameter in Gaussian RBF kernel functions

CHAPTER 1

INTRODUCTION

Construction projects are carried out by several teams and contractors of various expertise in a continuously changing and complex environment plagued with high levels of risks and uncertainties. Resulting in unique products, construction projects also need to satisfy a wide variety of stakeholder requirements while dealing with geography, site conditions, communities, challenging physical environments, and existing infrastructure at the same time. Moreover, they generally require usage of high-end technology and sophisticated equipment besides integration of diverse engineering disciplines such as civil, structural, electrical, mechanical, geotechnical, etc. (Project Management Institute (PMI), 2016). Thus, multiple discrete parties having different specialties come together temporarily to work in a coordinated manner with the main goal of completing the project successfully within the planned schedule and budget (Harmon, 2003). However, these parties also have varying goals and expectations as well as they seek to maximize their own benefits simultaneously, which cause perception differences and conflicting goals (Cheung and Suen, 2002). The primary goal of the client is usually to get the maximum quality and functionality at the minimum cost, while contractors expect to achieve financial goals and client satisfaction (Howard et al., 1997). In addition to these differences, stemming from their large and complex nature, construction projects involve large number of uncertainty sources (i.e. ground conditions) that makes encountering conflicting situations in the construction industry more common than many other domain (Dalton and Shehadeh, 2003). Because of these conflicts, the relationship between parties of a construction project is inherently adversarial, which derails projects from the main goal of successful completion (Jones, 2006; Harmon, 2003; Fenn et al., 1997). Conflicts do not only cause adversarial relationships, but also disrupts the success of construction projects by potentially creating additional costs for all parties (Thompson et al., 2000). When a conflict is not satisfactorily settled, it can quickly escalate to become a claim and ultimately a dispute (Cheung and Suen, 2002).

In the literature, construction disputes are characterized as unwanted, unpleasant, time consuming, and expensive (Harmon, 2003). It is emphasized that construction disputes cause waste of scarce resources and damage participant relationships (i.e. client-supplier relationship), which are built up in many years with difficulty (Fenn, 2007). Disputes may also result in significant delays and additional costs that contradict the goals of successful project completion, which are completing the project on time, within the budget and with the desired quality (Office of Government Commerce (OGC), 2002; Chen and Hsu, 2007). In other words, with a potential to result in delayed schedules, budget overruns, poor quality and performance, increased tension, and damaged long-term business relationships, construction disputes can be detrimental (Cheung and Suen, 2002).

The severity of disputes in construction has been well understood and documented (Gebken and Gibson, 2006). However, the construction industry still struggles to find methods to resolve them justly and economically. Parallel to this, the industry has acquired a bad reputation for being contentious and is overwhelmed by increasing number of dispute occurrences (Arditi et al., 1998; Cheng et al., 2009). Moreover, the increasing complexity of construction projects causes an additional increment in the complexity and number of disputes (Cheung and Suen, 2002; Harmon, 2003; Acharya et al., 2006). Corroborative data can be found in the literature. In a study analyzing the application of dispute review boards (DRB) in construction projects covering the period 1975-2007 in the United States (U.S.), it is revealed that the number of projects with DRB panel per year is increasing rapidly (Menassa and Peña Mora, 2009). Moreover, among the 60 billion U.S. Dollars spent every year on lawsuits in the U.S., nearly 5 billion is related solely to construction industry and litigation expenses are steadily increasing (Pulket and Arditi, 2009a). A more recent study is taken from the American Arbitration Association (AAA) that shows the number of construction cases submitted to the AAA in 2017 were up by 4%, which involves 12% increase in the number of cases with claim amounts of at least 0.5 million U.S. Dollars and 13% increase with amounts of at least 1 million U.S. Dollars (AAA, 2018). Examples from other countries are also similar. In the period of 2002-2009, the dispute rate in public-private partnership (PPP) projects undertaken by the Taiwan Public Construction Commission (TPCC) was 23.6% (Chou et al., 2013a). According to annual reports of Hong Kong International Arbitration Centre (HKIAC), of the 429 cases handled by HKIAC in 2009, there were 93 cases of construction disputes covering 21.6% of all cases (HKIAC, 2009). More recently, between 2015 and 2017, the average rate of construction disputes was as high as 20.2% among all HKIAC registered cases (HKIAC, 2018). In Turkey, there is approximately 12% increase annually in the number of litigious cases related to construction, expropriation, demolition and related works (Yılmaz and Dikbaş, 2013). Considering all these examples, it can be concluded that there is an increase in the number of construction disputes worldwide with disputed projects covering a significant portion in the industry.

The financial consequences of construction disputes should also be highlighted. The estimated additional direct costs of disputes range from 0.5% to 5% of the contract value (Love et al., 2010). On the other hand, there are also indirect costs due to decreased productivity, strained business relationships, loss of future business opportunities, damaged reputation of parties, etc. that amplify damages caused by disputes (İlter, 2012).

Supported by the statistical data, successful construction projects are not achieved as often as participants would like (Harmon, 2003). Disputes being one of the main reasons behind this, it is no surprise that the research on construction dispute drew significant attention. Any attempt to forestall possible disputes may have significant contributions to the industry in terms of time and cost savings, especially in the public projects (Yılmaz and Dikbaş, 2013). It is commonly accepted in the construction industry that the best solution against disputes is to avoid them (Fenn, 2007). Therefore, it is expected that attempts and efforts should be directed to avoid disputes before occurrence. However, there is a prevalent tendency in the current literature to perceive disputes as inevitable. (Cheung and Suen, 2002; Cheung et al., 2002; Yates, 2003; Gebken and Gibson, 2006; Kassab et al., 2006; Fenn, 2007; Ellis and Baiden, 2008; İlter and Dikbaş, 2009). Consequently, although there are limited research on identification of dispute likelihood of projects (Diekmann and Girard, 1995; Molenaar et al., 2000) and dispute prediction (Dalton and Shehadeh, 2003; Chou and Lin, 2012; Chou et al., 2013a; Chou et al., 2014), studies mainly focus on mitigation of impacts of disputes and determination of appropriate management and resolution methods. In addition, there are many researchers attempting to reduce the number of claims and disputes before the dispute reaches its final resort, the courthouse. Subjects of such studies involve, but not limited to, identification and classification of reasons, causes, sources, and types of claims and disputes (Revay, 1993; Watts and Scrivener, 1993; Kumaraswamy, 1997; Cheung and Pang, 2013); justification, quantification, and impact/outcome analysis of claims, disputes and resolution methods (Alshawi and Hope, 1989; Alkass et al., 1993; Arditi and Pattanakitchamroon, 2006; Gebken and Gibson, 2006; Chau, 2007; Chen and Hsu, 2007; Cheng et al., 2009; Arditi and Pulket, 2009; Pulket and Arditi, 2009b); investigation and selection of appropriate dispute management and resolution methods (Cheung, 1999; Cheung et al., 2002; Jones, 2006; Kassab et al., 2006; Kassab et al., 2010; İlter, 2010a; Marzouk et al., 2011; Chou, 2012; Chou et al., 2013b; Yılmaz and Dikbaş, 2013).

The above-mentioned studies provide valuable theoretical frameworks and practical models. However, as the number and impact of disputes increasing continuously, it can be said that these efforts are not sufficient to satisfy the needs of the construction industry. Therefore, new decision-support technologies are needed (Kassab et al., 2006). Considering the fact that developing deterministic mathematical models to solve construction management problems is difficult and costly, the research interest moves towards approximate inference as a fast and cost efficient alternative (Cheng and Wu, 2009). Consequently, the use of artificial intelligence (AI) is accepted as a strongly effective method for determining numerous complex and interconnected factors related to construction disputes along with hidden relationships that are difficult to rationalize (Chou and Lin, 2012). Establishment of dispute libraries using AI applications can have significant contributions to the construction industry (İlter and Dikbaş, 2009). Among various AI applications, data mining via machine learning (ML) techniques form an important research branch since 1960's as they enable gathering valuable information from large volumes of data that is difficult to understand and interpret (Liao et al., 2012). In addition to these, as a data mining subdomain, supervised learning in ML (i.e. classification and regression problems) is a potential tool in management domain (Chou et al., 2014). Such tools can be used for forecasting dispute occurrence and supporting resolution process.

Prediction of potential disputes using ML techniques will be valuable for management personnel as early planning to take necessary precautions will be possible, which may reduce the effort, time, and cost of dispute management actions considerably (Marzouk et al., 2011; Chou et al., 2014).

Dispute management is one of the main determinants of the performance of construction projects and consequently, the industry (İlter and Dikbaş, 2009). However, achieving successful dispute management is a complex and challenging problem (Chou et al., 2013b). This is because the best way to deal with construction disputes does not exist since disputes often vary in scale, complexity, and nature due to the fragmented and complex status of construction projects. Adding complexity of technical and financial matters related to disputes, various external (i.e. political) and internal factors (i.e. personnel related attributes) as well as their interrelations should be carefully considered in dispute management decision-making (Cheung and Suen, 2002; Harmon, 2003). The current tendency in the construction industry is to make these challenging decisions intuitively based on the experience of the decision-maker with limited available information of questionable quality (Chou et al., 2013b). Moreover, a research on Turkish construction industry revealed that majority of professionals characterizes their decision-making rationale for resolution processes and resolution method selection as unconscious, while pointing out the need for tools to make informed and logical decisions at the same time (İlter, 2010b). In other words, instead of depending on a systematical process, the current decision-making practice in dispute management is prone to subjectivity. AI applications, on the other hand, has the potential to minimize this subjectivity (Cheung et al., 2004a). Utilization of AI techniques enables to systematically select the resolution strategy (Cheung and Suen, 2002). Although dispute resolution methodology is usually chosen before occurrence of a dispute via contract documents, considering the nature of the dispute and factors such as the relationship between disputed parties, the methodology that best suits the needs of the participants should be selected (Harmon, 2003). While performing such a selection, experience and knowledge are invaluable. The merits of AI techniques include extraction of such tacit knowledge in an articulable and presentable way to the relevant personnel. As a result, an informed decision-making can be achieved during resolution method selection via appropriate decision-support systems based on AI techniques (Cheung et al., 2004a). Development of AI based models can also enable early warning of potential dispute resolutions (Chou et al., 2014).

In the light of these, the AI, specifically ML, based applications have drawn attention in the literature. As they yielded promising results, these methods are being used soundly in establishing decision-support systems (Pulket and Arditi, 2009a). However, the literature has proven that it is not possible to solve all data mining problems using a single ML technique because of the varying characteristics of real world datasets. Instead, in order to obtain accurate results, the bias due to learning technique should be compatible with the dynamics of the problem domain, which makes data mining an experimental process (Witten et al., 2016). In other words, the effectiveness of ML techniques depend on characteristics of the application domain and dataset, along with various enhancements that increase their performance. An accurate model that is proven to be well performing on a certain dataset does not necessarily have to perform well on another (Pulket and Arditi, 2009b). Therefore, the ML technique that performs best on construction dispute problems with the available data should be experimentally determined.

With the aim of demonstrating that models to predict dispute occurrence and to support decision-making for resolution processes can be generated, this thesis study compares the strengths and weaknesses of similar studies that utilize ML techniques. This detailed literature review will be given in Chapter 2 of the thesis study.

In short, people interested in construction dispute domain should aim to reduce the number of potential disputes to a minimum as well as to manage them in the most effective way upon inevitable occurrence. In doing this, they can benefit from AI applications, specifically ML techniques. Thus, this thesis study will focus on the ML techniques in the dispute management domain in order to perform dispute prediction and to support the decision-making during resolution processes.

1.1. PROBLEM STATEMENT

Producing numerous valuable theoretical and practical outcomes, the topics in construction dispute domain were discussed intensively. However, the number of disputes are still increasing worldwide along with disruptive effects on the construction industry. This is a clear proof that shows current avoidance, management, and resolution efforts do not meet industry's requirements. The high frequency of occurrence in the industry and the consensus on the inevitable nature should not necessarily imply that disputes cannot be avoided (Revay, 1993). Contrarily, the industry requires development and employment of adequate decision-support technologies in order to forestall and mitigate disputes.

It is commonly accepted in the construction industry that the best solution against disputes is to avoid them. Necessary actions to avoid disputes can only be achieved by prediction (Fenn, 2007). Therefore, in order to decrease the number of construction disputes, dispute prediction is an important research area. AI based applications, especially new technologies and algorithms available in the ML domain, makes developing data-specific prediction models possible. When the output variable is a categorical variable, prediction problems become data classification problems (Chou and Lin, 2012). For example, in the case of dispute occurrence prediction, the output variable is dispute occurrence, where ‘undisputed cases’ can be categorized as ‘0’ and ‘disputed cases’ as ‘1’. ML algorithms are well equipped to solve such data classification problems. A similar situation is also present in resolution method selection, where the resolution method as the output variable can again be categorical. For example, litigation can be categorized as one group, arbitration as another, mediation as another, and so on. Thus, it can be said that ML techniques are also suitable for decision-support systems in resolution method selection. Therefore, among various AI applications, this research will focus on establishment of prediction models via data classification by employing adequate ML techniques on construction dispute data.

Data classification problems model relationships between various input variables to an output, but it is not an easy task to determine which input variables affect the output variable. In pursuit of achieving this, the literature on dispute management focuses on diverse areas such as identification of dispute causes, factors affecting disputes and their resolution, analysis of resolution methods and alternative dispute resolution (ADR) techniques, and decision-support models and systems in dispute resolution, etc. However, there is no consensus in the literature. Instead of sharing a common wisdom, studies are conducted based on different causes and factors with varying levels of inclusion and detailing, which cause understanding and perspective differences. In addition, there is a confusion in the related terminology due to overlapping concepts and the distinction between causes, factors, types of disputes, etc. may not be very clear (İlter, 2012). Moreover, as the research progresses, scholars detect importance of new variables that should update the upcoming research efforts. In order to tackle these problems, a common ground should be established that involves findings of previous studies as distinct input variables.

Another problem of dispute management models and systems is the level of representation. Although global-scaled research can be found to some extent, studies are mainly based on local industries. Thus, they are not capable of representing the construction industry as a whole. Moreover, instead of a general approach, it is observed that the main preference is to conduct the research for public or PPP projects only. Despite the fact that there are various parties from many domains in a construction project, most of the studies also fail to represent or target these various professions since they merely review certain groups. The results of a study analyzing the literature on dispute resolution is parallel. Another important highlight of that study was the tendency of the current literature to produce general insights and statistical outcomes rather than establishing supporting models or systems (İlter and Dikbaş, 2008).

Considering all these problems, this thesis study will focus on ML techniques in the dispute management domain in order to perform dispute prediction and to support the decision-making process during dispute resolution. However, prediction problems are complex to model since these problems contain substantial uncertainty and vagueness besides the questionable quality of the available data (Chou et al., 2014). In addition, it is difficult to select the best single or combined ML algorithms that suit the prediction problem at hand. The conventional approach in ML domain is to experimentally compare the classification performances of promising single ML algorithms with each other as base classifiers and select the best performing one in that dataset. In addition to this, enhancements to these base classifiers can be pursued by creating ensemble classification schemes systematically, which are basically adding or combining base classifiers to improve classification (prediction) performance (Arditi and Pulket, 2009). Therefore, such problems require an experimental data mining approach.

In the light of all these, it is claimed that in dispute prediction and resolution method selection, there is a need for sound decision-support technologies that are capable of representing the industry inclusively with a general approach that can be benefited by various project participants. The research at the core of this thesis study addresses these needs systematically by collecting and processing data from various construction projects and classifying them by models derived from utilizing ML algorithms. Performing numerous experiments on the collected data with several alternative ML algorithms, the aim is to present models with best prediction (classification) performances in order to fill the mentioned gaps in the construction dispute literature.

1.2. SCOPE, OBJECTIVES AND LIMITATIONS

Based on defined problems, the primary objective of this research is to utilize state-of-the-art ML techniques on real-world construction project data to extract the invaluable tacit knowledge of previous dispute cases and to present the results to relevant personnel beforehand. By doing this, for new cases, it would be possible to achieve dispute prediction (occurrence and potential compensation prediction) and appropriate resolution method selection that will function as an early-warning mechanism. Such an achievement may be beneficial to management personnel in avoiding or mitigating possible disputes by highlighting the complex interrelations between disputes and projects as well as helping them to take necessary precautions.

In short, this research argues that in order to forestall and mitigate construction disputes, prediction models should be developed via utilizing alternative ML techniques on datasets capable of representing variations in the construction industry. For this purpose, the research suggests developing three different prediction models: (1) a model for predicting dispute occurrence, (2) a model to help the decision-maker in understanding what compensation can be acquired out of the disputed case, and (3) a model for supporting the decision-making process during resolution method selection.

All prediction models will link several input variables related to a construction project to an output. In the case of dispute occurrence prediction, input variables will have impact on the dispute occurrence as an output such that projects in the dataset will be classified as “disputed projects” and “undisputed projects”. In the case of potential compensation prediction, input variables will have impact on the type of compensation that can be acquired out of the disputed case as an output such that “no compensation”, “cost compensation only”, “time compensation only”, and “time and cost compensation”. In the case of resolution method selection, input variables will have impact on the method to be chosen as an output variable such that "litigation" should be preferred, "arbitration" should be preferred, etc.

The lack of a common ground in disputes domain should also be tackled during development of prediction models. As stated earlier, instead of sharing a common wisdom, the literature is composed of studies that are conducted based on different causes and factors with varying levels of inclusion and detailing, which cause understanding and perspective differences. The research will address this problem by establishing a conceptual model firstly. The conceptual model will be developed as a result of an extensive literature survey with an aim to identify causes of disputes, factors affecting dispute development and potential compensations, mechanisms of resolution strategies and alternative resolution methods. As a result of the literature survey, numerous variables will be identified that can be used in prediction models.

Following the literature survey, empirical data is collected via questionnaires in order to establish a construction project dataset. All prediction models are based on data from construction projects collected specifically for this research. The questionnaire is designed according to the developed conceptual model. In other words, empirical data on variables identified in the conceptual model are collected via questionnaires. The dispute occurrence prediction model utilizes a dataset composed of 108 construction projects (38 undisputed projects and 70 disputed projects), while the potential compensation model utilizes 82 cases (12 cases with no compensation, 38 cases with only cost compensation, 5 cases with only time compensation, and 27 cases with both cost and time compensation) out of the 108 cases collected. Notice that compensation model utilizes 82 cases, which is more than the number of disputed projects (70 disputed projects) in the dataset. However, some projects experienced more than one disputed issue. This is the reason why there are more disputed cases than disputed projects in the dataset. These 82 cases are the cases in which participants declared satisfaction with the compensation. Finally, the resolution method model utilizes 54 cases coming from 82 disputed cases. These 54 disputed cases are the ones that are resolved satisfactorily according to participants.

Subsequent to collecting empirical data for development of prediction models, the next objective is to determine which input variables affect the output variable for each prediction model. For this purpose, the data is analyzed in terms of significance. In other words, the significance of associations between input variables and the output variable are analyzed for all input variables in all prediction models. The insignificant input variables are removed from the original conceptual model and three different prediction models are established with fewer input variables. This kind of variable elimination is known as attribute or feature selection/elimination in practice and it helps achieving better algorithm generalization in ML applications (Drucker et al., 1999). Another reason for attribute elimination by significance analysis is that the performance of ML algorithms is generally affected negatively by the irrelevant or insignificant attributes (variables) (Pulket and Arditi, 2009b). Therefore, elimination of insignificant attributes and selection of the ones impacting the model outcomes improve generalization performance of ML algorithms (Arditi and Pulket, 2009; Sönmez and Sözgen, 2017).

At the end of attribute elimination, experiments are conducted using alternative ML algorithms to determine the ML algorithm that gives the highest prediction performance. Each model is tested using its own dataset with the same set of ML algorithms and the best performing classifier is determined as the final prediction model.

To summarize objectives of this research so far; the primary goals are establishing a dispute occurrence prediction model, a potential compensation prediction model, and a resolution method selection model. However, in order to achieve these primary goals, a conceptual model involving findings of previous studies in the literature as distinct input variables should be established initially as a secondary objective. Then, empirical data is collected according to input variables identified in the conceptual model and insignificant variables are eliminated to establish the prediction models. Various alternative ML algorithms are tested to come out with the best prediction performance for all three models. However, the scope of this research does not include determining the best method (i.e. not all ensemble model combinations are experimented) but instead; the aim is to demonstrate that models using alternative ML techniques can be generated for purposes of dispute prediction and decision-support during dispute resolution.

The data specific nature of this research is regarded as its main limitation. In other words, the research is based on data from a finite number of construction projects. Although collected datasets can said to be quite representative, the number and variety of projects are still limited due to limitations on access to such information, research duration, and budget. The number and variety of projects can be increased so that the level of representation of the construction industry and generalization capabilities of presented models will be improved. In addition, further research can be performed to establish a combined decision-support system utilizing presented models via a user interface.

1.3. RESEARCH METHODOLOGY

In accordance with research objectives that aim to resolve defined problems in the literature about construction dispute management, the research methodology can be summarized in three steps (Figure 1.1).

Step 1 – Conceptual Model Development: There is no consensus in the literature about causes of disputes or factors that affect them. A similar case is also valid about factors that affect the dispute resolution decision-making. Instead of sharing a common wisdom, studies are conducted based on different causes and factors with varying levels of inclusion and detailing, which cause understanding and perspective differences. In addition, there is a confusion in the related terminology due to overlapping concepts and the distinction between causes, factors, types of disputes, etc. may not be very clear (İlter, 2012). In order to overcome these problems with a systematical approach, the first step involves an extensive analysis of literature on construction conflicts, claims, disputes, compensations, and resolution methods with the aim of synthesizing findings of the previous research. Findings of the literature survey will be used to develop a conceptual model (Chapter 2). The conceptual model will be composed of input and output variables identified in accordance with the literature survey. In other words, the conceptual model is developed to identify input and output variables for prediction models. Although there are numerous input variables, only three output variables are identified. These output variables are (1) dispute occurrence, (2) potential compensation, and (3) resolution method.

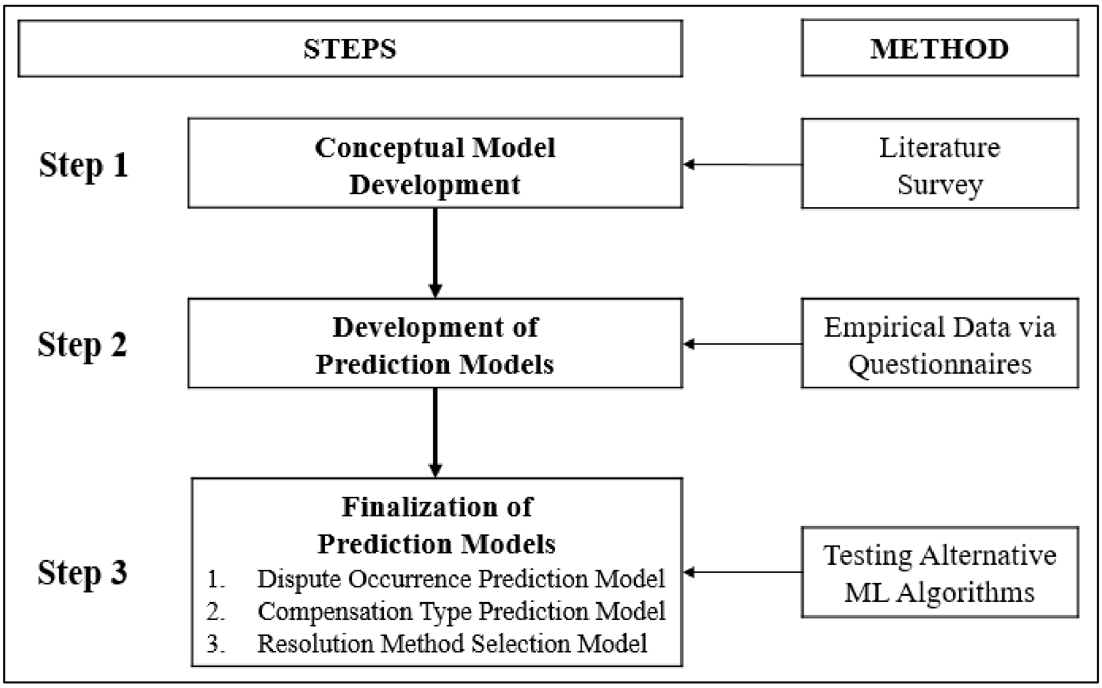


Figure 1.1. Research Methodology

<u>Step 2 – Development of Prediction Models:</u> The second step is the development of prediction models as (1) dispute occurrence prediction model, (2) potential compensation prediction model, and (3) resolution method selection model. For this purpose, empirical data on construction projects are needed. In order to collect construction project data, a questionnaire is designed (Chapter 3) based on identified variables in the conceptual model. With the goal of reflecting variations in construction types, contract documents, participants, delivery systems, business environments, etc., the data is collected from a wide variety of construction projects. The collected dataset involves data related to all variables identified in the conceptual model. However, the impact of input variables on outputs are not the same. Some variables may impact the outcome more than the others, while the impact of some variables can even be statistically insignificant. Therefore, the significance of association between input and output variables should be analyzed. Firstly, the collected data is processed and variables are turned into categorical variables (i.e. nominal and ordinal). Then, in order to understand whether there is a statistically significant relationship between input and output variables, Chi-Square statistics is utilized. Chi-Square statistics is a useful way of testing the existence of association relationship between categorical variables (Weisburd and Britt, 2007). The Chi-Square tests are performed in IBM SPSS Statistics version 22.0. Finally, according to contingency tables resulting from the Chi-Square tests, statistically insignificant variables are eliminated and prediction models are developed with significant variables only. In short, variable (attribute) elimination is performed on variables of the conceptual model via the Chi-Square tests to establish three prediction models.

Step 3 – Finalization of Prediction Models: The third step is finalizing prediction models via data classification. When the output variable is a categorical variable, prediction problems become data classification problems (Chou and Lin, 2012). This is the case in dispute occurrence prediction, potential compensation prediction, and resolution method selection. Thus, using established prediction models, the classification performance of alternative ML algorithms are tested. In these tests, various single and ensemble ML algorithms are experimented. The utilized single algorithms, which are also known as base classifiers, are (1) Naïve Bayes; (2) k-nearest neighbor (kNN); (3) J48, which is an algorithm that generates C4.5 decision trees; (4) multilayer perceptron (MLP), which is an enhanced artificial neural network (ANN) algorithm; and (5) support vector machines (SVM). Meanwhile, the ensemble models are developed by using (1) voting technique, (2) stacked generalization, and (3) the AdaBoost algorithm. Containing a collection of numerous inbuilt ML algorithms, Waikato Environment for Knowledge Analysis (WEKA) version 3.8.3, which is an open-source Java application produced by the University of Waikato in New Zealand (Frank et al., 2016), is used in data classification tests. The algorithm that produced the best classification performance is presented as the final prediction model.

In the light of these, Figure 1.2 summarizes the detailed research overview.

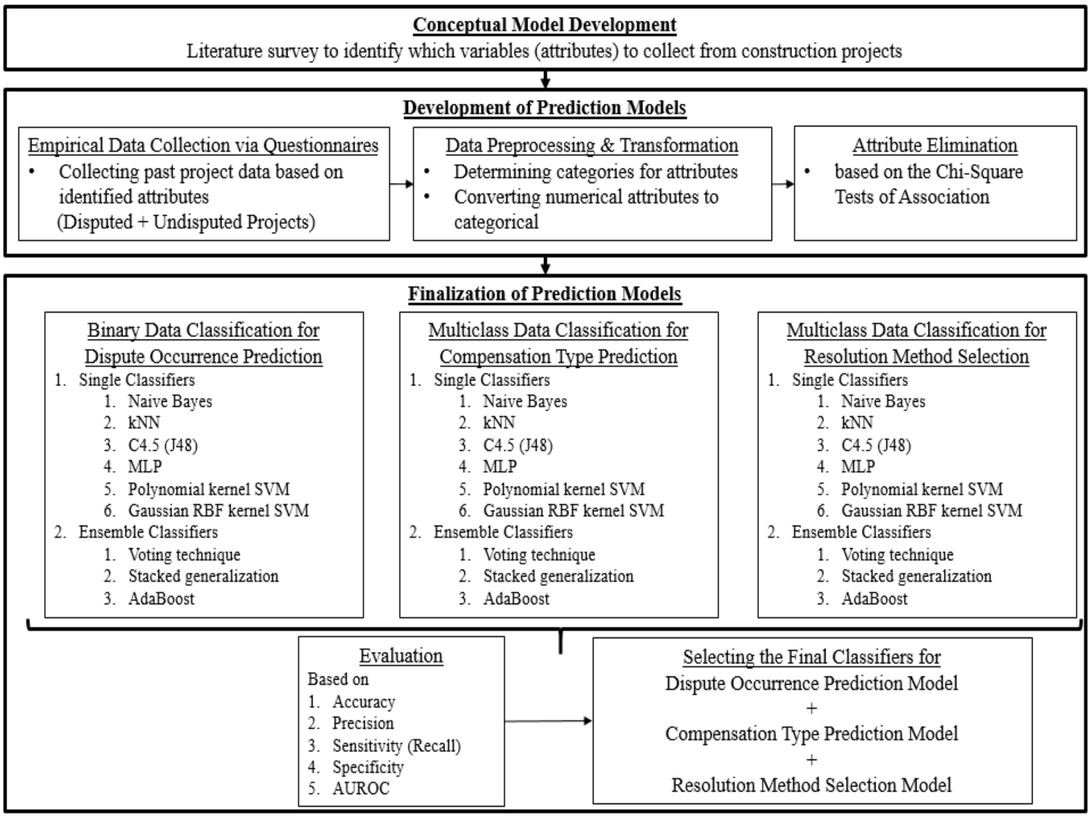


Figure 1.2. Detailed Research Overview

1.4. STRUCTURE OF THE RESEARCH

The order of contents of this research follows the steps in the detailed research overview (Figure 1.2). Chapter 2 starts with an introduction to research background on construction disputes. The basic terminology is explained from the research's point of view. Previous studies on construction disputes are highlighted and it is followed by a literature review on dispute prediction. Then, the research moves onto research background on dispute resolution methods and ADR techniques starting with an effort to generate an understanding of these techniques via comprehensive explanations of conventional resolution methods, which are litigation and arbitration, and ADR methods, which involve DRB, mediation, senior executive appraisal (SEA), and negotiation. This is followed by an investigation of previous studies on this domain and a literature review on decision-support systems for resolution method selection. Finally, resulting from this extensive analysis of literature, a conceptual model will be developed that will be used during data collection.

In Chapter 3, firstly, a questionnaire to collect empirical data is designed according to the conceptual model developed in Chapter 2. Secondly, empirical data is collected, processed, and initial findings are presented. Thirdly, attribute elimination is performed by using Chi-Square tests to identify the association relationship between input and output variables. Insignificant attributes, which may negatively affect the performance of ML algorithms, are eliminated from the conceptual model. Finally, utilizing the remaining significant attributes, established models for dispute occurrence and potential compensation prediction along with resolution method selection are given.

Chapter 4 starts with a detailed explanation of concepts in the ML domain. Differences in binary data classification and multiclass classification are emphasized. Once again, the importance of attribute selection before utilizing ML techniques is highlighted. Then, properties, characteristics, advantages, and disadvantages of the utilized single and ensemble ML techniques for data classification are reviewed. In addition, efforts for enhancing the performance of utilized techniques are explained.

In Chapter 5, results of binary data classification using ML techniques are given based on evaluations considering several performance metrics. According to these results, the final classifier for dispute occurrence prediction is selected. Similarly, results of multiclass data classification are presented and final classifiers for potential compensation prediction and resolution method selection are determined.

Chapter 6 includes concluding remarks, findings, contributions, and limitations along with recommendations for future research.

CHAPTER 2

RESEARCH BACKGROUND

In this chapter, research background on construction disputes will be given with explanation of the basic terminology and investigation of previous studies. It will be followed by a literature review on dispute prediction with a special emphasis on AI based applications and ML techniques. Then, research background on dispute resolution methods and ADR techniques will be given starting with an effort to generate an understanding of these techniques via comprehensive explanations of conventional resolution methods, which are litigation and arbitration, and ADR methods, which involve DRB, mediation, SEA, and negotiation. This is followed by an investigation of previous studies on this domain and a literature review on decision-support systems for resolution method selection with special emphasis on AI based applications. Finally, resulting from this extensive analysis of literature on construction conflicts, claims, disputes, and resolution methods, a conceptual model, which will be used for development of prediction models, is established.

2.1. RESEARCH BACKGROUND ON CONSTRUCTION DISPUTES

Before moving into details, the basic terminology related to construction disputes should be explained. Therefore, at this point, the perception of the terms “conflict”, “claim”, and “dispute” in this research should be identified, as there are variations in definitions of these terms in the literature and there is a tendency to use them in pairs or interchangeably without clear explanation of their meanings (Yates, 2003; Acharya et al., 2006; Younis et al., 2008; Ellis and Baiden, 2008).

2.1.1. The Basic Terminology – Conflicts, Claims, and Disputes

Project participants naturally have different expectations from a construction project and consequently, they have varying goals. For example, the client’s goal is usually to get the maximum quality and functionality at the minimum cost, while the contractor's goal is to fulfill its own financial expectations (Howard et al., 1997). Thus, goals of project parties are at conflict. In fact, conflict is a general behavioral concept that is common in every part of our lives. In general view, conflicts are serious disagreements or differences between two or more beliefs, ideas or interests (Kumaraswamy, 1997). Similarly, in construction domain, conflict refers to the clash of interests, values, or actions and if a party feels its position is threatened, conflicts arise (Love et al., 2011). Conflicts cause adversarial relationships among participants and may disrupt the success of a project (Thompson et al., 2000). When a conflict is not satisfactorily settled, it escalates further to become a claim and ultimately a dispute (Cheung and Suen, 2002).

In Project Management Body of Knowledge (PMBoK), a construction claim is defined as; "a demand for something due or believed to be due, usually as a result of a change in basis in the project execution; a variation or deviation in risk allocation; an action, direction, or requested change order against the agreed-upon terms and conditions of a contract or a part of the construction, which has failed or is not performing properly and cannot be economically resolved between the parties" (PMI, 2016). Resulting from contractual issues, claims on construction projects usually occur as an assertion for additional payment or extension of time (Kumaraswamy, 1997).

A construction dispute, on the other hand, is a rejected claim (Marzouk et al., 2011); a form of conflict that is made public and requires resolution (Cheung and Suen, 2002); an unreached agreement on a change occurrence or compensation claim that negatively affects the project and the participant relationships (PMI, 2016); a disagreement over issues that could have been resolved through resolution methods (Brown and Marriot, 1993); or simply any unsettled contractual difference or disagreement (Love et al., 2011). For a dispute to occur, conflicts between participants should exist and there has to be a claim by one party while the other is rejecting it (Younis et al., 2008).

In simple form, basic relationships between the terms “conflict”, “claim”, and “dispute” used in this research is based on Figure 2.1 (Kumaraswamy, 1997).

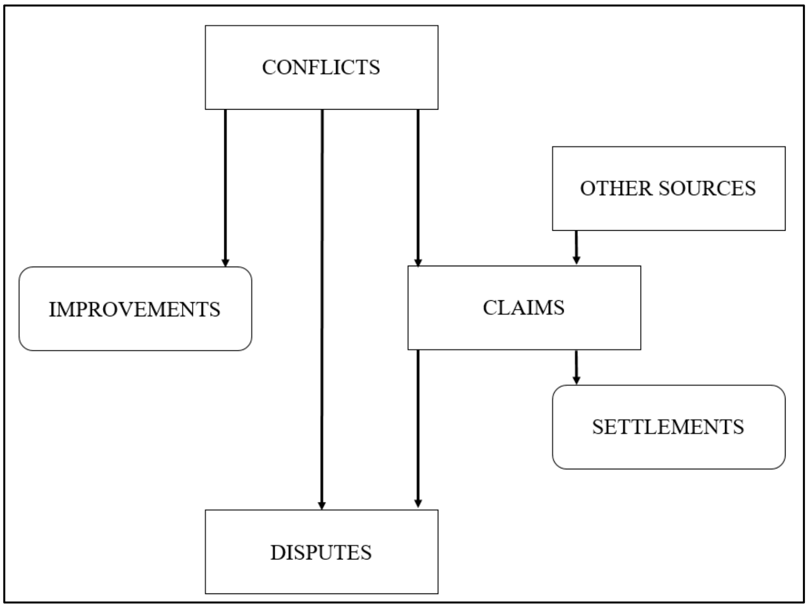


Figure 2.1. Conceptual Model for Basic Relationships between “Conflict”, “Claim”, and “Dispute” (Kumaraswamy, 1997)

Another important distinction between construction claims and disputes is that; a construction claim does not always indicates negativity or a bad situation. Upon satisfactory agreement on the claim, necessary modifications to contract are made and the problem is solved without further escalation. However, when parties cannot reach a satisfactory settlement on the claim, it proceeds to resolution methods such as negotiation, mediation, arbitration, litigation, etc. and becomes a construction dispute that negatively influences the construction project and the participant relationships. In construction industry, although it is a frequent practice to satisfactorily settle claims, disputes are also frequent (PMI, 2016).

2.1.2. Previous Research on Construction Disputes

In the introduction chapter, disputes and their negative impacts have been discussed in details along with statistical data proving that the number and severity of disputes are increasing in the industry. Moreover, the tendency in the current literature to perceive dispute occurrence as inevitable is mentioned. As a summary, if conflicts and claims are not satisfactorily settled, construction disputes may arise and resulting disputes are unwanted, time-consuming, and expensive as they can cause delays, budget overruns, tense relationships, decreased performance and quality, damaged long-term business relationships, etc. that damage the success of the construction project (Cheung and Suen, 2002; Harmon, 2003; Chen and Hsu, 2007).

In the light of foregoing observations, various topics on construction disputes research has attracted significant attention. As an initial step to understand mechanisms of dispute development, researchers tried to identify causes of disputes and project characteristics or attributes that impact dispute occurrence. Revay (1993) tried to find frequently re-occurring reasons for claims and ways to avoid those by analyzing 175 projects from the North American continent and highlighting seven different reasons. Common features of analyzed projects were their fast-track nature, incomplete bid documents, change orders, extra works, quantity fluctuations, and costs and extensions related to these. Kumaraswamy (1997) conducted an extensive research on conflicts, claims, and disputes in construction by establishing valuable frameworks. In this research, potential sources of conflicts, various time and cost related claim categories, and common sources of claims and disputes are identified. In a review on disputes settled by litigation in Australia, 117 different sources of disputes under 59 categories are discovered and the most frequent sources are highlighted as target areas for further research (Watts and Scrivener, 1993). Dalton and Shehadeh (2003) developed a statistically derived mathematical model based on past project data that links 14 independent variables to 3 dependent variables, which are cost overrun, time overrun, and number of claims. Cheng et al. (2009) indexed dispute cases based on eight attributes. Fenn et al. (1997) reviewed the literature to identify 46 sources of disputes, while Marzouk et al.'s (2011) literature review highlighted 44 causes. İlter and Dikbaş (2009) analyzed the frequency of disputes under 12 project related attributes, while İlter (2012) identified relationships between dispute factors and categories using 21 variables. Cheung and Pang (2013) developed an anatomy, which groups disputes in a five-leveled structure, linking 46 different attributes with dispute occurrence. In a study explaining the development of disputes, 24 construction disputes between states and contractors are comparatively analyzed to develop a process model and the proposed model examines the combined effect of project uncertainties, contractual aspects, business relationships, and problem solving effectiveness in order to develop a classification of problems while identifying resolution requirements and the potential dispute in each situation at the same time. Moreover, the study involves several recommendations to the construction industry within the light of the results that indicate dispute prevention is dependent on planning and problem solving capabilities of participants rather than contractual terms (Mitropoulos and Howell, 2002). In a similar research based on the case of Hong Kong's airport core program, the contract incompleteness, opportunism, and asset-specificity concepts are considered as root causes of disputes and it is claimed that actual extent of disputes is governed by the client through selection of procurement system, client organization, consultants, contracting team, and the time and cost prioritization (Yates, 2003).

There are many other studies identifying causes and attributes not as a primary goal, but as a secondary goal. For example, in an effort to predict the dispute proneness of projects, researchers categorized project characteristics, which affect dispute occurrence, under three main classes as people, project, and process related characteristics. People related characteristics involve participant related issues such as management capabilities, experience, etc. and their relationships with each other, while project related characteristics are composed of external and internal factors (i.e. site limitations, design complexity) and process related ones include planning and contract related aspects (i.e. financial planning, adequacy of technical plans) (Diekmann and Girard, 1995; Molenaar et al., 2000). Such an effort can be guiding as objective and subjective project characteristics that affect dispute occurrence are investigated at the same time. In another series of studies on dispute prediction and resolution method selection, researchers identified that 13 different attributes have an impact on dispute occurrence. While in resolution method selection, dispute category and occurring phase are also added as impacting factors to make 15 variables (Chou and Lin, 2012; Chou, 2012; Chou et al., 2013a; Chou et al., 2013b; Chou et al., 2014). In an effort to classify influential information to discover rule sets for construction disputes and resolutions, 13 input variables are linked with four output variables as dispute occurrence, dispute type, resolution method, and dispute stage (Chou et al., 2016). In pursue of finding a method to predict occurrence, frequency, and level of possible claims in Iran construction industry, 60 primary causes of claims are grouped under nine different categories based on their nature (Yousefi et al., 2016).

Another group of studies can be found in research aiming to analyze the outcome of claims, disputes, litigation, etc. There are studies focusing on dispute causes and project characteristics that affect court rulings. In another perspective, these efforts can be viewed as identification of causes and characteristics that influence dispute development. Chau (2007) declared 13 distinct project characteristics affect court rulings, while Chen and Hsu (2007) and Chen (2008) discovered 17 dispute causes and six project characteristics. In a series of studies to predict the outcome of construction litigation, researchers identified that 38 attributes have impact on litigation rulings (Arditi and Pulket, 2009; Pulket and Arditi, 2009a; Pulket and Arditi, 2009b). In a system that generates legal arguments for change order disputes based on past cases, El-adaway and Kandil (2009) utilized 12 different impacting factors to classify arbitration and litigation cases. Kilian and Gibson (2005) examined primary causes of litigation associated with construction contracts of the U.S. Naval Facilities and linked seven claim causes and 17 project characteristics with litigation rulings. In a review on disputes handled by the Turkish Public Procurement Authority (TPPA), rulings of the Authority are classified under attributes such as applicants' organizational structure, type of government agency in charge, and phase of occurrence (Gencer, 2005).

All these mentioned causes, reasons, characteristics, and attributes are reviewed and similar items are merged. Resulting items are used in order to establish a conceptual model that is composed of a pool of input variables, which will be used for the design of a questionnaire to collect empirical data and the development of prediction models proposed in this thesis study.

The research on construction dispute domain is not limited to identification of causes of disputes and attributes affecting dispute development. For example, Gebken and Gibson (2006) grouped the research on dispute management domain under three main areas; (1) dispute identification, which focuses on causes and identification methods; (2) dispute assessment, which addresses studies on determination and quantification of dispute frequency and severity; (3) dispute control, which deals with ADR methods and other management strategies. Unfortunately, the scope of this thesis study is limited to dispute prediction (occurrence and potential compensation) and resolution method selection. Therefore, the thesis study will skip several topics related to construction disputes and continue with the literature review on dispute prediction.

2.1.3. Literature Review on Dispute Prediction

One of the earliest studies on predicting the dispute likelihood of construction projects is conducted by Diekmann and Girard (1995). They developed a ‘dispute potential index (DPI)’ using Logistic Regression (LR) analysis on a dataset of 159 construction projects with dispute predictors categorized under project, people, and process related aspects. An improvement is achieved by analyzing the same dataset by the Structural Equation Modeling (SEM), a technique that creates surrogate variables integrating several qualitative and quantitative input variables in order to reveal undiscovered relationships between pairs of input variables via regression operations (Molenaar et al., 2000). Both studies aimed to predict dispute propensity of construction projects at early stages. However, construction disputes require consideration of numerous complex and interrelated factors that are difficult to rationalize (Chou, 2012). Therefore, results from techniques like LR and SEM, which have limited capability of modeling multiple correlations between variables, can be misguiding.

Dalton and Shehadeh (2003) claimed that knowing various factors related to a construction project, it is possible to forecast the number (occurrence) and value of claims (compensation). For this reason, they developed a statistically derived mathematical model based on past project data. However, their methodology also lacks an approach for mapping complex relationships among their variables. In addition, even with the best available information, it is impossible to quantify claims precisely due to uncertainties on factors such as quantity of impacts, indirect costs, cumulative effects (i.e. loss of productivity), etc. (Ren et al., 2001). In another statistical based study, the frequently re-occurring reasons for claims are presented by analyzing 175 construction projects to investigate if disputes can be avoided or not, especially when there is a change order (Revay, 1993). Analyzed projects are categorized based on various attributes such as type of the project, value of the contract, location, etc. in order to link some project attributes and dispute reasons with time and cost overruns. Although the scope does not include a model or a system, the study is valuable by giving statistical insights about associations among various project attributes and dispute reasons with dispute occurrence and potential compensations.

There are studies for assessing the occurrence likelihood of a dispute by utilizing AI applications. One of these studies developed an anatomy of construction disputes using fuzzy sets and fault tree analysis. The proposed anatomy identifies critical dispute related factors and assesses the occurrence likelihood of causes of disputes, which will gradually lead to dispute occurrence likelihood evaluation (Cheung and Pang, 2013). Specific to Iran construction industry, Yousefi et al. (2016) benefited from advantages of the Analytical Hierarchy Process (AHP) and MLP neural networks to predict the occurrence of claims along with estimations of time and cost compensations related to them.

A series of studies on a dispute dataset of 584 PPP projects undertaken by the TPCC prove the efficiency and effectiveness of ML techniques in dispute prediction problems. The first of these studies utilizes kNN, MLP, Naïve Bayes, SVM, and C4.5

algorithms as single (base) classifiers. Then, in pursuit of enhanced classification performance, ensemble classifiers are established combining k-means clustering technique, MLP classifier, and C4.5 algorithm, respectively, with the mentioned base classifiers one by one. Classification performances are compared with each other considering metrics such as accuracy, false positive (FP) and false negative (FN) rates, and the area under the receiver operating characteristic curve (AUROC). It is highlighted that the prediction performance of ensemble models outperformed the classification performance of single classifiers (Chou et al., 2013a). In the second study, the same dataset is analyzed by several techniques involving LR, discriminant analysis (DA), ANN, SVM, decision list (DL) algorithm, tree augmented Naïve Bayes (TAN) classifier, classification and regression trees (CART), quick-unbiased-efficient statistical trees (QUEST), C5.0 algorithm, and exhaustive Chi-Squared automatic interaction detection (exhaustive CHAID) algorithm. Then, combining these single classifiers, several ensemble models are established and their classification performances are experimented. According to 10-fold cross-validation performances, the highest dispute prediction accuracy rate was 84.33% resulted from the ensemble model that combines SVM, ANN, and C5.0 classifiers (Chou and Lin, 2012). The third study focuses on SVM algorithm in specific by trying to optimize its parameters utilizing genetic algorithm (GA) in order to enhance the prediction performance. The GA based SVM model resulted in a dispute prediction rate of 89.30% (Chou et al., 2014). Finally, in the fourth study, the dispute occurrence and the dispute type are predicted by C5.0 algorithm with an average 10-fold cross-validation accuracy of 83.92% and 77.00%, respectively, and the dispute stage with an average 10-fold cross-validation accuracy of 79.77% by ANN classifiers (Chou et al., 2016). Although these studies are project and industry specific, they are pioneering studies in dispute prediction research.

Similar to predicting the dispute likelihood of construction projects, there are studies predicting the litigation likelihood. Based on 340 litigated cases from the period 1972-2002 in the U.S., a hybrid model is developed combining ANN and case-based reasoning (CBR) techniques in order to classify construction projects with change orders according to their litigation likelihood. The ANN part of the model achieved a classification accuracy rate of 84.61% on possible litigation likelihood, while the CBR part presented similar past litigious cases as examples (Chen and Hsu, 2007). The same dataset of change order related litigious cases is classified by a kNN-based model that achieved 84.38% accuracy of predicting the litigation likelihood (Chen, 2008). Although both studies can be used as early warning systems for change order related disputes, datasets are composed of litigious cases only, which ignores valuable knowledge that could have come from cases solved by other resolution methods. In addition, the research is specific to the U.S. construction industry and change order related disputes.

Respecting the state of the mentioned existing research on dispute prediction in construction industry, it can be observed that studies have limitations such as being industry, project type, dispute type, and contracting strategy specific. In other words, datasets are composed of projects from a certain construction industry (i.e. cases from the U.S. only), exhibiting a certain project characteristic (i.e. public projects only), focusing on certain dispute types (i.e. change order related disputes), and having a certain contracting strategy. Global-scaled models that consider various project, dispute, and contract types do not exist in the literature. Although considering the differences between various construction industries such as law systems, etc., a local or characteristic-specific modeling approach may seem appropriate. However, a global-scaled model with considerations on such variations would definitely be a better approach.

Due to their limited capabilities in discovering complex and interrelated factors between input variables, prediction studies that do not utilize ML techniques can be viewed as insufficient. Considering the multitude of participants, various sources of uncertainties, and numerous variables in construction industry, the utilization of ML techniques in construction dispute domain is a necessity. Leaving the studies that do not utilize ML techniques aside, another observation on the dispute prediction research is that there are only few studies experimenting the usage of various ensemble models. Instead, majority of results are based on single algorithms or specific combinations of algorithms. However, ML is an experimental science that requires diverse experiments to find models with better generalization performance (Arditi and Pulket, 2009; Witten et al., 2016). Moreover, performances of such algorithms should be evaluated considering many performance metrics such as TP and TN rates, AUROC, kappa statistics, etc. Contrarily, the current trend in the literature is to evaluate the performance of an algorithm depending on the accuracy measure only. Chou et al. (2013) supports all these ideas by stating that previous research is generally focused either on specific change order disputes or on conventional contracting projects, which means ignoring variations in the project environment and characteristics, using a single accuracy performance measure. Therefore, this thesis study aims to fill the mentioned gaps of the research by developing ML based dispute prediction models that utilizes construction project data from various construction industries. Moreover, projects in the dataset will be diverse in terms of project types, contracting strategies, dispute types, etc. Finally, not only performances of several single algorithms will be experimented using various performance measures in the ML domain, but also various ensemble models combining these single classifiers will be experimented.

2.2. RESEARCH BACKGROUND ON DISPUTE RESOLUTION METHODS AND ADR TECHNIQUES

Resources in construction projects are limited and disputes divert them from the primary goal, which is successful project completion (Harmon, 2003; Fenn, 2007). Considering these scarce resources, construction projects should pursue effective dispute management (Cheung et al., 2010). This is because dispute management is one of the main determinants of the performance of construction projects and consequently the industry (İlter and Dikbaş, 2009). The current complex state of the construction industry requires project managers to be equipped with necessary skills to enforce effective dispute management and resolving disputes has become an inevitable duty for managers (Cheung, 1999). The content of this duty involves diverse activities ranging from the selection of a resolution method to active participation in negotiations; thus, a manager is obliged to have a comprehensive understanding of various dispute resolution methods (Cheung et al., 2010). Therefore, this section will start with an effort to generate an understanding of resolution methods with special emphasis on ADR techniques. Besides the conventional resolution methods, which are litigation and arbitration, a comprehensive explanation of ADR techniques, which involve DRB, mediation, SEA, and negotiation, will be given. The ADR techniques to be reviewed in this thesis study is limited to the techniques used in the projects from the collected dataset.

2.2.1. Conventional Dispute Resolution Methods

In construction industry, litigation is the traditional form of dispute resolution (Jones, 2006). However, besides litigation, this research will consider arbitration process as another method of conventional dispute resolution. Although arbitration is available in the industry since the late 1800's as an ADR method, with the growing dissatisfaction on arbitration processes, its perception as an ADR technique is under discussion (Harmon, 2003; McGeorge et al., 2007). The recent consensus in the literature is to regard arbitration as a replicate of litigation due to increased procedural complexity over the years (Cheung et al., 2002). Similarly, in Construction Extension to PMBOK Guide, it is stated that arbitration is one-step short of litigation (PMI, 2016). Therefore, arbitration can said to be like the private enterprise version of the court system (Jones, 2006).

2.2.1.1. Litigation

Litigation is the process that involves the determination of the dispute in a court, generally, before a judge. Regardless of resolution processes defined in a construction contract, if the parties do not comply with these procedures, they have a right to appeal to courts ultimately (Jones, 2006). Litigation is a rigid process that is subject to formal rules and procedures set out by courts (Cheung, 1999). It is the final destination to settle disputes and known as the most formal way (Çevikbaş and Köksal, 2018).

Although litigation is usually perceived as the final resort that is preferred after failing to achieve desired outcomes from previous methods, in some cases, parties may prefer proceeding to courts directly with a perception of litigation as the best-defense mechanism (McGeorge et al., 2007). In other words, upon not satisfactorily resolving a conflicting situation or a dispute and remaining inconclusive after exhausting the non-binding options, litigation, which is the traditional form of reaching binding solutions, can be employed as a compulsory form of dispute resolution (Jones, 2006). Among various attributed advantages, being a formal process and having a binding nature are the main advantages of litigation process (Mahfouz and Kandil, 2011).

There are serious criticisms on litigation processes, but it can still be claimed that it is the most effective way of resolving a dispute when a party is not willing to achieve a resolution. The same claim applies when there are substantial legal implications and allegations of dishonesty (McGeorge et al., 2007). As details will be explained in upcoming sections, although benefits of ADR techniques are clear, they generally require a negotiated agreement between disputant and disputed parties. However, there can be frustrations in reaching such agreements that will lead the way to litigation (Lipsky and Seeber, 1998). Litigation is one of the most common methods used to resolve disputes (Çevikbaş and Köksal, 2018). In addition, it is a sensible resolution method for parties pursuing binding and enforceable solutions (Jones, 2006). Another advantage is its usefulness in determination of the right and wrong in a formalized way that is dependent to facts and laws instead of emotions (Cheung, 1999; McGeorge et al., 2007).

Despite mentioned advantages of litigation, researchers highlighted numerous shortcomings of the process in the literature. To begin with, litigation is a complex process that requires usage of significant resources and, generally, legal representations (McGeorge et al., 2007). These legal representatives generally use every tactic available to them, usually in an adversarial way, with the primary concern to win the case, which may damage relationships between plaintiff and defendant parties (Harmon, 2003). Moreover, characteristics and structure of judicial justice vary from country to country due to differences in judicial systems, enforcing laws, and organizational structures (Çevikbaş and Köksal, 2018). This may cause problematic issues specifically in international construction projects, as parties may not be familiar with the legal system of the host country. Another disadvantage of litigation is the lack of confidentiality. Litigation is not a confidential process, which is open for public and media viewing (Harmon, 2003). In addition, when litigation is preferred, parties delegate their controls over the case entirely to a third party, generally a judge, who will enforce the outcomes of the hearings to both parties (Cheung, 1999; Gebken and Gibson, 2006).

Unsettled disputes from large and complex construction projects usually cause complex construction litigation due to factors such as the high technical and financial complexity of matters related to these disputes, numerous parties involved and affected in such projects, thousands of activities, documents, and facts associated with such cases, etc. (Harmon, 2003). These technical and complex engineering disputes are being resolved by judges (Jones, 2006). In order to review disputes in this nature appropriately, people with combined expertise and experience both in legal and construction domains are needed. However, lawyers competent in engineering subjects or engineers with legal expertise are not easy to find (Cheng et al., 2009). This skill set combination can rarely be found and experts of this nature require high salaries (Mahfouz and Kandil, 2011). Thus, the nature of the construction litigation can be characterized as specialized, complex, and expensive (Arditi and Pulket, 2009). Besides these negative characteristics associated with litigation, another negative aspect of the process is the increased hostilities between parties (Gebken and Gibson, 2006). The ruling of the court will declare a winner and a loser to a dispute and such a situation will make parties compete with each other, which will result in an adversarial process. This adversarial environment can eliminate job profits and the possibility of future work by damaging good working relationships with the counter party and the long-term business relationships. As a result, unrecoverable costs may be generated even if the outcome of the litigation was in favor, which turns the winner of the case into a loser in reality (Harmon, 2003).

Among all these disadvantages, the most striking negative aspects are the costs associated with the litigation process and the length of the duration required for the hearings in the court. Litigation is usually the most expensive and time-consuming solution (Chen, 2008). Besides the indirect costs, the direct costs associated with litigation process is also high that makes it an expensive process (Kassab et al., 2006; Chau, 2007). Although there may be situations where avoiding litigation is more costly than preferring it, this preference is usually purchased at great costs (Arditi and Pulket, 2009). In addition, when the case comes to the court, a protracted period for discovery is required due to the great volume of documents involved in a typical construction dispute to be heard (Harmon, 2003; Jones, 2006). Corroborative statistical data can be found in the literature. For example, in a study on Turkish construction litigation, it is revealed that the average duration of the proceedings is almost equal to the duration required to complete an average construction project (İlter, 2010a). In another research, it is claimed that depending on the jurisdiction, the proceedings of a complex construction dispute may take between two to six years (Mahfouz and Kandil, 2011).

Litigation is generally viewed as the worst resolution method in most countries due to mentioned disadvantages (PMI, 2016). There is a consensus in the literature on litigation avoidance and several researchers recommended to avoid litigation (Chen and Hsu, 2007; Chau, 2007; Chen, 2008; Pulket and Arditi, 2009b; Cheung et al., 2010). However, it may be the only way to resolve a dispute in some cases. Nevertheless, it can be said that instead of hoping for a favorable judgment through litigation, it is usually better to resort to other resolution methods even when the alternative method is not the most fruitful one for the disputant party (PMI, 2016).

In short, positive aspects of construction litigation can be listed as (1) bindingness and enforceability, (2) being just, and (3) reaching a solution regardless of the willingness of parties. On the other hand, negative aspects associated with litigation can be listed as (1) high costs, (2) low resolution speed, (3) adversarial nature (not preserving relationships between parties), (4) having no confidentiality in the process, (5) no flexibility in rulings, and (6) no control over the process or proceedings.

Subsequent to giving an understanding on overall concepts related to construction litigation along with its advantages and disadvantages, the thesis study will continue with the second conventional dispute resolution method, which is the arbitration.

2.2.1.2. Arbitration

The OGC (2002) defines arbitration as a formal, private, and binding process where the dispute is resolved by the decision of a nominated third party, which can be an arbitrator or a panel of arbitrators. According to the AAA, arbitration is the binding resolution of a dispute outside the court by the decision of an impartial third party and this method is faster and less costly compared to litigation (AAA, 2019a). In litigation, a judge reviews the case. Contrarily, in arbitration, an expert or a panel of experts, which generally include construction experts, attorneys, or retired judges, review the dispute as a judge and jury to receive arguments and testimony, assess facts and documents, and reach an impartial judgment based on facts and evidences (PMI, 2016). In a typical construction arbitration, a panel of three arbitrators are involved; both parties select one arbitrator each and the third one is selected either by mutual agreement of parties or by the administrative organization. However, parties can determine issues related to arbitration processes such as the number of arbitrators, administrative organization, location, regulations to be followed, etc. in their contracts (Peña-Mora et al., 2003). In order to be able to use arbitration as a resolution method, parties should agree on the terms of arbitration via contract documents (İlter, 2010a). Arbitration clauses are involved in standard forms of contracts and they are being widely used in both private and public construction contracts (Harmon, 2003). Rules of arbitration are generally set out by international courts or administrative organizations (PMI, 2016).

Similar to litigation, arbitration is generally preferred as a final option that is employed upon exhausting the non-binding resolution methods. Another similarity is that the judgments resulting from arbitration processes are binding on both parties. In addition, outcomes are enforceable both domestically via host country's law system and internationally via New York Convention, which is an accepted agreement by a large number of countries (Jones, 2006). There are other international organizations (i.e. International Center for Settlement of Investment Disputes (ICSID)) and agreements (i.e. Geneva Convention) that contributes to domestic and international arbitration (İlter, 2010a).

High costs and prolonged proceeding durations associated with litigation encouraged the construction industry to utilize the less formal adjudicatory process, the binding arbitration (Mitropoulos and Howell, 2002). Indeed, it is a faster and cost-effective alternative to litigation (Peña-Mora et al., 2003). Moreover, instead of resorting to local law system of the host country, arbitration is a more attractive preference in international construction projects as it can form a common ground for international parties of different law systems (Jones, 2006). In addition, the arbitration process is superior to litigation in terms of confidentiality since arbitral proceedings can be kept confidential (OGC, 2002; Harmon, 2003)

Despite mentioned advantages, arbitration process also has several shortcomings. To begin with, the conventional resolution methods are characterized as costly and time consuming (Cheung et al., 2010). Although simplifying arbitral proceedings contributes to mitigate the time and cost related criticisms, cost and speed are still the main disadvantages of arbitration, especially in complex disputes (Jones, 2006). Moreover, the current state of the arbitration is no different from litigation with regard to confrontational and adversarial processes involved. There is a declaration of a winner and a loser upon judgments of a third party, which eliminates the possibility to achieve constructive remedies for the disputed case (Harmon, 2003). Within the light of these, it can be said that while litigation is generally viewed as the worst resolution method in most countries, it is followed closely by arbitration (PMI, 2016).

In short, positive aspects of construction arbitration can be listed as (1) being a more attractive preference for international construction projects, (2) bindingness and enforceability, (3) experienced experts with combined skill sets in engineering and law can review the cases (4) confidentiality, (5) impartiality, (6) relatively low cost compared to litigation, and (7) relatively fast resolution compared to litigation. On the other hand, negative aspects associated with arbitration can be listed as (1) high costs compared to ADR techniques, (2) low resolution speed compared to ADR techniques, (3) adversarial nature (not preserving relationships between parties), (4) no flexibility in the rulings, and (5) very little control over the process or proceedings (i.e. selection of arbitrators, administrative organization, location, etc.).

As a final remark, primary shortcomings of conventional resolution methods, which are high costs, prolonged resolution durations, and adversarial environment, encouraged a rapid growth in ADR processes (Cheung et al., 2002). To give an insight of these processes, the thesis study will continue with details on the ADR methods.

2.2.2. Alternative Dispute Resolution (ADR) Methods

In the literature, there are several different definitions with little variations for ADR methods. These variations result from the perception of the researcher on resolution methods. For example, Gebken and Gibson (2006) defines ADR as any binding or non-binding method of resolving disputes, which involves both self-deterministic and third party intervened methods, other than litigation. According to this definition, ADR methods involve all methods excluding litigation only. This is compatible with the definition by the Construction Extension to PMBoK Guide, which defines ADR methods as less expensive alternative techniques to litigation in a court of law (PMI, 2016). With another perspective, Marzouk et al. (2011) defines ADR as any binding, non-binding, and preventive process or procedure contributed by an impartial third party excluding adjudication by a judge. Thus, this definition excludes self-deterministic methods in addition to litigation. Jones (2006) considers non-binding and less formal (compared to litigation and arbitration) administrative resolution processes as ADR techniques where parties reach a solution by themselves or by the help of a third party. The perception of ADR in this definition is the same as Cheung (1999), who categorizes all methods as ADR except formalized methods, which are litigation and arbitration. Considering all these differences, this thesis study should state its own perception of ADR.

In this thesis study, ADR techniques refer to all non-adversarial processes that aim to allow parties to apply their own applicable solutions without establishing formal frameworks so that good business relationships can be preserved. While applying ADR techniques, third party assistance can be acquired if it is necessary. More simply, any dispute resolution method will be referred as an ADR method in this thesis study other than the conventional ones, which are arbitration and litigation. In other words, ADR methods involve techniques ranging from self-deterministic methods such as negotiation and SEA to third party intervened methods such as mediation and DRB.

Conventional dispute resolution methods in construction industry are characterized as inefficient in terms of cost and duration (Cheung et al., 2010). Besides the associated high costs and long durations, conventional methods also cause adversarial relationships between project participants (Gebken and Gibson, 2006). Moreover, in order to avoid a bad reputation, construction companies will try to avoid involvement in prolonged disputes, especially in public projects. In such a case, resolving disputes by ADR techniques without resorting to conventional methods would be more advantageous for companies (Jones, 2006). Based on a research conducted by Günay and Birgönül (2001) on Turkish construction industry, it is revealed that contractors to public projects do not prefer to solve disputes in courts due to the fear of damaging their reputations and the possibility of future works (İlter et al., 2007). Therefore, although conventional methods are well-developed formal techniques for construction dispute resolution, the industry is in search of methods to solve disputes equitably, economically, and quickly without damaging the relationships (Cheung, 1999; Arditi and Pulket, 2009). Although prevention techniques do not guarantee total dispute avoidance (Cheung, 1999), this search can be addressed by ADR techniques.

ADR is composed of contractual dispute resolution mechanisms. This means that there should be a contractual agreement between parties to apply ADR methods in dispute resolution (Harmon, 2003). Such an agreement before occurrence of a dispute indicates the good will and the commitment of parties to achieve cooperative project success. In addition, unlike the conventional resolution methods, ADR techniques are generally not binding and the parties can leave these methods to resort to conventional ones without losing their legal rights any time they feel uncomfortable or inconclusive. ADR outcomes can only be enforceable upon signing a settling agreement (Cheung et al., 2002). Thus, parties' willingness to reach a settlement is an important determinant of success in resolution through ADR methods.

The driving force in establishing ADR methods was to develop cost efficient and fast alternatives to conventional resolution methods (Harmon, 2003). Similarly, studies that recommend utilization of ADR techniques claim that ADR can overcome shortcomings of conventional methods by offering the chance to reach prompt and economic resolutions (Cheung et al., 2010). Legal fees, management resources to be used, and costs of resolution will be decreased in ADR methods compared to conventional ones. Moreover, ADR techniques do not require strict procedural rules and involvement of legal professionals that result in considerable timesaving (Cheung, 1999). Another time related advantage is that they enable discussion of disputes without interrupting the course of construction and delaying the works (Rubin and Quintas, 2003).

There are several other benefits of ADR methods. To begin with, ADR processes tend to be less formal than conventional resolution methods (Jones, 2006). Moreover, unlike the case in conventional methods where the right and wrong parties in a dispute are declared along with consequences to be enforced, the goal of ADR techniques is to assist parties in reaching their own workable, agreeable, and commercial solutions in a cooperative way (Jones, 2006). ADR techniques can provide tailor-made solutions that suit best to the dispute at hand and consequently, they can enable flexibility in dispute resolution (Rubin and Quintas, 2003). They enhance the communication, teamwork, and harmony between parties so that hostilities and grievances resulting from the adversarial environment of conventional resolution methods are avoided (Cheung, 1999). In addition, the control of parties over the process is increased in ADR methods compared to conventional ones that delegate the control entirely to a third party judgment (Gebken and Gibson, 2006). Users of ADR methods can also benefit from confidentiality and privacy in resolution processes (Cheung et al., 2002; Gebken and Gibson, 2006).

In the light of foregoing observations, ADR methods can be associated with greater efficiency, lower costs, fewer formal procedures, and improved relationships built on consensus rather than conflict (Rubin and Quintas, 2003). In addition, research has proven that upon prompt and appropriate utilization of ADR methods, relationships and trust between parties can be strengthened and, more importantly, win-win outcomes can be generated (Kassab et al., 2010).

In upcoming sections, ADR techniques that are utilized in projects from the collected dataset will be introduced. These techniques are DRB, mediation, SEA, and negotiation. Figure 2.2 is the review of dispute resolution techniques from this research’s point of view.

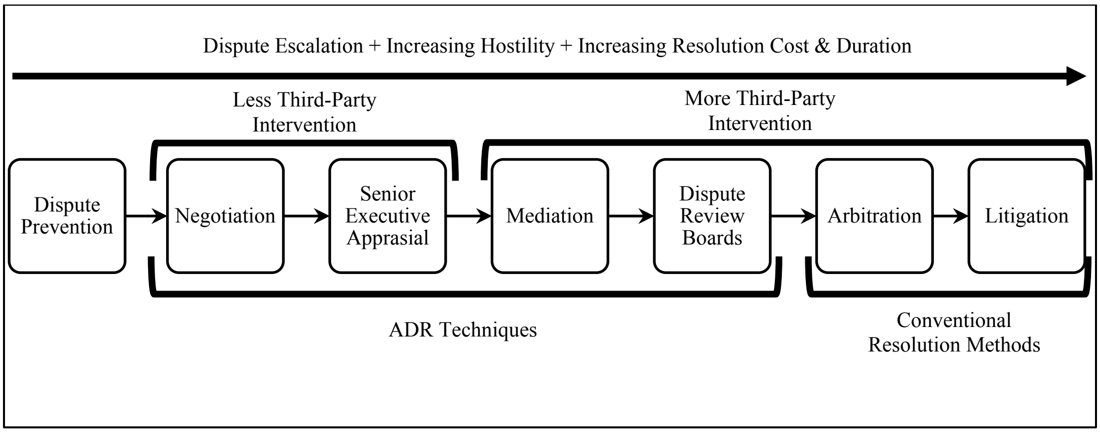


Figure 2.2. Review of Dispute Resolution Techniques

2.2.2.1. Dispute Review Boards (DRB)

DRB is a unique, proactive, and non-adversarial contractual dispute resolution method that mitigates or resolves disputes during the course of construction by the decision of a board, which is established at the beginning of the project before any dispute occurrence (Harmon, 2003). In other words, DRB is a resolution process in which parties submit summarized versions of their cases to a preselected panel of experts with the authority to settle disputes (PMI, 2016). Starting first in major infrastructure projects in the U.S., DRB has been extensively used since 1970's as an ADR technique in construction industry (Menassa and Peña Mora, 2009). The method gained attention especially after being presented as the principle method of dispute resolution within the 'International Federation of Consulting Engineers (FIDIC) Standard Forms of Contract' and becoming mandatory in the 'Procurement of the Works' by the World Bank in 1999 (McGeorge et al., 2007). The most common usage of DRB is in the FIDIC Conditions of Contract for Construction under the 'Clause 20 (Claims, Disputes and Arbitration)', which is named as 'dispute adjudication board (DAB)', and under sub-clauses 'Sub-Clause 20.2 (Appointment of the DAB)', 'Sub-Clause 20.3 (Failure to Agree DAB)', and 'Sub-Clause 20.4 (Obtaining DAB's Decision)'. The World Bank contracts employ DRB under the name of 'dispute resolution boards' (İlter, 2010a).

Project participants set details of the DRB, the scope of its authority, and its procedural rules via contracts according to their needs and project requirements (Jones, 2006). A typical DRB is composed of three independent and neutral experts with no financial or business relationships with project participants (PMI, 2016). Typically, each party selects one expert and selected experts determine the third one. There can be more experts in the panel if required (İlter, 2010a). There should be no questions with regard to DRB's integrity and impartiality among project participants. Therefore, the panel of experts is chosen by mutual agreements of parties. The DRB is established before the start of the construction and their duration of work generally covers the entire project duration unless otherwise is stated in the contract (Harmon, 2003). Generally, both parties equally share the associated costs (Peña-Mora et al., 2003).

DRB regularly visits the construction site and project participants regardless of dispute occurrence. By doing this, DRB keeps itself updated with a good understanding of the project, progress, and parties. Depending on the desired level of involvement and the rules set by the contracts, DRB can organize meetings periodically (Peña-Mora et al., 2003). Additionally, they are informed regularly by documents from both parties such as written progress reports, minutes of meetings, etc. and they are present on site upon request of any party or at times of critical construction events; thus, the major strength of DRB method is its always up-to-date informed nature (Harmon, 2003). As a result, the DRB not only responds to disputed issues, but also identifies emerging problems and conflicts by making recommendations, facilitating negotiations, etc. that makes it a dispute prevention method (Jones, 2006). Upon inevitable occurrence of a dispute, both parties can request a confidential hearing where either party is given a reasonable opportunity to present their cases. These hearings are not formally structured, but in order to be able to present the case to the DRB, parties should exhaust resolution procedures defined in the contract and remain inconclusive (Harmon, 2003).

There are three main approaches with regard to bindingness of DRB decisions. In the first approach, the DRB can serve in an advisory role with no bindingness (PMI, 2016). In the second one, judgments of the DRB are binding, but objection and further review are possible through arbitral or court proceedings. In other words, DRB decisions are binding unless it is overturned by arbitration or litigation. However, objection to DRB decisions are not common as they are based on informed decisions of highly experienced and qualified experts. Finally, in the third approach, decisions made by the DRB is final and binding on parties (Jones, 2006). The selection of bindingness of the DRB is regulated in contract agreements.

The driving force in establishment of DRB was the requirement of timely resolution of the issue of responsibility in complex disputes and projects. DRB is the prime candidate to fulfill this requirement thanks to its rapid determination of responsibility and cooperative problem solving assistance (Mitropoulos and Howell, 2002). Another goal is the early response to conflicts before they turn into disputes or the early response to disputes for timely settlement. When a conflict or a dispute arise, the DRB simultaneously responds to the problem and its success is considered as having no unresolved issue at the project completion (Harmon, 2009). In addition, as construction disputes involve complex technical issues rather than legal ones, the usage of DRB is advantageous since experts can provide complex engineering solutions on these technical issues (Peña-Mora et al., 2003). As stated earlier, DRB method is a non-adversarial and cooperative one that facilitates good working relationships and future business opportunities by avoiding hostilities (Harmon, 2003). Moreover, DRB is a less formal process and even hearings can be held in the form of site meetings that aim to have an investigating approach rather than being adversarial (Jones, 2006). DRBs usually search for commercially viable, acceptable, and workable solutions to problems. In addition, research has proven the effectiveness of DRB method. In a study analyzing the application of DRB in construction projects covering the period 1975-2007 in the U.S., it is revealed that the number of projects with DRB panel per year is increasing rapidly with a considerable prevention and resolution effectiveness. Among DRB utilizing projects, 51% ended up with no disputes, which is a statistic proving the prevention effectiveness. On the other hand, among the dispute occurred projects, DRB decisions settled disputes in more than 90% of all cases (Menassa and Peña Mora, 2009).

In short, positive aspects associated with the DRB method can be listed as (1) almost simultaneous timely response, (2) domestically and internationally applicable nature, (3) informed decisions by experienced experts, (4) flexibility in the process and procedures depending on needs of the parties and the project, (5) adjustable degree of bindingness, (6) dispute preventive mechanisms, (7) confidentiality, (8) impartiality, and (9) associated high costs can be recovered by dispute prevention or mitigation.

Despite mentioned advantages, the DRB method is also criticized in the literature. In a case study reviewing the effectiveness of the DRB method, it is revealed that barriers to DRB effectiveness can be listed as (1) elongations in the resolution process, (2) viewing the DRB process as adversarial, (3) length of time required to get prepared for DRB hearings, and (4) receiving unconvincing recommendations. Although, in the literature, there are assertions that presence of DRBs as a resolution method reduces bid prices due to better risk allocations, the same research also claims that bid prices are generally not reduced as a result of DRB inclusion (Harmon, 2009). Indeed, the process is expensive and corroborative data is present in Jones (2006), which gives the estimation of American Society of Civil Engineers (ASCE) on DRB costs being between 0.04% and 0.51% of the project costs. However, increased costs due to inclusion of the DRB can save substantial amounts later depending on problematic issues during the construction.

2.2.2.2. Mediation

According to the AAA, mediation is an informal form of negotiation in which a third party encourages participants in a dispute to reach their own solutions in order to preserve business relationships (AAA, 2019b). The third party is known as the mediator and although it is common to mediate disputes by one mediator, more mediators can be involved in the process if required (Arıcı, 2012). In a more comprehensive definition, mediation is defined as non-binding negotiation sessions facilitated by an impartial third party in pursue of mutual agreements of participants for dispute settlement (Harmon, 2003).

Mediation is a very common method of ADR in construction industry. Construction professionals have been utilizing mediation technique since 1980's and it is the most rapidly growing form of ADR (Harmon, 2003). This method is especially attractive in industries like construction industry where preserving continuous business relationships are important (Özer, 2012). Indeed, construction attorneys also perceive mediation as the most effective method for achieving a wide variety of goals ranging from preserving relationships to enhancing communication (Peña-Mora et al., 2003).

Mediation is generally the last step before resorting to arbitration or litigation (Peña-Mora et al., 2003). Unlike the conventional dispute resolution methods, mediation is a voluntary and non-binding process (Cheung et al., 2004b). Thus, the initial condition of mediation is the agreement between parties to resort to a mediator (Arıcı, 2012). The most important actor in mediation process is naturally the mediator and the mediator is nominated by mutual agreement of parties or by a nominating organization (McGeorge et al., 2007). The mediator helps to overcome difficulties faced during disputes by facilitating negotiations between parties, evaluating alternatives, and guiding parties to develop their own creative solutions with the purpose of mutual agreement (Yiu et al., 2006). While doing this, the mediator should be impartial. Parties should not suspect the impartiality of the mediator, otherwise the trust cannot be established and the process will be inconclusive (Jones, 2006). Therefore, the mediator should be objective, convincing, trustworthy, and respected at all times during the process (Özer, 2012). For this reason, mediators are usually from the field of law or social working industries with a training in communications and negotiation skills, instead of being an expert in the dispute-relevant field (McGeorge et al., 2007).

Another important point is the role of the mediator in the process. Researchers claim that if mediators offer opinions, they will distort the process; therefore, a good mediator does not offer opinions on ways to settle the dispute, instead guides the parties to a mutual agreement (Harmon, 2003). Unlike conventional techniques that depend on judgments of third parties and limit the control of participants over the process, mediation establishes a flexible environment depending on consensual agreements and parties can control the process at the same time (Özer, 2012). A mediator has no binding authority to make or enforce a decision on any matter and its only role is to guide the parties through the process (Peña-Mora et al., 2003). In other words, mediator is not an advisor making suggestions, but an assistant in negotiations by encouraging parties in identification of issues and solutions (McGeorge et al., 2007).

One of the primary concerns in development of the mediation technique was to present a speedy and low-cost alternative to conventional resolution methods. Indeed, mediation is a fast resolution method thanks to its informal nature and decreased procedural rules. Resulting from being a speedy process that does not require formal representations, mediation is a cheap undertaking (Cheung et al., 2004b). Corroborative data can be found in the literature. In Jones (2006), based on the Australian Commercial Disputes Centre, the cost of mediation is estimated as 5% of the cost of arbitration or litigation. Peña-Mora et al. (2003) reveals that among mediations reported to the AAA, more than 50% costs 3,000 U.S. Dollars or less, while less than 10% costs more than 20,000 U.S. Dollars. In addition, 50% of mediations are conducted in 2 days and less than 10% lasted longer than 6 days. İlter (2010a) stated that majority of mediations are settled in one day by only one session. Therefore, it can be said that mediation is a wise alternative to conventional methods in terms of cost and resolution duration when parties are willing to reach a settlement.

Mediation is appropriate when the parties pursue an informal, objective, and analytical assessment of their cases and there is a continuing business relationship (Jones, 2006). Despite the informal nature, mediation process has become more formalized and structured in recent years with the introduction of guidelines and codes of practice. There is a surge of efforts in establishing dispute resolution clauses in construction contracts to define and regulate mediation recently; but the actual mediation process is not strictly defined in any system (McGeorge et al., 2007).

In short, positive aspects of mediation can be listed as (1) low costs compared to conventional methods and DRB, (2) speedy resolutions compared to conventional methods and DRB, (3) informal nature, (4) creative solutions and win-win outcomes, (5) consensual and flexible agreements, (6) increased control of parties over the process, (7) enhanced communication between parties, (8) preserved business relationships and reputations, (9) decreased hostilities, and (10) confidentiality.

Considering all these advantages, the rapid growth in preference of mediation technique around the world can be understood (Cheung et al., 2004b). However, similar to other resolution methods, criticisms are also present. Firstly, the method is a voluntary and non-binding one that depends on willingness of participants to settle (McGeorge et al., 2007). Mediation can be successful only if parties are eagerly willing to settle (Jones, 2006). Secondly, parties can perceive compromising agreements as weakness and consequently, they are not eligible to find solutions through mediation. Therefore, the process can remain inconclusive that generates a waste of time and money. Finally, qualifications of the mediator is a key factor in successful mediation. An unqualified or incapable mediator may distort the process causing more problems (Peña-Mora et al., 2003). Despite these criticisms, mediation is still one of the first alternatives to conventional resolution methods (PMI, 2016). Organizations like the AAA encourage parties to mediate before taking their cases to arbitration or litigation (Harmon, 2003).

2.2.2.3. Senior Executive Appraisal (SEA)

According to National Alternative Dispute Resolution Advisory Council (NADRAC) of Australia, SEA is a form of case appraisal presentation where the facts of a case are presented to senior executives of organizations in dispute (NADRAC, 2011). This method has been defined under various names in the domain such as executive board appraisal, etc. Regardless of the name, the primary goal of this method is to include top-level management of disputed parties to resolution process.

The driving force in establishment of SEA method is problems arising from handling dispute negotiations by middle management. In such a case, senior executives make decisions without being fully informed as the accuracy, timeliness, and completeness of the information on the progress of negotiations they receive from middle management is questionable. Thus, this method involves senior executives in resolution process in order to avoid information related problems and to make informed negotiation decisions (Jones, 2006). In addition, resolution process can benefit from presence of unprejudiced and amicable senior executives with authority, instead of unauthorized, prejudiced, and tense middle management personnel (İlter, 2010a).

In SEA method, a panel of senior executives are established. In a typical panel, both parties select one (or more for larger panels) senior executive each, who is not involved in the resolution process of the dispute under review. These selected executives can invite an impartial expert as a third party if required. The impartial member is usually an experienced expert on technical and/or legal issues related with the reviewed dispute (İlter, 2010a). This panel reviews presentations of the parties' cases, documents, evidences, and testimonies at a desired level of formality ranging from hearings resembling mini-trials to informal sessions. Accordingly, the position of the impartial third party ranges from a judge-like role to a mediator-like role. Regardless of these, the impartial expert assists senior executives of both parties in reaching creative and agreeable solutions. This method enables early action in addition to speedy, economic and confidential resolution (Cheung and Yeung, 1998). Other advantages can said to be the decreased procedural formality and commercial pressure. However, the bindingness of decisions of the panel depends on agreements between parties. In addition, the method relies on willingness of parties to settle (Jones, 2006).

2.2.2.4. Negotiation

In general view, negotiation is a basic human activity that is performed in situations ranging from solving daily problems to international diplomacy (Kassab et al., 2006). In dispute management point of view, negotiation is a consensual process where parties willingly attempt to arrive a solution themselves (Bruce et al., 2004). It is a method encouraged by numerous law systems and organizations in which parties or their representatives conduct meetings upon dispute occurrence to search for solutions to their problems (İlter, 2010a). It is the most common method of ADR (OGC, 2002). Indeed, negotiation is usually the first attempt in dispute resolution (Marzouk et al., 2011). Settling through negotiation is a necessity considering the current state of the constantly shifting business environment in construction industry (Kassab et al., 2006).

Among all conventional and ADR methods, negotiation is the one with least procedural complexity. The actual process is not defined by any strict rules and parties come together directly or via representatives in an unstructured way (Arıcı, 2012). In a typical negotiation, there is no need for third party intervention. Thus, unlike mediation and DRBs, there is generally no third party to facilitate the negotiation and since there is no third party, participants can schedule and structure the course of negotiations according to their needs and requirements, which give them maximum level of control over the process (Bruce et al., 2004). This kind of negotiation is known as the direct negotiation, which helps parties in retaining their confidentiality and independence. Direct negotiation starts at the project level among project team and if they cannot reach a settlement, it can be terminated or continued at higher management levels (Mitropoulos and Howell, 2002). Upon continuation of negotiation in the next level of management, the process is called a stepped negotiation. In stepped negotiation, vertical and lateral management levels must be identified to guarantee the rapid escalation of issues in the right direction so that unresolved problems are escalated to the next level of management until a higher level resolves it (McGeorge et al., 2007). Ultimately, if parties cannot solve the dispute within themselves, they can acquire the assistance of a third party, in other words, a negotiator (Mitropoulos and Howell, 2002). This kind of negotiation is known as third party negotiation and its success depends on skills of the negotiator. An effective negotiator adopts required strategies and styles that suit best to the situation (Cheung et al., 2006). Similarly, the success of direct and stepped negotiations depend on skills of individuals involved in the process. Therefore, all construction professionals, especially the ones in the managerial positions, should have the necessary negotiation and communication skills (Cheung et al., 2006). With the increasing demand in these skills among construction professionals, negotiation has become a popular research area and decision-support systems to enhance the process are commercially available (Kassab et al., 2006). People lacking negotiation skills can benefit from such systems.

Negotiations for a predefined duration are often required before starting formal resolution methods (Bruce et al., 2004). Most standard forms of contracts involve negotiation as a first step in dispute resolution (McGeorge et al., 2007). In general, parties should mutually agree on resorting to negotiation as an initial step in order to start the process (Arıcı, 2012). In addition, it is a non-binding method unless parties establish a legally binding agreement at the end (İlter, 2010a). Therefore, the success of negotiation in settling disputes is relying on participants' willingness and motivations to settle (Jones, 2006). Besides the willingness in resolving the issues, negotiation requires participant openness to understand the standpoint of the other party through continuous communication. Therefore, if there are hostilities among parties, negotiation may not be the ideal resolution method, because such an adversarial environment will be a barrier in front of effective communication and understanding (McGeorge et al., 2007). Consequently, negotiation efforts in adversarial environments cause waste of time and this wasted time will cause additional financial loses while parties remain inconclusive (Marzouk et al., 2011). Thus, the relationship between parties is another critical factor in achieving successful negotiation (Mitropoulos and Howell, 2002). Unsuccessful negotiation will result in a business environment with increased hostilities and less chance of early settlement, which will lead to more expensive methods of resolution such as the conventional ones (Cheung et al., 2006). Another factor in achieving successful negotiation is the level of authority of the individuals. These individuals are less likely to make compromises, as they have to clarify their concessions to senior management (Jones, 2006). This lack of authority will threaten the success of the negotiation process. With involvement of authorized management personnel, chances of settlement will be higher. Finally, there can be individuals who perceive negotiation as a sign of weakness, which will cause an inconclusive process (Mackie et al., 2011).

In short, barriers in front of successful negotiation are (1) adversarial environment among negotiating parties, (2) lack of communication and negotiation skills of participants, (3) lack of authority of involved individuals, (4) being not motivated and willing to resolve the dispute, and (5) perception of negotiation as a weakness. If these barriers can be removed, negotiation will become an effective dispute resolution method, especially when failing to settle will have serious consequences to both parties. Moreover, if the process can be appropriately carried out, it will reduce costs of resolution and enhance the communication between parties. In addition, the current and long-term business relationships will be preserved (Jones, 2006).

There are several other advantages of negotiation process. To begin with, it is the most cost efficient method of dispute resolution in construction industry due to its informal, less complex, and fast nature (Cheung et al., 2006). In addition, as there are generally no third party, the process is free of additional costs associated with third party involvement. The OGC (2002) also states that negotiation is the most efficient form of ADR in terms of resolution cost and duration, while giving a range of possible solutions instead of being stuck to definite judgments. Thus, favorable outcomes can be achieved for all parties (Bruce et al., 2004). Finally, as mentioned earlier, the process can be controlled and regulated flexibly according to the needs of parties and issues under review.

In short, benefits of negotiation process can be listed as (1) lowest cost of resolution among all techniques, (2) speedy resolution, (3) flexibility, (4) degree of control over the process, (5) confidentiality, (6) wide range of possible solutions, (7) preserved business relationships, (8) enhanced communication and relationships, (9) informal nature, and (10) least procedural complexity among resolution methods.

All conventional and ADR methods used in projects in the dataset collected for this thesis study have been reviewed. Now, previous studies on dispute resolution method selection will be examined.

2.2.3. Previous Studies on Dispute Resolution Method Selection

The thesis study reviewed preventive actions, especially based on predicting dispute occurrence and potential compensations, so far. However, prevention techniques do not guarantee dispute avoidance completely and upon inevitable occurrence of disputes, these problems should be resolved (Cheung, 1999). Consequently, the construction industry has utilized several resolution techniques to solve these problems and therefore, the research continued with the introduction of these methods to the reader to highlight properties of each technique along with their advantages and disadvantages. Besides the level of knowledge on resolution methods, successful dispute resolution management also depends on the level of comprehension of factors affecting the dispute development. Although there are numerous studies on dispute resolution domain in the literature, there are only a few studies focusing on interrelations between disputes and various project characteristics based on empirical data (İlter and Dikbaş, 2009). Moreover, researchers argue that parties in a dispute fail to analyze possible gains and losses associated with each resolution method in a case-based approach (İlter, 2010b). Thus, it is claimed that the literature is insufficient to provide methods on how to systematically determine which dispute resolution strategy to adopt depending on the case characteristics (Cheung and Suen, 2002).

The dispute resolution method and strategy selection literature involves numerous efforts in order to decrease the number of disputes while avoiding inefficient resolution techniques, which are generally the court involved ones (Pulket and Arditi, 2009a). However, the systematic selection of the most appropriate resolution methodology is a difficult task because of its dependence on project and dispute characteristics, disputed parties' relationships with each other, and other factors that are not known prior to dispute occurrence (Harmon, 2003). Although such characteristics can significantly shape the resolution method selection, there are only few studies with such considerations. Thus, the remaining limited studies that identify mentioned characteristics are reviewed in this section to find out which of them influence resolution processes. This review will be used later in generation of the conceptual model that involves a pool of input variables comprised of project characteristics affecting dispute development and resolution strategies. Data related to these input variables will be collected via questionnaires in order to develop the resolution method selection model proposed in this thesis study.

Before reviewing previous studies on resolution method selection, some common studies should be highlighted. Some of the studies mentioned earlier in Section 2.1.2 of this thesis study focused on relationships between various factors, characteristics, and attributes with dispute prediction and resolution strategies simultaneously (e.g., studies aiming to predict the litigation outcomes and understand the mechanisms of court rulings, etc.). Therefore, findings of these studies are reviewed for both dispute prediction and resolution method selection models (Dalton and Shehadeh, 2003; Watts and Scrivener, 1993; Mitropoulos and Howell, 2002; Chau, 2007;Cheng et al., 2009; Arditi and Pulket, 2009; Pulket and Arditi, 2009a; Pulket and Arditi, 2009b; Marzouk et al., 2011). Besides these studies that generate common variables for both dispute prediction (occurrence and potential compensation) and resolution method selection models, there are various studies focusing on factors affecting resolution method selection only.

One of the first studies focusing on critical factors related to dispute resolution strategies is conducted by York (1996) and 17 factors are associated with strategic decision-making. Cheung (1999) identified 12 critical ADR attributes influencing resolution processes, which are bindingness, economy, confidentiality, control over process, creative solutions, enforceability, fairness, flexibility, privacy, speed, relationships, and width of remedy. Cheung and Suen (2002) reviewed the literature to collect 16 critical dispute resolution method selection criteria. In Cheung et al.'s (2002) literature review, 19 ADR attributes that are important in determination of the resolution strategy are identified. In Harmon (2003), various factors (i.e. magnitude of work, future work together, etc.) are associated with resolution methods and strategies. İlter and Dikbaş (2008) reviewed the literature to come up with 32 key ADR attributes, while İlter (2010a) used 56 criteria for a decision-support system to enhance resolution method selection. In addition, İlter (2010b) interviewed legal professionals in Turkish construction industry to highlight 16 factors they consider during resolution method selection. In a CBR-based decision-support system that retrieves similar dispute cases and presents to decision-makers, 11 project attributes, such as contract sum, type of contract, involvement of a claim consultant, etc., are utilized in similarity measurements (Cheung et al., 2004a). Moreover, there are studies using text-mining techniques to retrieve similar dispute cases that can be used as examples showing what can happen upon utilization of a specific resolution technique. These studies involve numerous attributes to perform text-mining operations (Yılmaz and Dikbaş, 2013; Fan and Li, 2013). Kassab et al. (2010) developed a decision-support system using the graph model to suggest resolution strategies and the system makes decisions based on attributes such as project delays, cost increase, contractor reputation, and presence of continuing projects. In a series of studies aiming to forecast dispute resolutions, 15 project and dispute characteristics have been associated with the method selection (Chou, 2012; Chou et al., 2013b). Mahfouz and Kandil (2011) reviewed links between 15 project related legal factors and litigation rulings. Çevikbaş and Köksal (2018) reviewed litigious cases from Turkish construction industry and investigated judgments on these cases by focusing on characteristics namely dispute types, project types, scope of construction, and specific articles in the contract. Cheung et al. (2010) claimed that with satisfactory dispute resolution, the industry would be less adversarial; thus, by focusing on factors affecting resolution method satisfaction, their study indicated that 14 variables under four categories are determinants of the resolution method satisfaction.

All these mentioned factors, characteristics, and attributes are reviewed and similar items are merged. Resulting items are used in order to establish a conceptual model that is composed of a pool of input variables, which will be used for the design of a questionnaire to collect empirical data and the development of prediction models proposed in this thesis study.

2.2.4. Literature Review on Decision-Support Systems for Resolution Method Selection

An appropriate resolution process can lead the project to success and as a result, selection of the resolution strategy is crucial (Cheung et al., 2004a). Although dispute resolution methodology is usually chosen by contract documents before occurrence of a dispute, considering the nature of the dispute and factors such as the relationship between disputed parties, the methodology that best suits the needs of the participants should be selected (Harmon, 2003). The decision-making approach should involve consideration of various interrelated factors such as technical, financial, social, contractual, etc. However, there is a lack of an approach of this nature and current processes depend on experience and qualitative assessments that cause subjectivity problem (Cheung and Suen, 2002). While performing such a selection, experience and knowledge are invaluable. The merits of AI techniques include extraction of such tacit knowledge in an articulable and presentable way to decision-makers resulting an informed decision-making process that is free of subjectivity during resolution method selection. Thus, developing dispute databases based on AI applications should be a primary goal in the domain (İlter and Dikbaş, 2009). Moreover, development of AI based models can also enable early warning of potential dispute resolutions (Chou et al., 2014). Therefore, in this section, decision-support systems, especially the ones based on AI techniques, will be reviewed.

There is a tendency in the resolution method related research to focus on the litigation technique specifically, because if parties can know the decision of the court beforehand with some certainty, they might be more likely to settle out of the court rather than facing undesired outcomes with serious financial consequences and damaged business relationships (Pulket and Arditi, 2009a). Therefore, there are various efforts to predict outcomes of litigation processes and court rulings. Chau (2007) developed a combined model using Particle Swarm Optimization (PSO) based ANN to predict the outcome of construction claims if they were taken to courts for projects in Hong Kong. By using 13 attributes for classification, 550 litigious cases covering the period of 1991-1995 are used to train the ANN model, 275 cases covering the period of 1996-1997 are used to test the trained ANN model, and finally, 280 cases covering the period of 1998-2000 are used to validate the model. The validated model resulted in a prediction accuracy of 80.00%. Cheng et al. (2009) enhanced the CBR technique by integrating with fuzzy set theory based on a new similarity measurement approach that combines Euclidean distance and cosine angle distance. Then, in order to retrieve similar litigated cases via CBR model, a dataset involving 153 dispute cases is analyzed based on eight attributes as similarity measures. El-adaway and Kandil (2009) established a multi-agent system for resolution of change order related disputes. The proposed system simulates legal proceedings and provides similar past cases including their supporting and counter arguments based on an algorithm that derives logic rules considering 12 factors related to change order disputes. In a series of studies to predict court rulings and litigation outcomes, various AI techniques are applied on a dataset composed of cases from Illinois Appellate Court. In these studies, the litigation outcome is predicted by an ANN model that achieved an accuracy rate of 66.67% (Arditi et al., 1998), by CBR with 83.33% (Arditi and Tokdemir, 1999), by boosted decision trees (BDT) with 89.59% (Arditi and Pulket, 2005), by their integrated prediction model (IPM) with 91.15% (Arditi and Pulket, 2009), and finally, by their universal prediction model (UPM) with 96.02% accuracy rate (Pulket and Arditi, 2009b). In a more recent study, litigation outcomes of differing site condition related disputes are predicted using ML techniques that involve polynomial SVM, Naïve Bayes, J48 decision trees, ADTree BDT, and Projective Adaptive Resonance Theory (PART). All algorithms achieved significant classification accuracy success (minimum 93.00%) along with other successful measures such as precision, recall, etc. (Mahfouz and Kandil, 2011). Although knowing litigation outcomes with some certainty can avoid parties from resorting to courts, these studies do not suggest which resolution method to use as an alternative. Thus, the need for a systematical approach that will suggest the most appropriate resolution method is not addressed through these studies.

There are several other efforts aiming to develop decision-support systems for resolution method selection besides studies predicting the litigation outcome and forecasting court rulings. Cheung (1999) evaluated critical factors affecting the use of ADR techniques using scale rating, percentage rating, and Factor Analysis methods to find attribute importance ratings with the aim of discovering the concerns of decision-makers during their resolution method selection. In a similar study, all attributes related to ADR processes are identified from the literature, a hierarchical model of ADR processes are developed using AHP method, and critical attributes are determined through prioritization (Cheung et al., 2002). Another decision-making model combining AHP and Multi-Attribute Utility Technique (MAUT) is developed with the same purpose (Cheung and Suen, 2002). Cheung et al. (2004a) proposed a CBR based model to select a resolution method by resorting to previous similar cases. Gebken and Gibson (2006) investigated impacts of cost of resolution methods, disputing party, and dispute complexity on resolution method selection via ANOVA analysis. Based on case studies, a systematical approach for conflict resolution utilizing graph model theory, named graph model for conflict resolution (GMCR II), is developed that considers possible decision-makers, decision options, feasible actions, and outcomes at a dispute negotiation (Kassab et al., 2006). Abilities of GMCR II are further improved in a new decision-support system, which is capable of performing uncertainty analysis when preferences of decision-makers are not certain. In addition, the new system can suggest possible solutions (Kassab et al., 2010). However, both tools are developed to enhance the decision-making process during negotiations and the scope does not cover entire options of resolution methods. In another study, based on results of a survey on decision-makers to highlight factors affecting their decision-making rationale, a dispute resolution method selection tool is developed using AHP and rule based algorithm for construction disputes in Egypt

(Marzouk et al., 2011). A more sophisticated decision-making model for resolution method selection is established using multi-criteria decision-making method (MCDM). Through validation of the model with industry professionals, it is observed that suggestions resulting from this model is highly compatible with decisions of the professionals (İlter, 2010a). MCDM is an appropriate method for decision-making especially when there are conflicting criteria. However, MCDM models require criteria weights and rankings to be externally determined by the developer. On the other hand, AI applications, specifically ML techniques, are easier to establish, as they do not require significant preprocesses. In addition, ML algorithms can model relationships of numerous variables and their impacts on the resolution method selection objectively with a comparable performance to MCDM method.

Considering benefits of ML techniques, studies utilizing these algorithms can be found recently among efforts of developing decision-support systems for resolution method selection. The effectiveness of ML algorithms has already been observed in studies aiming to predict the dispute occurrence and to forecast litigation outcomes.

A previously analyzed dataset of 584 PPP projects undertaken by the TPCC for dispute prediction is analyzed again to forecast resolution method selection. For this reason, disputed projects among the dataset (152 disputed cases) are extracted and a multiclass classification is performed on these cases to forecast the resolution method in both the project initiation and the dispute occurred phases depending on 15 project and dispute related characteristics. Various single algorithms, which involve SVM, ANN, TAN, CART, QUEST, Exhaustive CHAID, and C5.0 algorithms, are trained and tested. The best test set accuracy of 83.82% is achieved through Exhaustive CHAID algorithm for project initiation phase, while 69.05% accuracy is achieved through CART algorithm for dispute occurred phase. In addition, several ensemble models are also experimented. The best ensemble model was the triple combination of QUEST, Exhaustive CHAID, and C5.0 algorithms with an accuracy of 84.65% on the test set for project initiation phase. For dispute occurred phase, the best test set accuracy was obtained as 69.05% from three ensemble models, which are CART and Exhaustive

CHAID, SVM and CART, and the triple combination of SVM, CART, and Exhaustive CHAID (Chou, 2012). In another study, the performance of single SVM is enhanced by integrating with fast and messy genetic algorithm (fmGA) and then, enhancing the integrated algorithm even more by combining it with fuzzy logic. The average test set classification accuracy for resolution method forecasts is 61.75% with single SVM, 69.21% with fmGA based SVM, and 77.04% with the combination of SVM, fmGA, and fuzzy logic (Chou et al., 2013b). In a final attempt for selecting the most appropriate resolution method, the same dataset is experimented with the ANN, SVM, Naïve Bayes, CART, CHAID, QUEST, and C5.0 algorithms. The best average 10-fold cross-validation accuracy result is obtained from the SVM classifier with 81.12% (Chou et al., 2016). Although all three of these studies are specific to an industry and a project type, they are still valuable as the proposed classification approaches can enable early warning of potential dispute resolutions.

Cheung et al. (2010) claimed that resolution method selection should be based on the satisfaction with the method. Therefore, their study predicted the dispute resolution satisfaction on a dataset of 48 construction projects from Hong Kong using LR technique and results are compared with a previous Multi Discriminant Analysis (MDA) model. In another effort to classify past projects with respect to their dispute resolution satisfaction, an MLP model is developed. With this model, it is possible to distinguish adverse and favorable ADR methods (Cheung et al., 2002).

Finally, there are studies using text-mining techniques to retrieve similar dispute cases that can be used as examples showing what can happen upon utilization of a specific resolution technique. Among these studies, Yılmaz and Dikbaş (2013) compared classification performances of kNN, Naïve Bayes, SVM, and decision tree algorithms on a dataset of 49 documents received from Directorate of High Technics Board of Turkish Ministry of Environment and Urbanization. In another text mining effort in construction industry, Fan and Li (2013) tried to retrieve similar cases of ADR use in construction accidents using text-mining techniques. However, these text-mining based studies aim to call similar cases upon request to be used as examples and leave the decision-making process to the decision-maker. Instead of a text-mining based approach, a more enhanced decision-making process considering various project and dispute related characteristics would be more beneficial for decision-makers.

Respecting the state of the mentioned existing research on decision-support systems for dispute resolution method selection in construction industry, it can be observed that studies have limitations such as being industry, project type, dispute type, and contracting strategy specific. Global-scaled models that consider various projects, dispute types, contract types, and judicial systems during resolution method selection do not exist in the literature. This problem is similar to the problem that was previously observed in dispute prediction research. Development of a global-scaled model that considers various resolution methods along with considerations on variations in characteristics related to dispute, project, judicial system, etc. would be a better approach. For this reason, firstly, the dataset that will be used in such decision-support systems should involve projects reflecting these variations. Secondly, these systems should be based on ML techniques as these techniques can said to be containing the most appropriate approaches for such systems considering their superiority in modeling problems with complex and interrelated attributes. Finally, in order to achieve better classification performance with ML techniques, single and ensemble algorithms should be experimented on the dataset with respect to measures such as accuracy, TP rates, etc. Therefore, this thesis study aims to fill the mentioned gaps of the research by developing an ML based decision-support model for dispute resolution method selection that is using construction project data from various construction industries. Moreover, projects in the dataset will be diverse in terms of project types, contracting strategies, dispute types, etc. Finally, not only performances of several single algorithms will be experimented using various significant performance measures in the ML domain, but also various ensemble models combining these single classifiers will be experimented.

2.3. CONCEPTUAL MODEL DEVELOPMENT

Various research studies on construction conflicts, claims, disputes, and resolution methods are reviewed along with their contributions, strengths and weaknesses so far.

In this section, findings from these reviews will be assessed in order to develop a conceptual model with the purpose of identifying and determining input and output variables (attributes) for prediction models. The efforts will start with determination of output variables. The number of prediction models that should be established is equal to the number of output variables. Therefore, studies are grouped with respect to their output variables or contribution potentials to an output variable. Then, according to these output variables, predictors will be identified as input variables.

2.3.1. Determination of Output Variables for Prediction Models

Previous studies from construction dispute domain are grouped with respect to their output variables or contribution potentials to an output variable (Table 2.1). These studies are summarized in Table 2.2, where they are organized in chronological order. There are short remarks in the table explaining each study briefly. According to this grouping, it is observed that the literature mainly focuses directly or indirectly on three outputs: (1) Dispute Occurrence (Dispute Likelihood), (2) Potential Compensation (categorical or quantitative), and (3) Resolution Method.

Table 2.1. Number of Publications per Output

Output Variable	Output ID	Identified Publication Number
Dispute Occurrence	O1	15
Compensation	O2	19
Resolution Method	O3	24

Table 2.2. Literature Review According to Output Variable(s)

Year	Author(s)	Remarks	Output(s)
1993	Revay	Can construction claims be avoided? Identification of frequently reoccurring reasons for claims and investigation of ways to avoid those considering their financial consequences	Dispute Occurrence & Compensation
1995	Diekmann & Girard	Are some construction projects more prone to contract disputes than others? If so, can these projects be identified before construction begins?	Dispute Occurrence
1997	Fenn, Lowe & Speck	Do certain contract types cause more disputes than others?	Dispute Occurrence
1998	Arditi, Oksay & Tokdemir	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings using neural networks	Compensation & Resolution Method Selection
1999	Arditi & Tokdemir	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings using CBR	Compensation & Resolution Method Selection
1999	Cheung	Identification of critical factors that affect resolution method selection in construction disputes	Resolution Method Selection
2000	Molenaar, Washington & Diekmann	Are some construction projects more prone to contract disputes than others? If so, can these projects be identified before construction begins?	Dispute Occurrence
2002	Cheung & Suen	A decision-making model that combines AHP and MAUT for dispute resolution method selection	Resolution Method Selection

Table 2.2. Literature Review According to Output Variable(s) (Continued)

Year	Author(s)	Remarks	Output(s)
2002	Cheung, Suen & Lam	Identification of critical factors that affect selection of ADR techniques in construction dispute resolution	Resolution Method Selection
2002	Cheung, Tam & Harris	Analysis of factors affecting the outcome of resolution processes in Hong Kong using ANN method so that the decision-maker can decide on the method to use	Resolution Method Selection
2002	Mitropoulos & Howell	Development of a process model that classifies potential problematic situations and identifies resolution requirements by analyzing potential for dispute occurrence, possible compensations, and factors affecting dispute resolution	Dispute Occurrence & Compensation & Resolution Method Selection
2003	Dalton & Shehadeh	Knowing various factors for a project and using statistical models, it is possible to predict the number and value of claims.	Dispute Occurrence & Compensation
2004a	Cheung, Au-yeung & Wong	Establishment of a CBR based model for resolution method selection	Resolution Method Selection
2005	Arditi & Pulket	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings using boosted decision trees	Compensation & Resolution Method Selection
2005	Kilian & Gibson	Identification of primary causes of litigation associated with the U.S. Naval Facilities construction contracts and analysis of root causes of disputes with the aim of avoiding occurrence	Dispute Occurrence

Table 2.2. Literature Review According to Output Variable(s) (Continued)

Year	Author(s)	Remarks	Output(s)
2006	Gebken & Gibson	Quantification of costs for dispute resolution procedures in the construction industry with the aim of achieving quantitative comparisons of various resolution methods during selection	Resolution Method Selection
2006	Kassab, Hipel & Hegazy	Development of a decision support tool based on graph model that can predict the outcomes of negotiations between parties	Compensation
2007	Chau	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings	Compensation
2007	Fenn	Are disputes predictable?	Dispute Occurrence
2007	Chen & Hsu	Constructing an early warning model to prevent potential litigation due to project changes by using a hybrid ANN and CBR model (The ANN model predicts the litigation likelihood and the CBR model presents similar litigious cases)	Dispute Occurrence & Compensation
2008	Chen	Constructing a kNN based model to identify the litigation likelihood of change order disputes	Dispute Occurrence
2009	Arditi & Pulket	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings using an integrated AI model	Compensation & Resolution Method Selection
2009	Cheng, Tsai & Chiu	Establishment of a fuzzy CBR model that identifies similar construction disputes resolved by litigation to present the possible outcomes to decision-makers	Compensation

Table 2.2. Literature Review According to Output Variable(s) (Continued)

Year	Author(s)	Remarks	Output(s)
2009	El-adaway & Kandil	Development of a system that presents the user similarities, differences, strengths, and weaknesses of the disputed case and previous arbitrated cases with the aim to support defense preparations for acquiring possible compensations	Compensation
2009	Pulket & Arditi	Prediction of outcomes of construction litigation according to the characteristics of the individual dispute and the corresponding past court rulings using ant colony optimization	Compensation & Resolution Method Selection
2009	Pulket & Arditi	Prediction of outcomes of construction litigation according to characteristics of the individual dispute and corresponding past court rulings using ML based universal prediction model	Compensation & Resolution Method Selection
2010	Cheung, Yiu & Chan	Identification of favorable resolution methods based on logistic regression analysis by predicting the potential satisfaction with the resolution method	Resolution Method Selection
2010a	İlter	Establishment of a multi-criteria decision support model using MAUT in selecting the most appropriate dispute resolution method at various construction stages	Resolution Method Selection
2010	Kassab, Hegazy & Hipel	Development of a computerized decision support system to forecast possible compensation(s) that can be acquired out of the negotiation process	Compensation
2011	Mahfouz & Kandil	Prediction of outcomes of construction litigation of differing site condition disputes according to the characteristics of the individual dispute and the corresponding past court rulings using ML algorithms	Compensation & Resolution Method Selection

Table 2.2. Literature Review According to Output Variable(s) (Continued)

Year	Author(s)	Remarks	Output(s)
2011	Marzouk, El-Mesteckawi & El-Said	Development of a computerized model to support the decision-making process during resolution method selection for construction projects in Egypt	Resolution Method Selection
2012	Arıcı	Identification of factors affecting resolution method selection based on empirical data with special emphasis on ADR techniques	Resolution Method Selection
2012	Chou & Lin	Proactively forecasting dispute occurrence in the initiation phase of PPP projects using ML techniques on empirical project data	Dispute Occurrence
2012	Chou	Early forecasting of potential dispute resolutions using ML techniques for PPP projects	Resolution Method Selection
2012	İlter	Identification of impacts of various dispute factors through empirical analysis of the associations between dispute factors and categories (may highlight possible compensations)	Compensation
2013	Cheung & Pang	Evaluation of dispute occurrence likelihood of construction projects by identifying factors that contribute to disputes based on an anatomy of disputes	Dispute Occurrence
2013	Chou, Tsai & Lu	Proactively forecasting dispute occurrence of PPP projects using ML techniques on empirical project data	Dispute Occurrence
2013	Chou, Cheng & Wu	Early forecasting of potential dispute resolutions using ML techniques for PPP projects	Resolution Method Selection

Table 2.2. Literature Review According to Output Variable(s) (Continued)

Year	Author(s)	Remarks	Output(s)
2013	Fan & Li	Development of a text mining approach that presents the user similar past cases of construction accidents (can be used to understand compensations and potential resolutions)	Compensation & Resolution Method Selection
2013	Yılmaz & Dikbaş	Development of a text mining approach that presents the user similar past cases from Directorate of High Technics Board of Turkey database with the aim to support defense preparations for acquiring possible compensations and to decide whether to resort to the board or not	Compensation & Resolution Method Selection
2014	Chou, Cheng, Wu & Pham	Proactively forecasting dispute occurrence of PPP projects using ML techniques on empirical project data	Dispute Occurrence
2016	Chou, Hsu, Li & Chang	Proactively forecasting dispute occurrence of PPP projects, Identification of type of dispute, Identification of phase of dispute occurrence, Early forecasting of potential resolutions	Dispute Occurrence & Resolution Method Selection
2016	Yousefi, Yakhchali, Khanzadi, Mehrabanfar& Saparauskas	Proposal of an NN model that predicts time and cost claims in construction projects	Compensation

In short, the literature can be categorized under three output variables (O1, O2, and O3) that will lead the research to development of three distinct prediction models namely: (1) Dispute Occurrence Prediction Model, (2) Potential Compensation Prediction Model, and (3) Resolution Method Selection Model. For this purpose, predictors (input variables) related to each output should be identified from the literature as well. The next section will present findings of these efforts.

2.3.2. Determination of Input Variables for Prediction Models

Reviewed studies from the literature pointed out that numerous factors have been associated with dispute occurrence, potential compensations, and resolution method selection. However, level of inclusion and detailing of these factors are different in these studies. In addition, there is a confusion in the related terminology due to overlapping concepts and the distinction between causes, factors, types of disputes, etc. may not be very clear (İlter, 2012). In order to tackle these problems, this research identified its own set of input variables related to defined outputs, which are 'O1-Dispute Occurrence', 'O2 – Potential Compensation', and 'O3 – Resolution Method'.

The findings of the literature survey are assessed further to identify input variables that impact the mentioned outputs. Firstly, variables are grouped under different categories. Variables related to project and contract related characteristics of a construction project are categorized as 'Project Characteristics'. As the details will be given in following pages, 'Project Characteristics' involve variables such as project location, contract value, planned project duration, etc. Variables depending on parties involved in the project and their organizational characteristics are categorized as 'Skills'. This category involves variables such as working culture, communication skills, project management and coordination skills, etc. The third category is the 'Changes' that involves occurrence of variations, changes, or unexpected events in a construction project and the fourth category is the 'Delay' category. Variables related to characteristics of the dispute are categorized as 'Dispute Characteristics'. 'Dispute Characteristics' involve variables such as disputant party, financially disputed amount, presence of EoT claim, etc. Common causes of disputes are collected under the 'Dispute Sources' sub-category. Finally, variables of resolution strategy related characteristics are categorized as 'Resolution Method Characteristics'. This category is composed of variables related to expectations from the resolution method (i.e.

importance of preserving relationships between parties) and consequences of the selected method (i.e. cost of resolution). The level of knowledge of the decision-maker about specific resolution methods are also collected as variables under the category of 'Level of Knowledge on Resolution Method'. Following the categorization, input variables for each category are identified. In the literature, same or similar factors can be observed under different names. Therefore, while assigning variables to each category, attributes with similar names and descriptions are merged.

2.3.2.1. 'Project Characteristics' Attributes

The first category of input variables is the 'Project Characteristics' that involves attributes related to project and contract related characteristics of a construction project. There are 11 attributes in this category according to the findings of the literature survey. These attributes are;

- PC1 – Project Location
- PC2 – Project or Contract Value
- PC3 – Planned Project Duration
- PC4 – Type of Construction
- PC5 – Type of Contractor (i.e. joint venture, consortium)
- PC6 – Type of Employer (i.e. public, private, PPP)
- PC7 – Type of Contract
- PC8 – Payment Method (i.e. unit price, fixed price)
- PC9 – Project Delivery System (i.e. DBB, DB)
- PC10 – Level of Design Complexity
- PC11 – Level of Construction Complexity

Table 2.3 involves 'Project Characteristics' attributes, studies highlighting the importance of these attributes in the literature and the number of publications mentioning them. Studies are given in chronological order.

Table 2.3. 'Project Characteristics' Attributes

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
PC1	Project Location	1993	Revay	2013a	Chou et al.	9
		1995	Diekmann & Girard	2013b	Chou et al.	
		2000	Molenaar et al.	2014	Chou et al.	
		2012	Chou	2016	Chou et al.	
		2012	Chou & Lin			
PC2	Project or Contract Value	1993	Revay	2009	Cheng et al.	26
		1995	Diekmann & Girard	2009	İlter & Dikbaş	
		1998	Arditi et al.	2010a	İlter	
		1999	Arditi & Tokdemir	2010b	İlter	
		2000	Molenaar et al.	2010	Love et al.	
		2003	Harmon	2012	İlter	
		2003	Yates	2012	Chou	
		2004a	Cheung et al.	2012	Chou & Lin	
		2005	Arditi & Pulket	2013a	Chou et al.	
		2005	Kilian & Gibson	2013b	Chou et al.	
		2007	Chau	2014	Chou et al.	
		2007	Chen & Hsu	2016	Chou et al.	
		2008	Chen	2018	Çevikbaş & Köksal	
PC3	Planned Project Duration	2005	Kilian & Gibson	2012	Chou	14
		2007	Chen & Hsu	2012	Chou & Lin	
		2008	Chen	2013a	Chou et al.	
		2009	İlter & Dikbaş	2013b	Chou et al.	
		2010a	İlter	2014	Chou et al.	
		2011	Marzouk et al.	2016	Chou et al.	
		2012	İlter	2016	Yousefi et al.	
PC4	Type of Construct.	1993	Revay	2010a	İlter	25
		1995	Diekmann & Girard	2010	Love et al.	
		2000	Molenaar et al.	2011	Mahfouz & Kandil	
		2005	Arditi & Pulket	2011	Marzouk et al.	
		2005	Kilian & Gibson	2012	İlter	
		2006	Acharya et al.	2012	Chou	
		2007	Chen & Hsu	2012	Chou & Lin	
		2008	Chen	2013a	Chou et al.	
		2009	Arditi & Pulket	2013b	Chou et al.	
		2009	Cheng et al.	2014	Chou et al.	
		2009	İlter & Dikbaş	2016	Chou et al.	
		2009a	Pulket & Arditi	2018	Çevikbaş & Köksal	
		2009b	Pulket & Arditi			
PC5	Type of Contractor	1993	Watts & Scrivener	2007	Chau	19
		1995	Diekmann & Girard	2007	Chen & Hsu	
		1997	Kumaraswamy	2008	Chen	
		1998	Arditi et al.	2009	Arditi & Pulket	
		1999	Arditi & Tokdemir	2009	İlter & Dikbaş	
		2000	Molenaar et al.	2009a	Pulket & Arditi	
		2003	Harmon	2009b	Pulket & Arditi	
		2005	Arditi & Pulket	2010a	İlter	
		2005	Gencer	2012	İlter	
		2005	Kilian & Gibson			

Table 2.3. 'Project Characteristics' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
PC6	Type of Employer	1993	Watts & Scrivener	2008	Chen	26
		1995	Diekmann & Girard	2009	Arditi & Pulket	
		1997	Kumaraswamy	2009	İlter & Dikbaş	
		1998	Arditi et al.	2009a	Pulket & Arditi	
		1999	Arditi & Tokdemir	2009b	Pulket & Arditi	
		2000	Molenaar et al.	2010a	İlter	
		2003	Dalton & Shehadeh	2012	Chou	
		2003	Harmon	2012	Chou & Lin	
		2005	Arditi & Pulket	2012	İlter	
		2005	Gencer	2013a	Chou et al.	
		2005	Kilian & Gibson	2013b	Chou et al.	
		2007	Chau	2014	Chou et al.	
		2007	Chen & Hsu	2016	Chou et al.	
PC7	Type of Contractor	1997	Fenn et al.	2009	İlter & Dikbaş	26
		1997	Kumaraswamy	2009a	Pulket & Arditi	
		2002	Mitropoulos & Howell	2009b	Pulket & Arditi	
		2003	Dalton & Shehadeh	2011	Mahfouz & Kandil	
		2003	Harmon	2011	Marzouk et al.	
		2003	Yates	2012	Chou	
		2004a	Cheung et al.	2012	Chou & Lin	
		2005	Kilian & Gibson	2013a	Chou et al.	
		2007	Chau	2013b	Chou et al.	
		2007	Chen & Hsu	2014	Chou et al.	
		2008	Chen	2016	Chou et al.	
		2009	Arditi & Pulket	2016	Yousefi et al.	
		2009	Cheng et al.	2018	Çevikbaş & Köksal	
PC8	Payment Method of Contract	1995	Diekmann & Girard	2007	Chau	12
		1997	Fenn et al.	2009	Cheng et al.	
		1997	Kumaraswamy	2009	Arditi & Pulket	
		2000	Molenaar et al.	2009a	Pulket & Arditi	
		2003	Revay	2009b	Pulket & Arditi	
		2005	Kilian & Gibson	2010a	İlter	
PC9	Project Delivery System	1995	Diekmann & Girard	2012	Chou	13
		2000	Molenaar et al.	2012	Chou & Lin	
		2005	Kilian & Gibson	2013a	Chou et al.	
		2006	Acharya et al.	2013b	Chou et al.	
		2007	Chau	2014	Chou et al.	
		2009	İlter & Dikbaş	2016	Chou et al.	
		2010a	İlter			
PC10	Level of Design Complex.	1995	Diekmann & Girard	2003	Harmon	7
		2000	Cheung et al.	2003	Yates	
		2000	Molenaar et al.	2011	Marzouk et al.	
		2002	Cheung et al.			
PC11	Level of Construct. Complex.	1995	Diekmann & Girard	2003	Harmon	7
		2000	Cheung et al.	2003	Yates	
		2000	Molenaar et al.	2011	Marzouk et al.	
		2002	Cheung et al.			

2.3.2.2. 'Skills' Attributes

'Skills' category involves attributes depending on parties involved in the construction project and their organizational characteristics. There are eight attributes in this category according to the findings of the literature survey. These attributes are;

- S1 – Relationship between Parties / Individuals
- S2 – Previous Experience with Each Other or Reputation (Credibility)
- S3 – Dispute Avoidance Incentives
- S4 – Communication between Parties
- S5 – Working Culture & Skills of Parties
- S6 – Response Rate & Communication Skills of Parties
- S7 – Experience of Parties (with the type of project)
- S8 – Project Management & Coordination Skills of Parties

Table 2.4 involves 'Skills' attributes, studies highlighting the importance of these attributes in the literature and the number of publications mentioning them.

Table 2.4. 'Skills' Attributes

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
S1	Relation. btw. Parties / Individuals	1995	Bristow&Vasilopoulos	2006	Acharya et al.	22
		1995	Diekmann & Girard	2008	İlter & Dikbaş	
		1997	Fenn et al.	2010	Cheung et al.	
		1997	Kumaraswamy	2010a	İlter	
		2000	Cheung et al.	2010b	İlter	
		2000	Molenaar et al.	2010	Kassab et al.	
		2002	Cheung et al.	2010	Love et al.	
		2002	Mitropoulos & Howell	2011	Marzouk et al.	
		2003	Dalton & Shehadeh	2012	İlter	
		2003	Harmon	2013	Cheung & Pang	
		2003	Yates	2016	Yousefi et al.	
S2	Prev. Exp. or Reputation	1995	Diekmann & Girard	2002	Mitropoulos & Howell	7
		2000	Cheung et al.	2008	İlter & Dikbaş	
		2000	Molenaar et al.	2010	Cheung et al.	
		2002	Cheung et al.			
S3	Dispute Avoidance Incentives	1995	Diekmann & Girard	2004a	Cheung et al.	7
		2000	Molenaar et al.	2010	Cheung et al.	
		2002	Cheung et al.	2016	Yousefi et al.	
		2003	Yates			

Table 2.4. 'Skills' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
S4	Commun. btw Parties	1994	Rhys-Jones	2006	Acharya et al.	21
		1995	Bristow&Vasilopoulos	2008	İlter & Dikbaş	
		1995	Diekmann & Girard	2008	Younis et al.	
		1997	Fenn et al.	2010	Cheung et al.	
		1997	Kumaraswamy	2010a	İlter	
		2000	Molenaar et al.	2010b	İlter	
		2002	Cheung et al.	2010	Love et al.	
		2002	Mitropoulos & Howell	2012	İlter	
		2003	Harmon	2013	Cheung & Pang	
		2003	Yates	2016	Yousefi et al.	
		2005	Kilian & Gibson			
S5	Working Culture & Skills of Parties	1994	Rhys-Jones	2003	Yates	15
		1995	Diekmann & Girard	2008	İlter & Dikbaş	
		1997	Fenn et al.	2008	Younis et al.	
		1997	Kumaraswamy	2010	Love et al.	
		2000	Molenaar et al.	2011	Marzouk et al.	
		2002	Mitropoulos & Howell	2013	Cheung & Pang	
		2003	Dalton & Shehadeh	2016	Yousefi et al.	
		2003	Harmon			
S6	Response Rate & Commun. Skills of Parties	1995	Diekmann & Girard	2002	Mitropoulos & Howell	12
		1997	Fenn et al.	2003	Yates	
		1997	Kumaraswamy	2008	Younis et al.	
		1998	Cheung & Yeung	2010	Cheung et al.	
		2000	Molenaar et al.	2012	İlter	
		2002	Cheung et al.	2016	Yousefi et al.	
S7	Experience of Parties	1995	Diekmann & Girard	2008	Younis et al.	18
		2000	Cheung et al.	2009	İlter & Dikbaş	
		2000	Molenaar et al.	2010	Cheung et al.	
		2002	Cheung et al.	2010a	İlter	
		2003	Dalton & Shehadeh	2010	Love et al.	
		2003	Rubin & Quintas	2011	Love et al.	
		2006	Acharya et al.	2011	Marzouk et al.	
		2007	McGeorge et al.	2012	İlter	
		2008	İlter & Dikbaş	2016	Yousefi et al.	
S8	Project Manage. & Coord. Skills of Parties	1995	Diekmann & Girard	2005	Arditi & Pulket	15
		1997	Fenn et al.	2008	Younis et al.	
		1997	Kumaraswamy	2009	Arditi & Pulket	
		1998	Arditi et al.	2009a	Pulket & Arditi	
		1999	Arditi & Tokdemir	2009b	Pulket & Arditi	
		2000	Molenaar et al.	2010	Cheung et al.	
		2003	Dalton & Shehadeh	2016	Yousefi et al.	
		2003	Yates			

2.3.2.3. 'Changes'

The third category of input variables is the 'Changes' category that involves occurrence of variations, changes, or unexpected events in a construction project. Table 2.5 involves studies mentioning the impact of changes on construction disputes and resolution strategies as well as highlighting the importance of these changes in the literature with the total number of publications mentioning them.

Table 2.5. Studies in the Literature Highlighting the Importance of Changes

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
C1	Changes	1993	Revay	2006	Arditi & Pattanakitchamroon	37
		1993	Watts & Scrivener	2007	Chau	
		1994	Heath et al.	2007	Chen & Hsu	
		1995	Bristow&Vasilopoulos	2008	Chen	
		1995	Diekmann & Girard	2008	Younis et al.	
		1997	Fenn et al.	2009	Arditi & Pulket	
		1997	Kumaraswamy	2009	Cheng et al.	
		1998	Arditi et al.	2009	El-adaway & Kandil	
		1999	Arditi & Tokdemir	2009a	Pulket & Arditi	
		2000	Cheung et al.	2009b	Pulket & Arditi	
		2000	Molenaar et al.	2010	Cheung et al.	
		2001	Ren et al.	2010	Love et al.	
		2002	Cheung et al.	2011	Love et al.	
		2002	Mitropoulos & Howell	2011	Mahfouz & Kandil	
		2003	Yates	2011	Marzouk et al.	
		2004a	Cheung et al.	2012	İlter	
		2005	Arditi & Pulket	2013	Cheung & Pang	
		2006	Acharya et al.			

2.3.2.4. 'Delays'

This category considers the impact of delays on construction disputes. In order to reflect the effect of delays in construction projects, this category will be utilized. Although there is a distinct research area focusing on identification, quantification, and mitigation of delays in construction literature, this research will focus on delays as an impacting factor on dispute occurrence, potential compensations, and resolution strategies. Table 2.6 involves such studies mentioning the impact and importance of delays on construction disputes and resolution strategies in chronological order.

Table 2.6. Studies in the Literature Highlighting the Importance of Delays

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
D1	Delays	1989	Alshawi & Hope	2007	Yiu & Cheung	35
		1993	Revay	2008	Chen	
		1993	Watts & Scrivener	2008	Younis et al.	
		1997	Fenn et al.	2009	Arditi & Pulket	
		1997	Kumaraswamy	2009	Cheng et al.	
		1998	Arditi et al.	2009	El-adaway & Kandil	
		1999	Arditi & Tokdemir	2009a	Pulket & Arditi	
		2001	Ren et al.	2009b	Pulket & Arditi	
		2002	Mitropoulos & Howell	2010	Kassab et al.	
		2003	Dalton & Shehadeh	2010	Love et al.	
		2003	Pena-Mora et al.	2011	Marzouk et al.	
		2004a	Cheung et al.	2012	Chou	
		2005	Arditi & Pulket	2012	İlter	
		2005	Kilian & Gibson	2013	Cheung & Pang	
		2006	Acharya et al.	2013	Chou et al.	
		2006	Arditi & Pattanakitchamroon	2016	Yousefi et al.	
		2006	Gebken & Gibson	2018	Çevikbaş & Köksal	
		2007	Chen & Hsu			

2.3.2.5. 'Dispute Characteristics' Attributes

The fifth category of input variables involves attributes related to characteristics of a dispute and it is named as 'Dispute Characteristics'. There are 11 attributes in this category according to the findings of the literature survey. These attributes are;

- DC1 – Disputant Party
- DC2 – Phase of Occurrence
- DC3 – Dispute Sources
- DC4 – Suspension of Works due to Disputes
- DC5 – Disputed Amount (Financially)
- DC6 – Settled Amount (Financially)
- DC7 – Success Rate (Financially)
- DC8 – Presence of EoT Claim
- DC9 – Disputed EoT Amount
- DC10 – Settled EoT Amount

- DC11 – Success Rate (EoT)

Table 2.7 involves ‘Dispute Characteristics’ attributes, studies highlighting the importance of these attributes in the literature and the number of publications mentioning them. Studies are given in chronological order. Among these attributes, success rates in financial and EoT claims are not taken from the literature, but they are included to give another perspective to claimed and settled amount variables.

Table 2.7. ‘Dispute Characteristics’ Attributes

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
DC1	Disputant Party	1993	Watts & Scrivener	2008	Chen	15
		1997	Kumaraswamy	2009	Arditi & Pulket	
		1998	Arditi et al.	2009	Cheng et al.	
		1999	Arditi & Tokdemir	2009	El-adaway & Kandil	
		2005	Arditi & Pulket	2009a	Pulket & Arditi	
		2006	Acharya et al.	2009b	Pulket & Arditi	
		2007	Chau	2016	Yousefi et al.	
		2007	Chen & Hsu			
DC2	Phase of Occurrence	2004a	Cheung et al.	2012	Chou	9
		2005	Gencer	2012	İlter	
		2009	Cheng et al.	2013	Chou et al.	
		2010a	İlter	2016	Chou et al.	
		2010b	İlter			
DC3	Dispute Sources	1993	Watts & Scrivener	2008	Chen	38
		1995	Diekmann & Girard	2008	Ellis & Baiden	
		1997	Fenn et al.	2008	Younis et al.	
		1997	Kumaraswamy	2009	Arditi & Pulket	
		1998	Arditi et al.	2009	Cheng et al.	
		1999	Arditi & Tokdemir	2009	El-adaway & Kandil	
		2000	Molenaar et al.	2009a	Pulket & Arditi	
		2002	Cheung et al.	2009b	Pulket & Arditi	
		2002	Mitropoulos & Howell	2010a	İlter	
		2003	Dalton & Shehadeh	2010b	İlter	
		2003	Harmon	2010	Love et al.	
		2004a	Cheung et al.	2011	Marzouk et al.	
		2005	Arditi & Pulket	2012	Chou	
		2005	Kilian & Gibson	2012	İlter	
		2006	Acharya et al.	2013	Cheung & Pang	
		2006	Gebken & Gibson	2013	Chou et al.	
		2007	Chau	2016	Chou et al.	
		2007	Chen & Hsu	2016	Yousefi et al.	
		2007	McGeorge et al.	2018	Çevikbaş & Köksal	
D4	Suspension Of Works due to Disputes	1997	Kumaraswamy	2006	Arditi & Pattanakitchamroon	6
		1998	Arditi et al.	2008	Younis et al.	
		2006	Acharya et al.	2016	Yousefi et al.	

Table 2.7. 'Dispute Characteristics' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
DC5	Disputed Amount Financial.	1993 1993 1995 1997 1998 1999 2000 2002 2003 2004a 2005 2006 2007 2008	Revay Watts & Scrivener Diekmann & Girard Kumaraswamy Arditi et al. Arditi & Tokdemir Molenaar et al. Mitropoulos & Howell Dalton & Shehadeh Cheung et al. Arditi & Pulket Gebken & Gibson Chen & Hsu Chen	2009 2009 2009 2009a 2009b 2010a 2010b 2010 2011 2012 2012 2016 2016	Arditi & Pulket Cheng et al. Harmon Pulket & Arditi Pulket & Arditi İlter İlter Love et al. Marzouk et al. Chong & Zin İlter Lee et al. Yousefi et al.	27
DC6	Settled Amount Financial.	1995 1997 2000 2002 2006 2007 2008	Diekmann & Girard Kumaraswamy Molenaar et al. Mitropoulos & Howell Gebken & Gibson Chen & Hsu Chen	2009 2010a 2010b 2010 2011 2016	Harmon İlter İlter Love et al. Marzouk et al. Yousefi et al.	13
DC8	Presence Of EoT Claim	1989 1994 1997 2001 2002 2004a	Alshawi & Hope Heath et al. Kumaraswamy Ren et al. Cheung & Suen Cheung et al.	2007 2008 2009 2010 2012 2016	Chen & Hsu Chen Cheng et al. Kassab et al. İlter Yousefi et al.	12
DC9	Disputed EoT Amount	1993 1997 2001 2002 2003 2004a 2007	Revay Kumaraswamy Ren et al. Mitropoulos & Howell Dalton & Shehadeh Cheung et al. Chen & Hsu	2008 2009 2009 2010 2012 2016	Chen Cheng et al. El-adaway & Kandil Kassab et al. İlter Yousefi et al.	13
DC 10	Settled EoT Amount	1997 2003 2004a 2007	Kumaraswamy Dalton & Shehadeh Cheung et al. Chen & Hsu	2008 2010 2016	Chen Kassab et al. Yousefi et al.	7

2.3.2.6. 'Resolution Method Characteristics' Attributes

The sixth category of input variables involves attributes related to resolution method of a dispute and therefore, the category is named as 'Resolution Method Characteristics'. According to the findings of the literature survey, there are 13 attributes in this category and they reflect the expectations from the resolution method and consequences of the selected method. These attributes are;

- RM1 – Resolution Cost
- RM2 – Resolution Duration
- RM3 – Level of Satisfaction with the Resolution Method
- RM4 – Importance of Preserving Relationships between Parties
- RM5 – Importance of Speed of Resolution
- RM6 – Importance of Cost of Resolution
- RM7 – Importance of Bindingness of the Process
- RM8 – Importance of Confidentiality of the Process
- RM9 – Importance of Fairness in the Process
- RM10 – Importance of Flexibility in Procedures
- RM11 – Importance of Control Over the Process
- RM12 – Importance of Reaching Creative or Remedying Solutions
- RM13 – Importance of Willingness of Parties in Reaching a Solution

Table 2.8 involves resolution method characteristic attributes, studies highlighting the importance of these variables in the literature and the number of publications mentioning them. Studies are given in chronological order.

Table 2.8. 'Resolution Method Characteristics' Attributes

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
RM1	Resolution Cost	1996 2002 2002 2002 2006	York Cheung & Suen Cheung et al. Mitropoulos & Howell Gebken & Gibson	2008 2010a 2012 2012 2016	İlter & Dikbaş İlter Arıcı Chong & Zin Lee et al.	10
RM2	Resolution Duration	1996 1999 2002 2002 2005	York Cheung Cheung & Suen Cheung et al. Kilian & Gibson	2008 2010a 2012 2012 2016	İlter & Dikbaş İlter Arıcı Chong & Zin Lee et al.	10

Table 2.8. 'Resolution Method Characteristics' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
RM3	Level of Satisfaction	2000	Cheung et al.	2006	Yiu et al.	7
		2002	Cheung et al.	2010	Cheung et al.	
		2002	Mitropoulos & Howell	2011	Marzouk et al.	
		2003	Pena-Mora et al.			
RM4	Importance of Preserving Relations btw. Parties	1996	York	2007	Fenn	22
		1998	Lipsky & Seeber	2007	McGeorge et al.	
		1999	Cheung	2008	İlter & Dikbaş	
		2000	Cheung et al.	2010	Cheung et al.	
		2002	Cheung et al.	2010a	İlter	
		2002	Cheung & Suen	2010b	İlter	
		2002	Mitropoulos & Howell	2010	Kassab et al.	
		2002	OGC	2011	Mackie et al.	
		2003	Harmon	2011	Marzouk et al.	
		2003	Pena-Mora et al.	2012	Arıcı	
		2003	Yates	2016	Lee et al.	
RM5	Importance of Speed of Resolution	1996	York	2003	Pena-Mora et al.	17
		1998	Lipsky & Seeber	2006	Yiu et al.	
		1998	Cheung & Yeung	2008	İlter & Dikbaş	
		1999	Cheung	2010a	İlter	
		2000	Thompson et al.	2011	Marzouk et al.	
		2002	Cheung & Suen	2012	Arıcı	
		2002	Mitropoulos & Howell	2012	Chong & Zin	
		2002	OGC	2016	Lee et al.	
		2003	Harmon			
RM6	Importance of Cost of Resolution	1996	York	2003	Harmon	15
		1998	Lipsky & Seeber	2006	Gebken & Gibson	
		1998	Cheung & Yeung	2008	İlter & Dikbaş	
		1999	Cheung	2010a	İlter	
		2002	Cheung & Suen	2012	Arıcı	
		2002	Cheung et al.	2012	Chong & Zin	
		2002	Mitropoulos & Howell	2016	Lee et al.	
		2002	OGC			
RM7	Importance of Binding. of the Process	1996	York	2006	Jones	20
		1997	Fenn et al.	2007	İlter et al.	
		1998	Lipsky & Seeber	2008	İlter & Dikbaş	
		1999	Cheung	2009	Harmon	
		2002	Cheung & Suen	2010a	İlter	
		2002	Cheung et al.	2010b	İlter	
		2003	Harmon	2011	Mackie et al.	
		2003	Pena-Mora et al.	2012	Arıcı	
		2003	Rubin & Quintas	2012	Chong & Zin	
		2006	Gebken & Gibson	2016	Lee et al.	

Table 2.8. 'Resolution Method Characteristics' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
RM8	Importance of Confident. of the Process	1996	York	2006	Yiu et al.	18
		1998	Cheung & Yeung	2007	McGeorge et al.	
		1999	Cheung	2008	İlter & Dikbaş	
		2002	Cheung & Suen	2010a	İlter	
		2002	Cheung et al.	2010b	İlter	
		2002	OGC	2011	Mackie et al.	
		2003	Harmon	2012	Arıcı	
		2003	Pena-Mora et al.	2012	Chong & Zin	
		2003	Rubin & Quintas	2016	Lee et al.	
RM9	Importance of Fairness in the Process	1996	York	2007	McGeorge et al.	17
		1998	Cheung & Yeung	2008	İlter & Dikbaş	
		1999	Cheung	2009	Harmon	
		2002	Cheung & Suen	2010a	İlter	
		2002	Cheung et al.	2011	Mackie et al.	
		2003	Harmon	2012	Arıcı	
		2003	Pena-Mora et al.	2012	Chong & Zin	
		2003	Rubin & Quintas	2016	Lee et al.	
		2006	Jones			
RM10	Importance of Flexibility in Procedures	1996	York	2006	Yiu et al.	20
		1998	Cheung & Yeung	2007	İlter et al.	
		1999	Cheung	2007	McGeorge et al.	
		2000	Thompson et al.	2008	İlter & Dikbaş	
		2002	Cheung & Suen	2010a	İlter	
		2002	Cheung et al.	2010b	İlter	
		2003	Harmon	2011	Mackie et al.	
		2003	Pena-Mora et al.	2012	Arıcı	
		2003	Rubin & Quintas	2012	Chong & Zin	
		2006	Jones	2016	Lee et al.	
RM11	Importance of Control Over the Process	1996	York	2008	İlter & Dikbaş	17
		1998	Lipsky & Seeber	2009	Harmon	
		1999	Cheung	2010a	İlter	
		2002	Cheung & Suen	2010b	İlter	
		2002	Cheung et al.	2011	Mackie et al.	
		2003	Harmon	2012	Arıcı	
		2003	Pena-Mora et al.	2012	Chong & Zin	
		2003	Rubin & Quintas	2016	Lee et al.	
		2007	McGeorge et al.			
RM12	Importance of Creative or Remedying Solutions	1996	York	2006	Cheung et al.	17
		1998	Lipsky & Seeber	2006	Jones	
		1999	Cheung	2008	Ellis & Baiden	
		2002	Cheung & Suen	2008	İlter & Dikbaş	
		2002	Cheung et al.	2010a	İlter	
		2002	Mitropoulos & Howell	2011	Mackie et al.	
		2003	Harmon	2012	Arıcı	
		2003	Pena-Mora et al.	2016	Lee et al.	
		2003	Rubin & Quintas			

Table 2.8. 'Resolution Method Characteristics' Attributes (Continued)

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
RM13	Importance of Willing. of Parties	1996	York	2007	McGeorge et al.	15
		1998	Lipsky & Seeber	2008	İlter & Dikbaş	
		1999	Cheung	2010a	İlter	
		2001	Ren et al.	2010b	İlter	
		2002	Cheung & Suen	2011	Mackie et al.	
		2002	Cheung et al.	2012	Arıcı	
		2003	Pena-Mora et al.	2016	Lee et al.	
		2006	Yiu et al.			

2.3.2.7. 'Level of Knowledge on Resolution Method'

This category focuses on the impact of the level of knowledge of the decision-maker about specific resolution methods on selection of resolution strategy or method. In chronological order, Table 2.9 involves studies mentioning the impact and importance of the level of resolution method knowledge on strategical decision-making related to dispute resolution.

Table 2.9. Studies in the Literature Highlighting the Importance of Resolution Method Knowledge

Attr. ID	Attribute Definition	Year	Author(s)	Year	Author(s)	Number of Publications
K1	Level of Resolution Method Knowledge	1998	Lipsky & Seeber	2008	İlter & Dikbaş	16
		1999	Cheung	2010a	İlter	
		2002	Cheung et al.	2010b	İlter	
		2002	Cheung & Suen	2011	Mackie et al.	
		2003	Pena-Mora et al.	2012	Arıcı	
		2004b	Cheung et al.	2016	Lee et al.	
		2005	Kilian & Gibson	2016	Yousefi et al.	
		2007	İlter et al.	2018	Çevikbaş & Köksal	

2.3.3. Finalization of the Conceptual Model

Following the determination of output and input variables for prediction models, the conceptual model can be finalized. For this reason, input variables associated with each output variable will be given in separate tables. Table 2.10 shows the dispute occurrence framework along with input variables associated with the output, 'O1 – Dispute Occurrence'.

Table 2.10. Dispute Occurrence Framework

Output Variable	**Predictor Input Variables**		
		ID	**Attribute**
O1 – Dispute Occurrence	**Project Characteristics**	PC1	Project Location
		PC2	Project or Contract Value
		PC3	Planned Project Duration
		PC4	Type of Construction
		PC5	Type of Contractor
		PC6	Type of Employer
		PC7	Type of Contract
		PC8	Payment Method of Contract
		PC9	Project Delivery System
		PC10	Level of Design Complexity
		PC11	Level of Construction Complexity
	Skills	S1	Relationship between Parties / Individuals
		S2	Previous Experience with Each Other or Reputation
		S3	Dispute Avoidance Incentives
		S4	Communication between Parties
		S5	Working Culture & Skills of Parties
		S6	Response Rate & Communication Skills of Parties
		S7	Experience of Parties
		S8	Project Management & Coordination Skills of Parties
	Changes	C1	Occurrence of Variations
			Occurrence of Changes
			Occurrence of Unexpected Events
	Delay	D1	Ratio of Extensions to Total Planned Project Duration

As it can be seen from Table 2.10, dispute occurrence is associated with attributes related to project characteristics, skills, changes, and delay. These attributes will be tested further in order to achieve attribute elimination through Chi-Square tests so that, only the attributes that significantly impact the output (dispute occurrence) will remain.

The next table (Table 2.11) shows the potential compensations framework, which includes input variables associated with the output, ‘O2 – Potential Compensation’.

Table 2.11. Potential Compensation Framework

Output Variable	**Predictor Input Variables**		
		ID	**Attribute**
O2 – Potential Compensation	**Project Characteristics**	PC1	Project Location
		PC2	Project or Contract Value
		PC3	Planned Project Duration
		PC4	Type of Construction
		PC5	Type of Contractor
		PC6	Type of Employer
		PC7	Type of Contract
		PC8	Payment Method of Contract
		PC9	Project Delivery System
		PC10	Level of Design Complexity
		PC11	Level of Construction Complexity
	Skills	S1	Relationship between Parties / Individuals
		S2	Previous Experience with Each Other or Reputation
		S3	Dispute Avoidance Incentives
		S4	Communication between Parties
		S5	Working Culture & Skills of Parties
		S6	Response Rate & Communication Skills of Parties
		S7	Experience of Parties
		S8	Project Management & Coordination Skills of Parties
	Changes	C1	Occurrence of Variations
			Occurrence of Changes
			Occurrence of Unexpected Events
	Delay	D1	Ratio of Extensions to Total Planned Project Duration
	Dispute Characteristics	DC1	Disputant Party
		DC2	Phase of Occurrence
		DC3	Dispute Sources
		DC4	Suspension of Works due to Disputes
		DC5	Disputed Amount (Financially)
		DC8	Presence of EoT Claim
		DC9	Disputed EoT Amount

As it can be seen from Table 2.11, potential compensation to a dispute is associated with attributes related to project characteristics, skills, changes, delay, and dispute characteristics. However, it should be noted that not all dispute characteristics attributes are included in this prediction model. The variables that cannot be known prior to selection of a resolution method (i.e. settled amount) are not included in the model. Similar to the dispute occurrence prediction model, given attributes will be tested further in order to end up with attributes significantly affecting the output (potential compensation).

The next table (Table 2.12) shows the resolution method framework along with input variables associated with the output, 'O3 – Resolution Method'.

Table 2.12. Resolution Method Framework

<table>
<tr><th rowspan="2">Output Variable</th><th colspan="3">Predictor Input Variables</th></tr>
<tr><th></th><th>ID</th><th>Attribute</th></tr>
<tr><td rowspan="19">O3 – Resolution Method</td><td rowspan="11">Project Characteristics</td><td>PC1</td><td>Project Location</td></tr>
<tr><td>PC2</td><td>Project or Contract Value</td></tr>
<tr><td>PC3</td><td>Planned Project Duration</td></tr>
<tr><td>PC4</td><td>Type of Construction</td></tr>
<tr><td>PC5</td><td>Type of Contractor</td></tr>
<tr><td>PC6</td><td>Type of Employer</td></tr>
<tr><td>PC7</td><td>Type of Contract</td></tr>
<tr><td>PC8</td><td>Payment Method of Contract</td></tr>
<tr><td>PC9</td><td>Project Delivery System</td></tr>
<tr><td>PC10</td><td>Level of Design Complexity</td></tr>
<tr><td>PC11</td><td>Level of Construction Complexity</td></tr>
<tr><td rowspan="8">Skills</td><td>S1</td><td>Relationship between Parties / Individuals</td></tr>
<tr><td>S2</td><td>Previous Experience with Each Other or Reputation</td></tr>
<tr><td>S3</td><td>Dispute Avoidance Incentives</td></tr>
<tr><td>S4</td><td>Communication between Parties</td></tr>
<tr><td>S5</td><td>Working Culture & Skills of Parties</td></tr>
<tr><td>S6</td><td>Response Rate & Communication Skills of Parties</td></tr>
<tr><td>S7</td><td>Experience of Parties</td></tr>
<tr><td>S8</td><td>Project Management & Coordination Skills of Parties</td></tr>
</table>

Table 2.12. Resolution Method Framework (Continued)

Output Variable	Predictor Input Variables		
		ID	Attribute
O3 – Resolution Method	**Changes**	C1	Occurrence of Variations
			Occurrence of Changes
			Occurrence of Unexpected Events
	Delay	D1	Ratio of Extensions to Total Planned Project Duration
	Dispute Characteristics	DC1	Disputant Party
		DC2	Phase of Occurrence
		DC3	Dispute Sources
		DC4	Suspension of Works due to Disputes
		DC5	Disputed Amount (Financially)
		DC6	Settled Amount (Financially)
		DC7	Success Rate (Financially)
		DC8	Presence of EoT Claim
		DC9	Disputed EoT Amount
		DC10	Settled EoT Amount
		DC11	Success Rate (EoT)
	Resolution Method Characteristics	RM1	Resolution Cost
		RM2	Resolution Duration
		RM3	Level of Satisfaction with the Resolution Method
		RM4	Importance of Preserving Relationships btw. Parties
		RM5	Importance of Speed of Resolution
		RM6	Importance of Cost of Resolution
		RM7	Importance of Bindingness of the Process
		RM8	Importance of Confidentiality of the Process
		RM9	Importance of Fairness in the Process
		RM10	Importance of Flexibility in Procedures
		RM11	Importance of Control Over the Process
		RM12	Importance of Reaching Creative or Remedying Soln.
		RM13	Importance of Willingness in Reaching Soln.
	Level of Res. Method Know.	K1	Level of Knowledge on Litigation
		K2	Level of Knowledge on Arbitration
		K3	Level of Knowledge on DRB
		K4	Level of Knowledge on Mediation
		K5	Level of Knowledge on Senior Executive Appraisal
		K6	Level of Knowledge on Negotiation

As it can be seen from Table 2.12, dispute resolution method selection is associated with attributes related to project characteristics, skills, changes, delay, dispute characteristics, resolution method characteristics, and level of resolution method knowledge. As details will be given in following chapters, level of resolution method knowledge includes information about six methods only. This is because the collected dataset is composed of dispute cases resolved by these six methods. Similar to dispute occurrence and potential compensation prediction models, given attributes will be tested further in order to end up with attributes significantly affecting the output (resolution method).

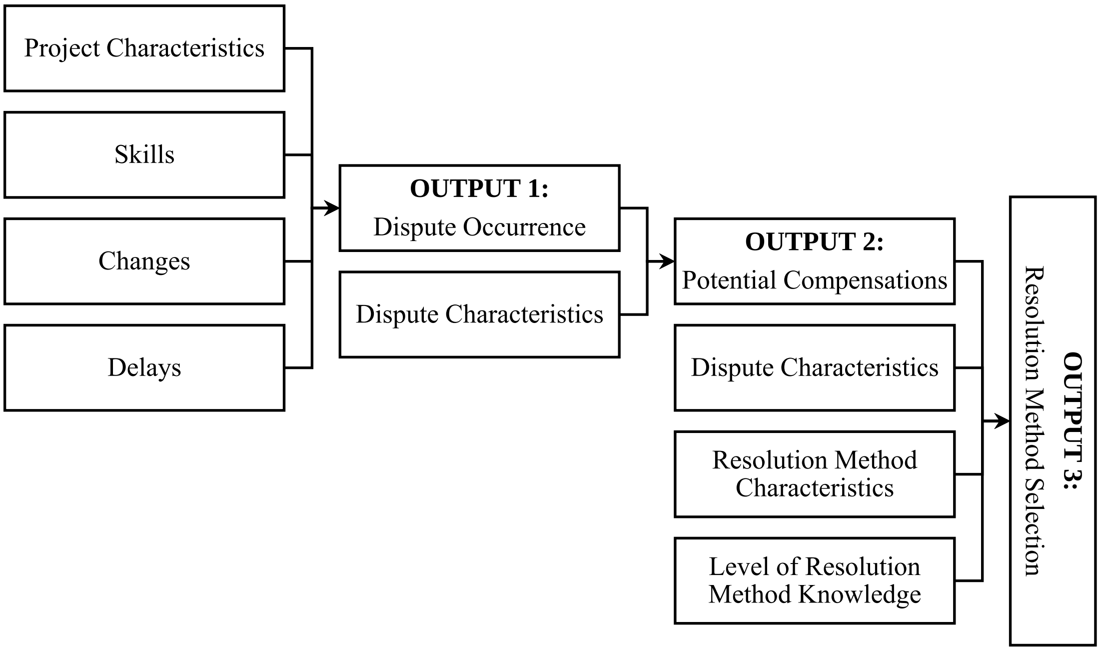


Figure 2.3. The Conceptual Model

The finalized conceptual model is shown in Figure 2.3. The model shows the dispute occurrence is associated with project characteristics, skills, changes, and delays attributes. Then, dispute characteristics are added to these attributes and they are linked to potential compensations all together. Finally, resolution method selection is associated with all the previous attributes in addition to the remaining dispute characteristics, resolution method characteristics and level of resolution method knowledge attributes.

In short, the conceptual model that will be used for development of prediction models is finalized. The conceptual model is based on findings of an extensive literature survey on construction dispute domain. According to these findings, firstly, the output variables are identified as ‘O1 – Dispute Occurrence’, ‘O2 – Potential Compensation’, and ‘O3 – Resolution Method’. This means that this thesis study should focus on development of three distinct prediction models as (1) Dispute Occurrence Prediction Model, (2) Potential Compensation Prediction Model, and (3) Resolution Method Selection Model. Following the determination of output variables for prediction models, input variables associated with each output are identified. In other words, the conceptual model is composed of identified input and output variables from the literature and the resulting conceptual model will be used for designing a questionnaire to collect empirical data related to these variables. Subsequent to empirical data collection, attributes of each prediction model will be tested further in order to identify the variables that have significant impact on outputs and to eliminate the insignificant ones. Starting with the next chapter, details related to questionnaire design depending on the developed conceptual model, empirical data collection, and tests of associations between inputs and outputs will be given.

CHAPTER 3

DEVELOPMENT OF PREDICTION MODELS

This chapter focuses on development of prediction models as (1) Dispute Occurrence Prediction Model, (2) Potential Compensation Prediction Model, and (3) Resolution Method Selection Model. In order to develop the mentioned prediction models, the following steps need to be accomplished:

1) Design of a questionnaire based on the conceptual model to collect past construction project data,
2) Empirical data collection via questionnaires,
3) Establishment of the dispute database by processing the collected data,
4) Performing attribute elimination by identifying input variables that are significantly associated with the outputs via Chi-Square statistics to achieve better generalization performance from ML algorithms.

In accordance with these steps, this chapter starts with the questionnaire design and explains the efforts for empirical data collection via designed questionnaires along with the general profile of the collected data. Then, data processing efforts will be explained such as data type conversions. Finally, Chi-Square tests of association will be performed on variables in order to eliminate the insignificant attributes. As a result of these activities, three distinct prediction models will be developed.

3.1. DESIGN OF THE QUESTIONNAIRE

There are numerous studies on construction dispute and resolution domains in the literature. However, there are only few studies focusing on interrelations between disputes and various project characteristics based on empirical data (İlter and Dikbaş, 2009). Thus, in order to fulfill this gap, this research aims to collect past project data via questionnaires conducted on decision-making authorities. The collected data will be used in development of prediction models later.

To collect the empirical data, this research designed its own questionnaire. The designed questionnaire aims to collect information related to input and output variables identified in the conceptual model (Section 2.3.3). Thus, the designed questionnaire will have questions related to three outputs (dispute occurrence, potential compensations, and resolution methods) and the associated impacting input attributes (attributes related to project characteristics, skills, changes, delays, dispute characteristics, resolution method characteristics, and level of knowledge on resolution methods).

The full version of the questionnaire can be found in Appendix A of this thesis study. Mainly, there are nine sections in the questionnaire. The first section aims to gather information about the participant. In order to obtain opinions of decision-making authorities from different professions, questionnaires are conducted with legal representatives (i.e. legal advisors, attorneys), architects, and engineers. Moreover, to understand the standing point of various management levels, authorized project participants with different roles ranging from the owner of the company to the project engineer are selected. Participants' experience in the construction industry and in their current position are also noted. In the second section, information related to project and contract characteristics are compiled. According to the conceptual model, there are 11 project characteristics related attributes and questions aiming to collect data about these 11 attributes are in the second section of the questionnaire. In the third section, data related to characteristics of parties and their organizational structures (eight attributes from the 'Skills' category of the conceptual model) are collected. In the fourth section, information related to changes are collected. In the fifth section, information related to dispute characteristics are compiled. According to the conceptual model, there are 11 dispute characteristics related attributes and questions aiming to collect data about these attributes are in the fifth section of the questionnaire. In the sixth section, information related to delays are collected. The seventh section contains questions related to resolution method characteristics. According to the conceptual model, there are 13 resolution method characteristics related attributes that should be collected in this section of the questionnaire. The eighth section identifies the level of knowledge of participants on various resolution methods. Finally, the ninth section identifies participants' interest in using prediction models that will be developed in this research.

Table 3.1 shows the list of attributes and the corresponding question(s) from the questionnaire.

Table 3.1. Attributes and Corresponding Question(s)

Attribute ID	Corresponding Question(s)	Attribute ID	Corresponding Question(s)	Attribute ID	Corresponding Question(s)
PC1	Q7	S7	Q26 - Q27	RM3	Q48
PC2	Q8	S8	Q28 - Q29	RM4	Q49a
PC3	Q9 - Q9a - Q9b	C1	Q30a-30b-Q30c	RM5	Q49b
PC4	Q10	D1	Q44a - Q44b	RM6	Q49c
PC5	Q11	DC1	Q33	RM7	Q49d
PC6	Q12	DC2	Q34	RM8	Q49e
PC7	Q13	DC3	Q35	RM9	Q49f
PC8	Q14	DC4	Q36	RM10	Q49g
PC9	Q15	DC5	Q37	RM11	Q49h
PC10	Q16	DC6	Q38	RM12	Q49i
PC11	Q17	DC7	Q39	RM13	Q49j
S1	Q18	DC8	Q40	K1	Q50
S2	Q19	DC9	Q41"	K2	Q51
S3	Q20	DC10	Q42	K3	Q52
S4	Q21	DC11	Q43	K4	Q53
S5	Q22 - Q23	RM1	Q46	K5	Q54
S6	Q24 - Q25	RM2	Q47	K6	Q55

3.2. EMPIRICAL DATA COLLECTION VIA QUESTIONNAIRES

Questionnaires are conducted via face-to-face and online meetings with construction professionals that have decision-making authority in order to collect empirical data related to past construction projects. With the goal of reflecting variations in construction types, contract documents, participants, delivery systems, business environments, etc., the data is collected from a wide variety of construction projects.

The collected dataset involves data related to all variables identified in the conceptual model.

Data related to 108 construction projects are collected via questionnaires. The dispute prediction model utilizes a dataset composed of all 108 construction projects (38 undisputed projects and 70 disputed projects), while the compensation model utilizes 82 cases (12 cases with no compensation, 38 cases with only cost compensation, 5 cases with only time compensation, and 27 cases with both cost and time compensation) out of the 108 cases collected. Notice that compensation model utilizes 82 cases, which is more than the number of disputed projects (70 disputed projects) in the dataset. However, some projects experienced more than one dispute. This is the reason why there are more disputed cases than disputed projects in the dataset. These 82 cases are the cases in which participants declared satisfaction with the compensation. Finally, the resolution method model utilizes 54 disputed cases coming from 82 disputes. These 54 disputed cases are the ones that are resolved satisfactorily according to participants.

3.2.1. Profile of Participants in the Dataset

As stated above, the dataset for this research involves 108 construction projects, which are executed in 19 different countries. These projects are obtained from 75 different construction companies from six different nationalities via face-to-face and online meetings with 78 individuals. Among these 75 companies, 16 of them (21.3%) are placed in the Engineering News-Record (ENR) Top 250 International Contractors List in 2018, which is an international index ranking the contractors all around the world according to their contracting revenues, and 21 project data (19.4%) is obtained from these companies (ENR, 2018). Thus, approximately 20% of the dataset reflects the top-level construction companies in the world (Figure 3.1).

In order to obtain opinions of different professions, participants are chosen from three groups of professions, which involve legal representatives (i.e. legal advisors, attorneys), architects, and engineers. The inner doughnut chart in Figure 3.2 shows that among 78 individuals, 58 of them were engineers (74.4%), 15 of them were legal representatives (19.2%), and 5 of them were architects (6.4%).

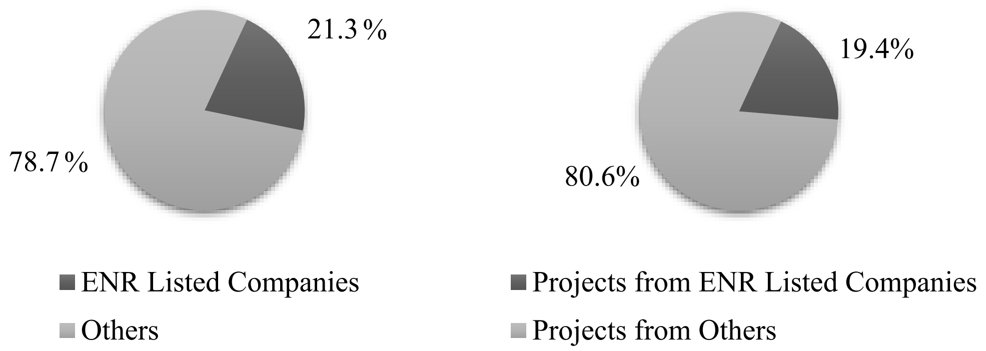


Figure 3.1. Overview of Companies and Projects in the Dataset

The outer doughnut chart in Figure 3.2 shows that the dataset is composed of 82 construction projects (75.9%) obtained from engineers, 20 projects (18.5%) from legal representatives, and 6 projects (5.6%) from architects.

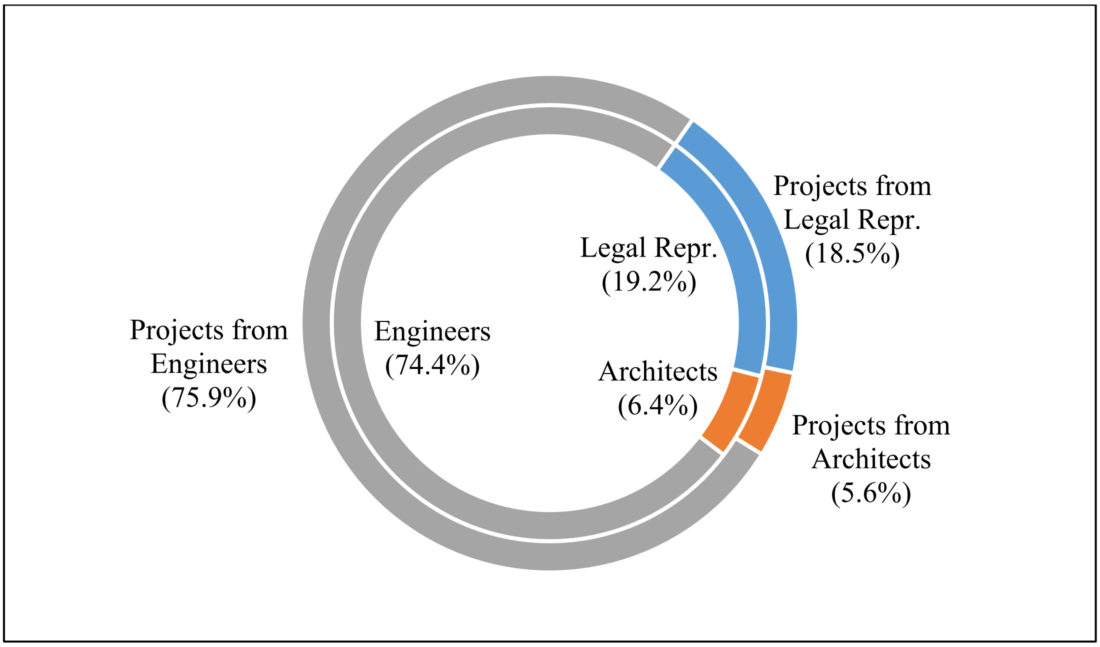


Figure 3.2. Overview of Professions of Participants in the Dataset

Dispute rates with respect to professions of participants are given in Figure 3.3. It can be observed that when the decision-maker is a legal representative, construction projects tend to experience more disputes with a rate of 80%. It is followed by engineers with 62% dispute rate and architects with 50%.

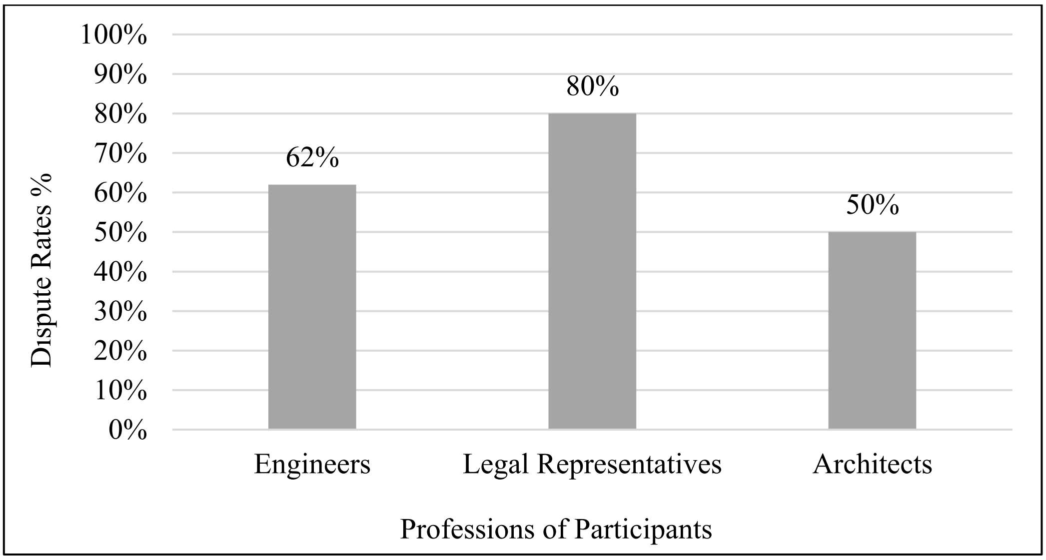


Figure 3.3. Dispute Rates with respect to Professions

In the light of the statistics given so far, it can be claimed that the dataset is dominated by individuals from engineering domain. With another perspective, dispute management decision-making in the dataset is performed by engineers mostly. In addition, the highest dispute rate belongs to legal professionals in charge of dispute management decisions. Considering that dispute management in construction industry requires both technical and legal backgrounds, the decision-support model for resolution method selection proposed in this thesis study can said to be beneficial for management personnel by overcoming this shortcoming. Moreover, regardless of the profession, the probability of encountering disputes is significantly high.

The decision-making rationale in dispute management can vary according to the level of management dealing with the dispute. Thus, the standing point of various management levels and their decision-making rationale should be understood. For this reason, the data is collected from participants having a wide variety of roles ranging from the owner of the company to the project engineer. Figure 3.4 shows the number of projects obtained from each role on the top and the distribution of roles of participants on the bottom.

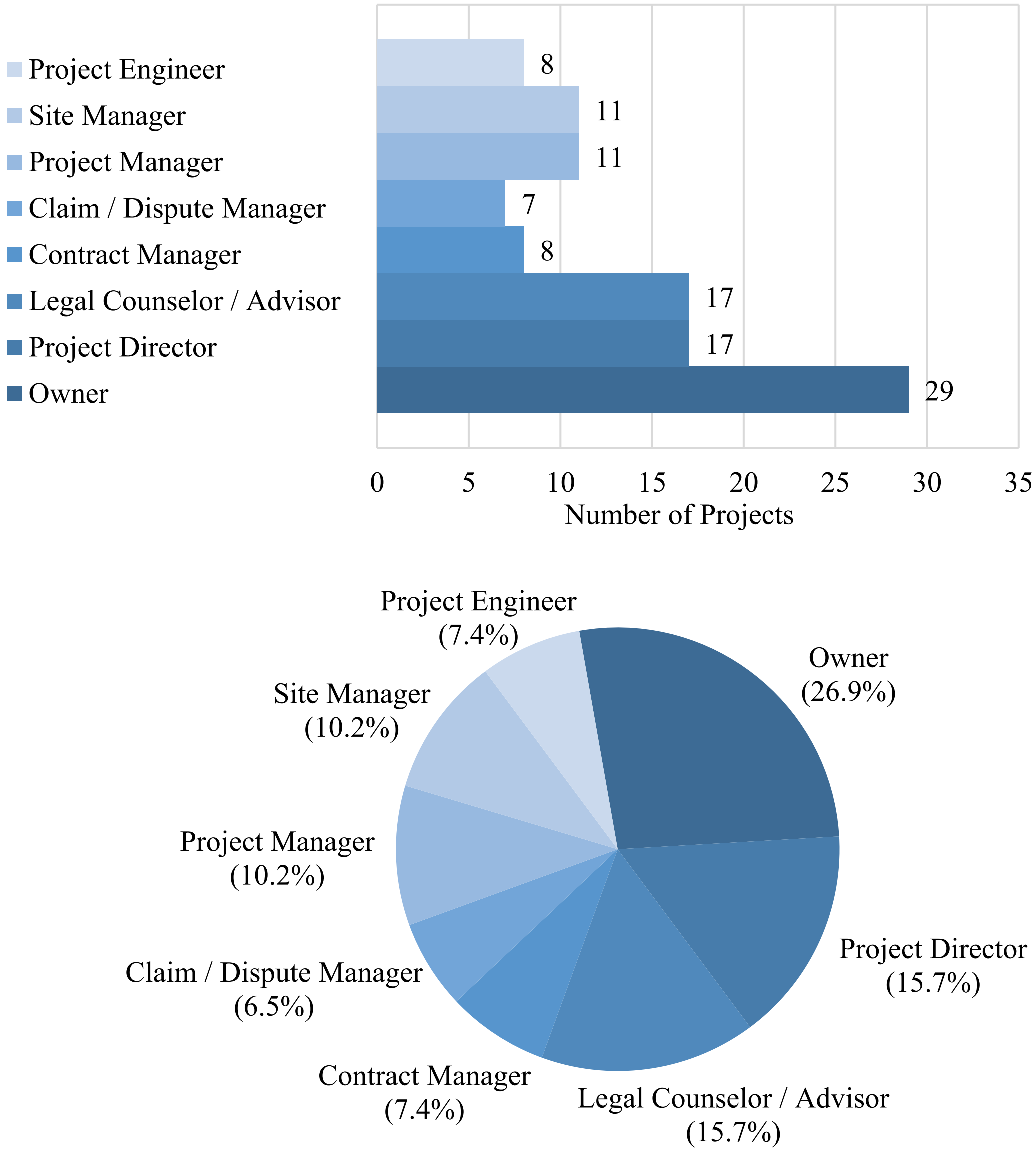


Figure 3.4. Overview of Roles of Participants in the Dataset

Dispute rates with respect to roles of participants are given in Figure 3.5. Regardless of the role in the industry, dispute rates are significantly high for all positions. Specifically, claim or dispute managers tend to arise disputes more than any other participants do (100%).

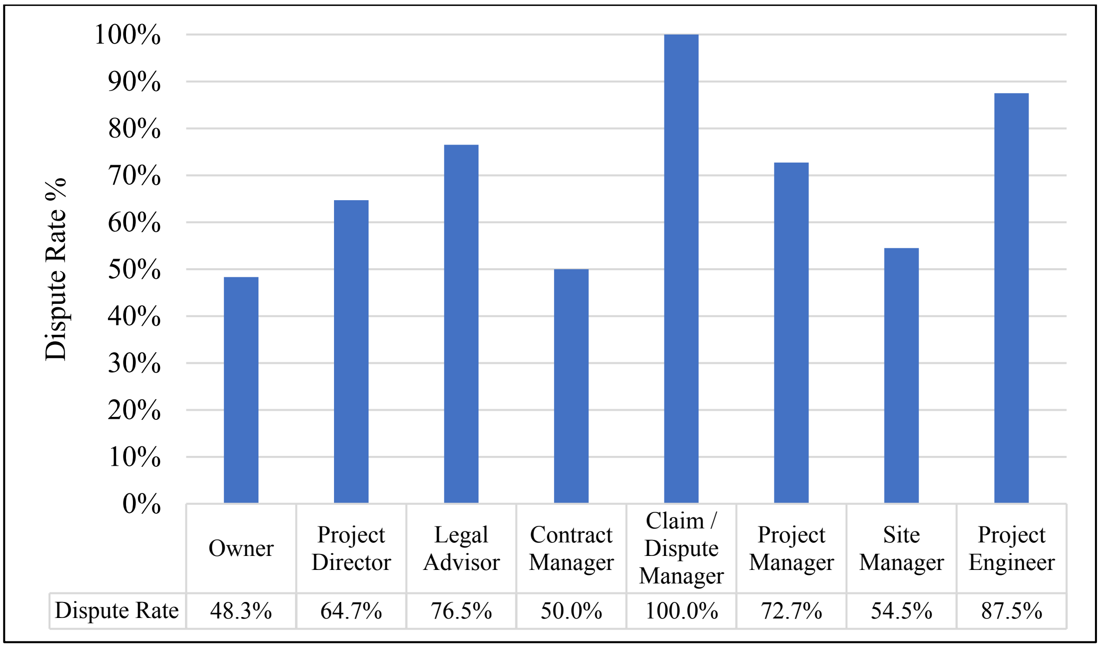
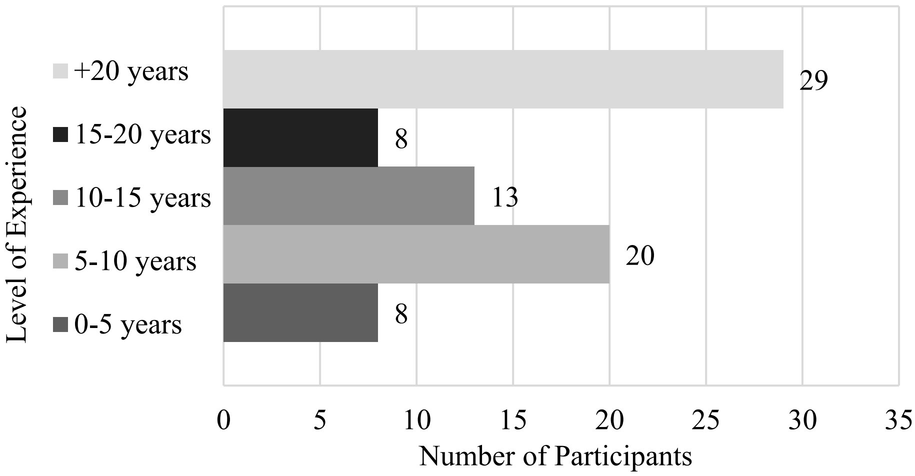


Figure 3.5. Dispute Rates with respect to Roles

Participants are selected with the purpose of reflecting opinions of decision-makers with different levels of experience. Thus, selected participants have varying levels of experience ranging from 2 years to 49 years. The average construction industry experience of participants in the dataset is approximately 18 years. Thus, it can be claimed that opinions of experienced professionals are reflected to the research mainly. Among participants, 10% can said to be inexperienced having worked in the industry less than 5 years and 53% have worked less than 15 years, while 47% can said to be experienced spending more than 15 years in the industry. There are participants having invaluable level of experience (over 20 years) in construction, which corresponds to 37% of participants. Figure 3.6 shows the level of experience of participants in the construction industry quantitywise on the top and percentagewise on the bottom.



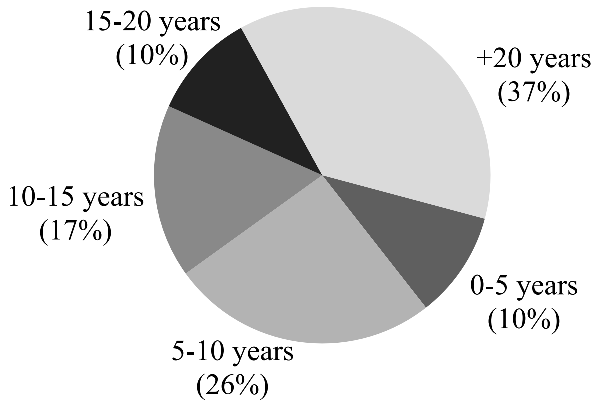


Figure 3.6. Level of Experience of Participants

3.2.2. Profile of Projects in the Dataset with respect to Output Variables

As stated earlier, the dispute occurrence prediction model utilizes the whole dataset of 108 projects. Among these projects, 38 of them did not experience any disputes, while 70 projects faced with at least one dispute. This shows that dispute occurrence in construction projects are dominant (65%) for this dataset (Figure 3.7).

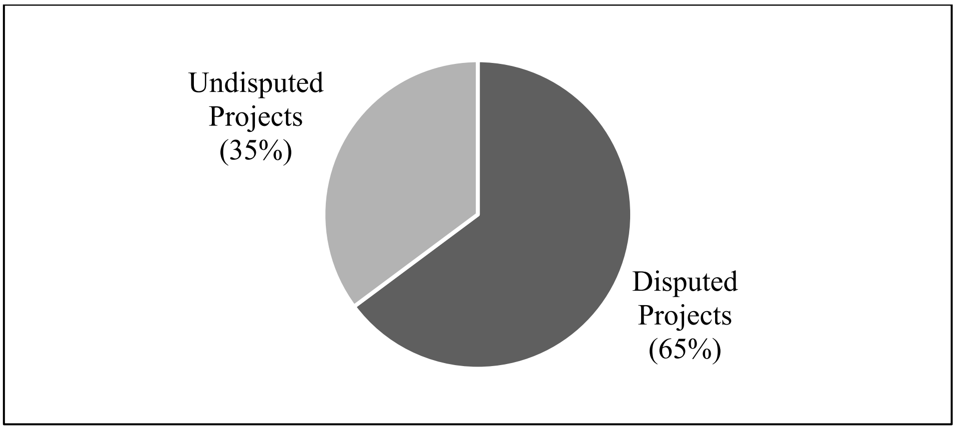


Figure 3.7. Overview of Projects with respect to Dispute Occurrence

The potential compensation prediction model utilizes 82 dispute cases coming from 70 disputed projects. Among these projects, there are 38 cases that ended up with cost compensation only (46%), 5 cases with time compensation only (6%), and 27 cases with both cost and time compensations (33%). Meanwhile, in 12 dispute cases (15%), no compensations were acquired. Figure 3.8 shows the overview of projects in the dataset with respect to compensations acquired from the disputed case.

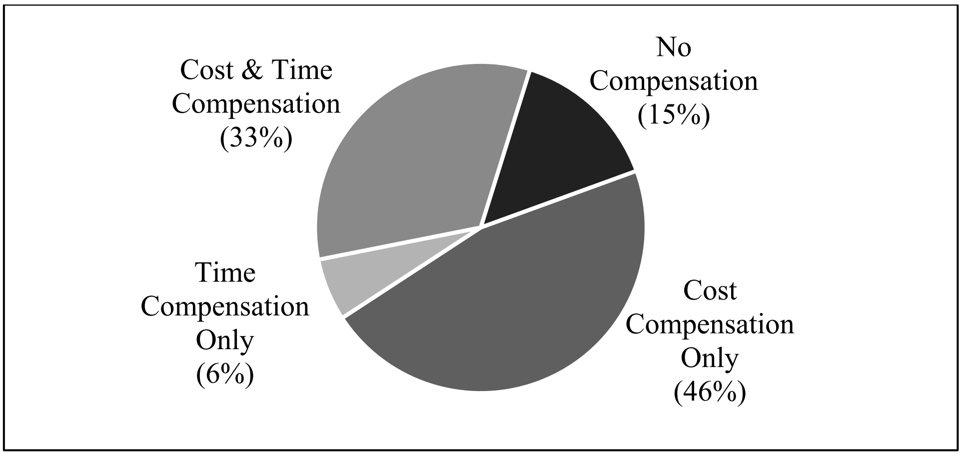


Figure 3.8. Overview of Projects with respect to Compensations

Finally, resolution method selection model utilizes 54 satisfactorily resolved disputes out of 82 disputed cases. In these 54 cases, decision-makers settled disputes utilizing litigation process in 9 cases, arbitration in 6 cases, DRB in 5 cases, mediation in 5 cases, SEA in 10 cases, and negotiation in 19 cases. Figure 3.9 shows the overview of projects in the dataset with respect to utilized resolution method. According to this, the most preferred method of resolution is the negotiation that achieved successful resolution in 35% of all cases. Considering this and advantages of negotiation mentioned in Section 2.2.2.4, it can be claimed that it is beneficial for the construction industry to utilize negotiation processes. Moreover, the SEA method, which can be viewed as a form of negotiation performed under supervision of senior executives of parties, is the second most preferred method of resolution. However, participants resorted to litigation in 17% of cases in the dataset, which makes it the third most preferred resolution method. Considering the claim that litigation should be avoided even with the best outcomes (PMI, 2016) and the consensus in the literature on litigation avoidance (Chen and Hsu, 2007; Chau, 2007; Chen, 2008; Pulket and Arditi, 2009b; Cheung et al., 2010), the situation in the dataset poses a contradiction. Similarly, another traditional method, the arbitration, follows these three methods with a preference rate of 11%. Other ADR techniques utilized for disputed cases are DRB (9%) and mediation (9%). Their preference rates are relatively low considering the need for utilizing ADR techniques in the construction industry.

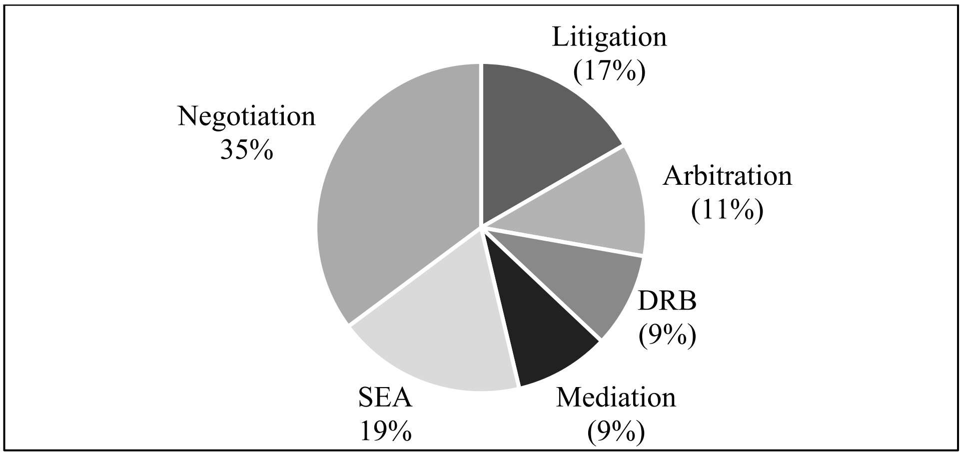


Figure 3.9. Overview of Projects with respect to Utilized Resolution Method

3.2.3. Profile of Projects in the Dataset with respect to Input Variables

Categorical labels and frequencies of each input variable (attribute) in the dataset will be given in this section. When these frequencies are reviewed, it can be seen that the dataset is capable of representing various characteristics. Thus, categorical labels and frequencies for each attribute category will be given in separate tables for further investigation.

Table 3.2 shows categorical labels and frequencies of project characteristics attributes in the dataset. Frequencies are given for all three models separately in the same table. Categorical labels for some of the attributes are organized. Firstly, project location attribute is divided into two categories as domestic and international projects. Domestic projects are the ones that are constructed in the home country of the construction company, while international ones involve parties from distinct countries. Secondly, although the project value attribute is numeric, it is converted into categorical values. Such a data type conversion is performed for computational reasons. In order to understand whether there is a statistically significant relationship between input and output variables, Chi-Square statistics will be utilized. Chi-Square statistics is a useful way of testing the existence of association between categorical variables (Weisburd and Britt, 2007) and it is one of the most effective methods in testing the hypothesis between two categorical variables (McHugh, 2013). Remembering that output variables in this research are all categorical variables, numeric input variables should also be converted into categorical values to be able to apply the Chi-Square tests. However, if the discretization process removes distinguishing features, the data type conversion may harm accuracy of classification algorithms. For this reason, special care should be given in such data type conversions. Projects in the dataset have values ranging from 172 thousand to 2.9 billion U.S. Dollars; however, they are discretized as (1) small projects with values less than 10 million U.S. Dollars, (2) medium sized projects with values between 10 to 100 million U.S. Dollars, and (3) large projects with values greater than 100 million U.S. Dollars.

Table 3.2. Project Characteristics Attributes - Categorical Labels & Frequencies

Attr. ID	Attribute	Categorical Label		Frequency in the Dataset					
				Dispute Occurrence		Potential Compensation		Resolution Method	
PC1	Project Location	1	Domestic	80	(74.1%)	50	(61.0%)	30	(55.6%)
		2	International	28	(25.9%)	32	(39.0%)	24	(44.4%)
PC2	Project or Contract Value	1	< 10 million $	44	(40.7%)	21	(25.6%)	13	(24.1%)
		2	10 - 100 million $	35	(32.4%)	27	(32.9%)	19	(35.2%)
		3	> 100 million $	29	(26.9%)	34	(41.5%)	22	(40.7%)
PC3	Planned Project Duration	1	< 1 year	31	(28.7%)	13	(15.9%)	8	(14.8%)
		2	1 - 2 years	37	(34.3%)	25	(30.5%)	17	(31.5%)
		3	2 - 3 years	21	(19.4%)	16	(19.5%)	14	(25.9%)
		4	> 3 years	19	(17.6%)	28	(34.1%)	15	(27.8%)
PC4	Type of Construct.	1	Housing	18	(16.7%)	16	(19.5%)	11	(20.4%)
		2	Commercial	10	(9.3%)	13	(15.9%)	6	(11.1%)
		3	Industrial	12	(11.1%)	7	(8.5%)	6	(11.1%)
		4	Transportation	17	(15.7%)	16	(19.5%)	13	(24.1%)
		5	Pow.Plants&Lines	8	(7.4%)	3	(3.7%)	2	(3.7%)
		6	WaterSupp.&Reser.	10	(9.3%)	9	(11.0%)	4	(7.4%)
		7	Sport&Cult.&Edu.	11	(10.2%)	8	(9.8%)	5	(9.3%)
		8	Medical	7	(6.5%)	3	(3.7%)	3	(5.6%)
		9	Public	6	(5.6%)	4	(4.9%)	3	(5.6%)
		10	Soil Works	9	(8.3%)	3	(3.7%)	1	(1.9%)
PC5	Type of Contractor	1	Single	88	(81.5%)	65	(79.3%)	43	(79.6%)
		2	Joint Venture	11	(10.2%)	11	(13.4%)	7	(13.0%)
		3	Consortium	9	(8.3%)	6	(7.3%)	4	(7.4%)
PC6	Type of Employer	1	Public	52	(48.1%)	42	(51.2%)	25	(46.3%)
		2	Private	43	(39.8%)	31	(37.8%)	22	(40.7%)
		3	PPP	13	(12.0%)	9	(11.0%)	7	(13.0%)
PC7	Type of Contract	1	Private Contracts	53	(49.1%)	43	(52.4%)	29	(53.7%)
		2	Public Procurement	36	(33.3%)	18	(22.0%)	8	(14.8%)
		3	FIDIC Red	10	(9.3%)	15	(18.3%)	11	(20.4%)
		4	FIDIC Silv./Yellow	9	(8.3%)	6	(7.3%)	6	(11.1%)
PC8	Payment Method	1	Fixed (Lump-Sum)	58	(53.7%)	44	(53.7%)	25	(46.3%)
		2	Unit Price	50	(46.3%)	38	(46.3%)	29	(53.7%)
PC9	Project Delivery System	1	DBB	67	(62.0%)	49	(59.8%)	34	(63.0%)
		2	DB	26	(24.1%)	22	(26.8%)	10	(18.5%)
		3	EPC	15	(13.9%)	11	(13.4%)	10	(18.5%)
PC10	Level of Design Complex.	1	Very Low	13	(12.0%)	14	(17.1%)	9	(16.7%)
		2	Low	16	(14.8%)	8	(9.8%)	6	(11.1%)
		3	Moderate	20	(18.5%)	14	(17.1%)	10	(18.5%)
		4	High	37	(34.3%)	28	(34.1%)	19	(35.2%)
		5	Very High	22	(20.4%)	18	(22.0%)	10	(18.5%)
PC11	Level of Construct. Complex.	1	Very Low	9	(8.3%)	9	(11.0%)	5	(9.3%)
		2	Low	15	(13.9%)	9	(11.0%)	7	(13.0%)
		3	Moderate	19	(17.6%)	13	(15.9%)	8	(14.8%)
		4	High	38	(35.2%)	30	(36.6%)	20	(37.0%)
		5	Very High	27	(25.0%)	21	(25.6%)	14	(25.9%)

A similar discretization is also performed for planned project duration attribute. This numeric attribute has values between 36 days to 2160 days. However, it is converted into a categorical variable as (1) very short projects with planned project duration less than 1 year, (2) short projects with duration between 1 to 2 years, (3) long projects with duration between 2 to 3 years, and (4) very long projects with planned project duration more than 3 years. All other project characteristics attributes are categorical variables and no further data type conversion is needed.

Some input variables reveal their patterns in the dataset at first glance. For example, the increase in the contract values cause a proportional increase in dispute rates (Figure 3.10). The highest dispute rate is observed for the group of projects with the highest contract values (86%), while projects with the lowest contract values have the lowest dispute occurrence rate (48%). The medium sized projects encounter disputes with a rate of 69%. Thus, it can be claimed that as the contract value increases, the dispute occurrence rate also increases for this dataset.

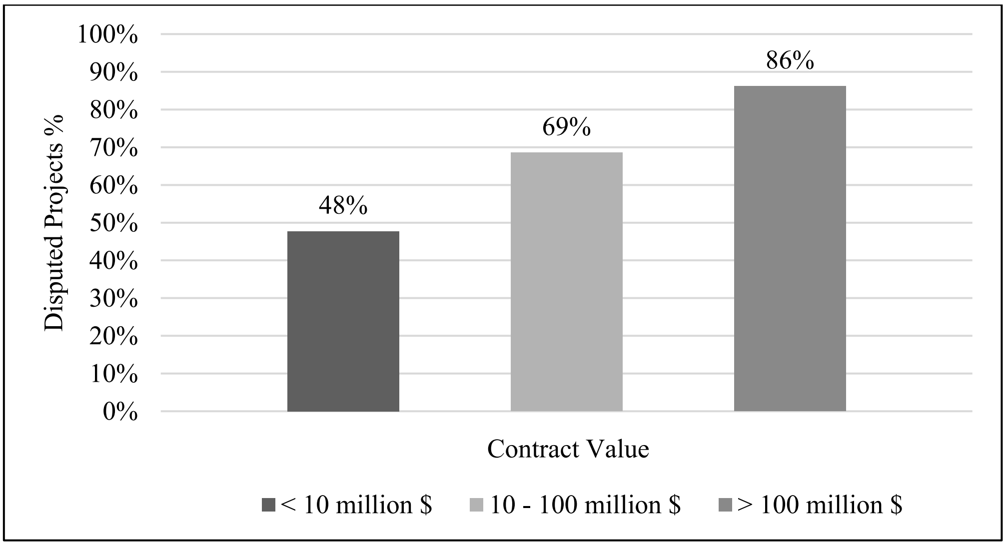


Figure 3.10. Dispute Rates with respect to Contract Values

Figure 3.11 shows dispute rates with respect to planned project duration. For this dataset, dispute rates indicate that the longer the project duration is, the higher the dispute rate. The highest dispute occurrence rate is observed for the group of projects with the longest planned project duration (100%) and this rate decreases with the decreasing project duration. Long projects with planned duration between 2 and 3 years exhibit 71% dispute rate, short projects with planned duration between 1 and 2 years encounter 62% dispute occurrence, and very short projects with planned duration less than 1 year have the lowest dispute rate with 42%.

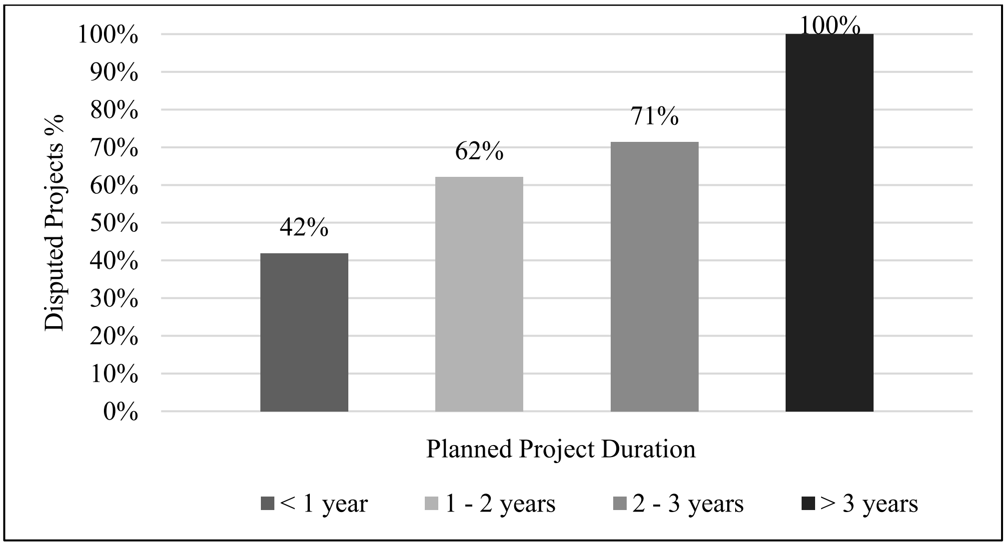


Figure 3.11. Dispute Rates with respect to Planned Project Duration

Figure 3.12 is dispute rates with respect to project location, which shows that encountering disputes in international projects (82%) is more likely compared to domestic projects (59%) for this dataset. Thus, it can be claimed that construction professionals should pay more attention to dispute management in projects with participants from distinct countries.

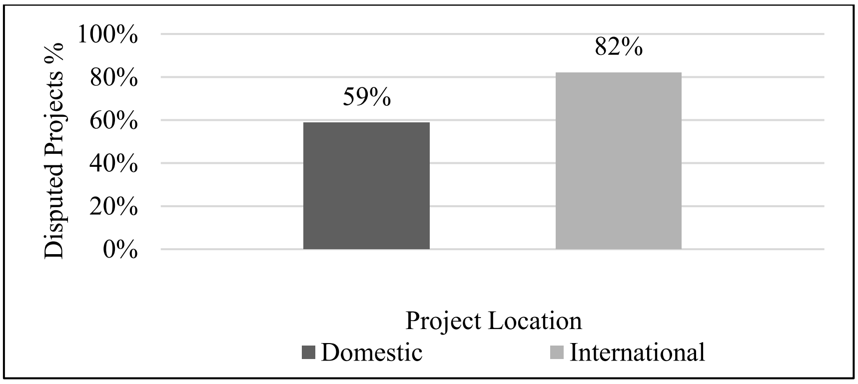


Figure 3.12. Dispute Rates with respect to Project Location

Table 3.3 shows categorical levels and frequencies of skills attributes in the dataset. Frequencies are given for all three models in the same table. For each attribute, Level 1 corresponds to weakest (worst) level and Level 5 to strongest (best) level.

Table 3.3. Skills Attributes – Levels & Frequencies

Attr. ID	Attribute	Levels	Frequency in the Dataset: Dispute Occurrence		Potential Compensation		Resolution Method	
S1	Relationship between Parties / Individuals	Level 1	10	(9.3%)	11	(13.4%)	3	(5.6%)
		Level 2	14	(13.0%)	15	(18.3%)	9	(16.7%)
		Level 3	12	(11.1%)	14	(17.1%)	11	(20.4%)
		Level 4	48	(44.4%)	35	(42.7%)	25	(46.3%)
		Level 5	24	(22.2%)	7	(8.5%)	6	(11.1%)
S2	Previous Experience with Each Other or Reputation	Level 1	2	(1.9%)	3	(3.7%)	1	(1.9%)
		Level 2	7	(6.5%)	9	(11.0%)	5	(9.3%)
		Level 3	20	(18.5%)	16	(19.5%)	11	(20.4%)
		Level 4	42	(38.9%)	31	(37.8%)	21	(38.9%)
		Level 5	37	(34.3%)	23	(28.0%)	16	(29.6%)
S3	Dispute Avoidance Incentives	Level 1	46	(42.6%)	37	(45.1%)	26	(48.1%)
		Level 2	3	(2.8%)	4	(4.9%)	4	(7.4%)
		Level 3	16	(14.8%)	15	(18.3%)	7	(13.0%)
		Level 4	21	(19.4%)	10	(12.2%)	9	(16.7%)
		Level 5	22	(20.4%)	16	(19.5%)	8	(14.8%)
S4	Communication between Parties	Level 1	7	(6.5%)	7	(8.5%)	2	(3.7%)
		Level 2	18	(16.7%)	18	(22.0%)	10	(18.5%)
		Level 3	25	(23.1%)	24	(29.3%)	16	(29.6%)
		Level 4	34	(31.5%)	22	(26.8%)	16	(29.6%)
		Level 5	24	(22.2%)	11	(13.4%)	10	(18.5%)
S5-1	Working Culture & Skills of Represented Party	Level 1	1	(0.9%)	1	(1.2%)	1	(1.9%)
		Level 2	7	(6.5%)	7	(8.5%)	4	(7.4%)
		Level 3	20	(18.5%)	21	(25.6%)	16	(29.6%)
		Level 4	45	(41.7%)	25	(30.5%)	17	(31.5%)
		Level 5	35	(32.4%)	28	(34.1%)	16	(29.6%)

Table 3.3. Skills Attributes – Levels & Frequencies (Continued)

Attr. ID	Attribute	Levels	Frequency in the Dataset: Dispute Occurrence		Potential Compensation		Resolution Method	
S5-2	Working Culture & Skills of Counter Party	Level 1	18	16.7%	22	26.8%	11	20.4%
		Level 2	17	15.7%	15	18.3%	8	14.8%
		Level 3	29	26.9%	21	25.6%	16	29.6%
		Level 4	27	25.0%	17	20.7%	14	25.9%
		Level 5	17	15.7%	7	8.5%	5	9.3%
S6-1	Response Rate & Communication Skills of Represented Party	Level 1	3	2.8%	4	4.9%	4	7.4%
		Level 2	6	5.6%	5	6.1%	3	5.6%
		Level 3	24	22.2%	18	22.0%	12	22.2%
		Level 4	42	38.9%	26	31.7%	19	35.2%
		Level 5	33	30.6%	29	35.4%	16	29.6%
S6-2	Response Rate & Communication Skills of Counter Party	Level 1	19	17.6%	22	26.8%	11	20.4%
		Level 2	21	19.4%	15	18.3%	14	25.9%
		Level 3	24	22.2%	19	23.2%	12	22.2%
		Level 4	26	24.1%	18	22.0%	15	27.8%
		Level 5	18	16.7%	8	9.8%	2	3.7%
S7-1	Experience of Represented Party	Level 1	1	0.9%	1	1.2%	1	1.9%
		Level 2	3	2.8%	3	3.7%	3	5.6%
		Level 3	15	13.9%	14	17.1%	8	14.8%
		Level 4	33	30.6%	25	30.5%	19	35.2%
		Level 5	56	51.9%	39	47.6%	23	42.6%
S7-2	Experience of Counter Party	Level 1	11	10.2%	10	12.2%	5	9.3%
		Level 2	16	14.8%	16	19.5%	13	24.1%
		Level 3	22	20.4%	16	19.5%	9	16.7%
		Level 4	30	27.8%	25	30.5%	20	37.0%
		Level 5	29	26.9%	15	18.3%	7	13.0%
S8-1	Project Management & Coordination Skills of Represented Party	Level 1	1	0.9%	1	1.2%	0	0.0%
		Level 2	4	3.7%	5	6.1%	5	9.3%
		Level 3	22	20.4%	21	25.6%	13	24.1%
		Level 4	50	46.3%	33	40.2%	21	38.9%
		Level 5	31	28.7%	22	26.8%	15	27.8%
S8-2	Project Management & Coordination Skills of Counter Party	Level 1	10	9.3%	9	11.0%	4	7.4%
		Level 2	26	24.1%	23	28.0%	15	27.8%
		Level 3	32	29.6%	31	37.8%	21	38.9%
		Level 4	31	28.7%	18	22.0%	13	24.1%
		Level 5	9	8.3%	1	1.2%	1	1.9%

Table 3.4 shows categorical labels and frequencies of changes. Frequencies are given for all three models separately in the same table. In this attribute category, changes, variations, and unexpected events are considered as indicators for changes. Occurrence of any one these events is considered as occurrence of a change and if none of them occurred, it is considered as no change.

Table 3.4. Changes – Categorical Labels & Frequencies

Attr. ID	Attribute	Categorical Label		Frequency in the Dataset: Dispute Occurrence		Potential Compensation		Resolution Method	
C1	Changes	0	No	67	(62.0%)	30	(36.6%)	22	(40.7%)
		1	Yes	41	(38.0%)	52	(63.4%)	32	(59.3%)

As stated earlier, the ratio of extensions to total planned project duration is considered as delay indicator. However, this numeric attribute should be converted into a categorical one. Thus, the delay attribute is categorized as (1) ratio equals to 0% (no extension), (2) ratio between 0% and 20%, (3) ratio between 20% and 40%, and (4) ratio greater than 40%. Table 3.5 shows categorical labels and frequencies related to delays. Frequencies are given for all three models separately in the same table.

Table 3.5. Delays – Categorical Labels & Frequencies

Attr. ID	Attribute	Categorical Label		Frequency in the Dataset: Dispute Occurrence		Potential Compensation		Resolution Method	
D1	Delays	1	Ratio = 0%	44	(40.7%)	21	(25.6%)	15	(27.8%)
		2	Ratio 0% - 20%	24	(22.2%)	21	(25.6%)	11	(20.4%)
		3	Ratio 20% 40%	17	(15.7%)	17	(20.7%)	9	(16.7%)
		4	Ratio > 40%	23	(21.3%)	23	(28.1%)	19	(35.2%)

Figure 3.13 is dispute rates with respect to delays (ratio of time extensions to planned project duration), which shows that the higher the delay, the higher the dispute rate in general. The highest dispute rate (94%) is observed for the group of projects with delay ratio between 20% and 40%, while the second highest dispute rate is observed for projects with delay ratio greater than 40% (78%). The dispute occurrence rate drops to 63% for projects with delay ratio between 0% and 20%. For projects with no extensions, the dispute occurrence rate is 48%.

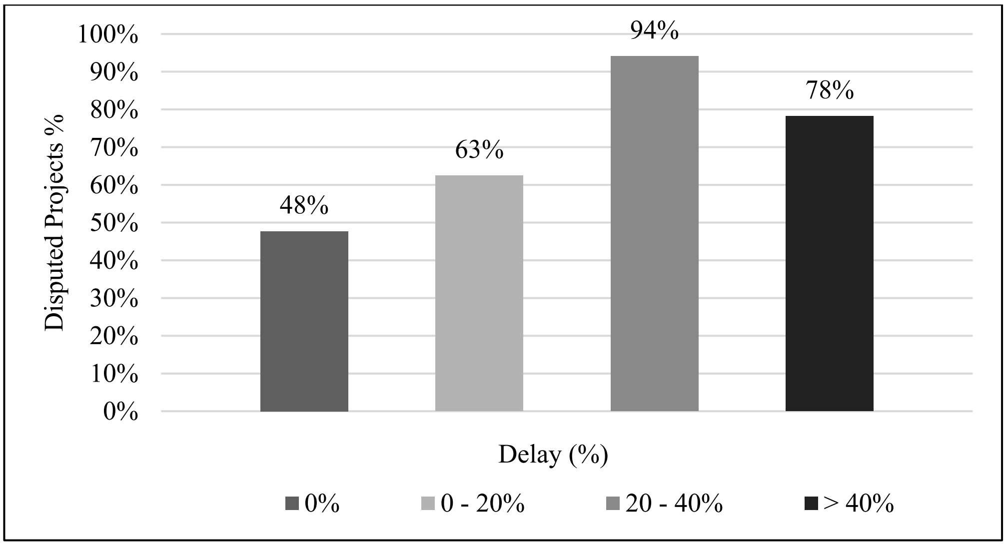


Figure 3.13. Dispute Rates with respect to Delays

Table 3.6 shows categorical labels and frequencies of dispute characteristics attributes in the dataset. Dispute characteristics attributes are not taken as inputs in dispute occurrence models. Therefore, frequencies are given for potential compensation and resolution method models only. In addition, DC6 – Settled Amount (Financially), DC7 – Success Rate (Financially), DC10 – Settled EoT Amount, and DC11 – Success Rate (EoT) are not used as inputs in potential compensation model. Therefore, corresponding cells in Table 3.6 are empty. Finally, numeric dispute characteristics attributes (disputed amount (financially), settled amount (financially), success rate (financially), disputed EoT amount, settled EoT amount, success rate (EoT)) are converted to categorical variables.

The lowest financial dispute amount in the dataset is 22 thousand U.S. Dollars and the highest one is 330 million U.S. Dollars. Meanwhile, the settlement amount is between 0 and 170 million U.S Dollars. The success rate in acquiring the disputed amount varies from 0% to 100%. Similarly, the shortest EoT dispute is 60 days and the longest is 1100 days, while EoT settlement is between 0 and 800 days. The success rate in acquiring the disputed EoT varies from 0% to 100%. All these numeric attributes are converted to categorical values for computational reasons explained earlier.

Table 3.6. Dispute Characteristics Attributes – Categorical Labels & Frequencies

Attr. ID	Attribute	Categorical Label		Frequency in the Dataset: Potential Compensation		Frequency in the Dataset: Resolution Method	
DC1	Disputant Party	1	Owner / Employer	11	13.4%	9	16.7%
		2	Contractor	71	86.6%	45	83.3%
DC2	Phase of Occurrence	1	Planning&Design&Tender&Procurement	2	2.4%	0	0.0%
		2	Construction	69	84.1%	47	87.0%
		3	Transfer & Repair & Maintenance	11	13.4%	7	13.0%
DC3	Dispute Source	1	Cost compensation of change orders	7	8.5%	3	5.6%
		2	Time & Cost compensation of change orders	21	25.6%	17	31.5%
		3	Measurement & valuation of contracted works	6	7.3%	6	11.1%
		4	Extended overhead due to extensions	1	1.2%	0	0.0%
		5	Delay in site handover & possession	4	4.9%	3	5.6%
		6	Defects, errors and poor quality	8	9.8%	6	11.1%
		7	Contractor fails to act as a prudent merchant	8	9.8%	6	11.1%
		8	Delays in payments	7	8.5%	3	5.6%
		9	Errors or substantial changes in BoQ	4	4.9%	4	7.4%
		10	Inadequate site or soil investigation	8	9.8%	5	9.3%
		11	Interpretation of contract clauses	8	9.8%	1	1.9%
DC4	Suspension of Works	0	No	52	63.4%	33	61.1%
		1	Yes	30	36.6%	21	38.9%
DC5	Disputed Amount (Financially)	1	< 5 million $	39	47.6%	24	44.4%
		2	5 - 25 million $	22	26.8%	15	27.8%
		3	25 - 75 million $	9	11.0%	8	14.8%
		4	> 75 million $	12	14.6%	7	13.0%
DC6	Settled Amount (Financially)	1	0 $	-	-	3	5.6%
		2	< 1 million $			15	27.8%
		3	1 - 5 million $			15	27.8%
		4	5 - 25 million $			9	16.7%
		5	> 25 million $			12	22.2%
DC7	Success Rate (Financially)	1	0%	-	-	1	1.9%
		2	0% - 25%			1	1.9%
		3	25% - 50%			14	25.9%
		4	50% - 75%			19	35.2%
		5	> 75%			19	35.2%
DC8	Presence of EoT Claim	0	No	46	56.1%	28	51.9%
		1	Yes	36	43.9%	26	48.1%
DC9	Disputed EoT Amount	1	0 days	46	56.1%	28	51.9%
		2	0 - 6 months	13	15.9%	10	18.5%
		3	6 months - 1 year	11	13.4%	6	11.1%
		4	> 1 year	12	14.6%	10	18.5%
DC10	Settled EoT Amount	1	0 days	-	-	29	53.7%
		2	0 - 6 months			9	16.7%
		3	6 months - 1 year			8	14.8%
		4	> 1 year			8	14.8%
DC11	Success Rate (EoT)	1	0%	-	-	29	53.7%
		2	0% - 25%			0	0.0%
		3	25% - 50%			0	0.0%
		4	50% - 75%			4	7.4%
		5	> 75%			21	38.9%

Table 3.7 shows categorical labels and frequencies of resolution method characteristics attributes in the dataset. Resolution method characteristics attributes are not taken as inputs in dispute occurrence and potential compensation models. Therefore, frequencies are given for resolution method model only. Moreover, as stated earlier, the resolution method model utilizes 54 disputes coming from 82 disputed cases. These 54 disputed cases are the ones that are resolved satisfactorily according to participants. In other words, 54 cases in this dataset are the cases that have level of satisfaction with the resolution method (attribute RM3) higher than Level 3. Here, Level 1 corresponds to the weakest (worst) level and Level 5 corresponds to the strongest (best) level. In addition, 10 resolution method characteristics are ranked according to their importance for the decision-maker. These 10 characteristics are; (1) RM4 – Preserving relationships between parties, (2) RM5 – Speed of resolution, (3) RM6 – Cost of resolution, (4) RM7 – Bindingness of the process, (5) RM8 – Confidentiality of the process, (6) RM9 – Fairness in the process, (7) RM10 – Flexibility in procedures, (8) RM11 – Control over the process, (9) RM12 – Reaching creative or remedying solution, and (10) RM13 – Willingness in reaching solutions. Same importance value cannot be given to two different attributes (each importance ranking can be given just once for an attribute). Here, Rank 1 corresponds to the most important attribute and Rank 10 corresponds to the least important attribute. Finally, numeric attributes (RM1 – Resolution Cost and RM2 – Resolution Duration) are converted to categorical variables for computational reasons. Costs of resolution for the cases in the dataset varies from 0 to 28 million U.S. Dollars. Meanwhile, the shortest resolution duration in the dataset is 3 days and the longest is 3650 days. These two numeric attributes are converted to categorical values as can be observed from Table 3.7.

Table 3.7. Resolution Method Characteristics Attributes – Categorical Labels & Frequencies

Attr. ID	Attribute	Categorical Label (Level or Rank)		Frequency in the Dataset Resolution Method	
RM1	Resolution Cost	1	0 $	29	53.7%
		2	0 - 100 000 $	5	9.3%
		3	100 000 - 350 000 $	7	13.0%
		4	350 000 - 1 000 000 $	8	14.8%
		5	> 1 000 000 $	5	9.3%
RM2	Resolution Duration	1	< 2 weeks	11	20.4%
		2	2 - 4 weeks	10	18.5%
		3	1 - 3 months	16	29.6%
		4	3 - 6 months	2	3.7%
		5	6 months - 2.5 years	8	14.8%
		6	> 2.5 years	7	13.0%
RM3	Level of Satisfaction with the Resolution Method	Level 1		0	0.0%
		Level 2		0	0.0%
		Level 3		26	48.1%
		Level 4		13	24.1%
		Level 5		15	27.8%
RM4	Importance of Preserving Relationships between Parties	Rank 1		18	33.3%
		Rank 2		3	5.6%
		Rank 3		3	5.6%
		Rank 4		4	7.4%
		Rank 5		9	16.7%
		Rank 6		5	9.3%
		Rank 7		3	5.6%
		Rank 8		0	0.0%
		Rank 9		3	5.6%
		Rank 10		6	11.1%
RM5	Importance of Speed of Resolution	Rank 1		12	22.2%
		Rank 2		11	20.4%
		Rank 3		6	11.1%
		Rank 4		12	22.2%
		Rank 5		3	5.6%
		Rank 6		3	5.6%
		Rank 7		5	9.3%
		Rank 8		2	3.7%
		Rank 9		0	0.0%
		Rank 10		0	0.0%
RM6	Importance of Cost of Resolution	Rank 1		5	9.3%
		Rank 2		8	14.8%
		Rank 3		9	16.7%
		Rank 4		9	16.7%
		Rank 5		3	5.6%
		Rank 6		3	5.6%
		Rank 7		9	16.7%
		Rank 8		6	11.1%
		Rank 9		2	3.7%
		Rank 10		0	0.0%

Table 3.7. Resolution Method Characteristics Attributes – Categorical Labels & Frequencies (Continued)

Attr. ID	Attribute	Categorical Label (Level or Rank)	Frequency in the Dataset	
			Resolution Method	
RM7	Importance of Bindingness of the Process	Rank 1	0	0.0%
		Rank 2	5	9.3%
		Rank 3	5	9.3%
		Rank 4	2	3.7%
		Rank 5	8	14.8%
		Rank 6	11	20.4%
		Rank 7	4	7.4%
		Rank 8	10	18.5%
		Rank 9	8	14.8%
		Rank 10	1	1.9%
RM8	Importance of Confidentiality of the Process	Rank 1	1	1.9%
		Rank 2	2	3.7%
		Rank 3	2	3.7%
		Rank 4	2	3.7%
		Rank 5	0	0.0%
		Rank 6	7	13.0%
		Rank 7	3	5.6%
		Rank 8	3	5.6%
		Rank 9	11	20.4%
		Rank 10	23	42.6%
RM9	Importance of Fairness in the Process	Rank 1	12	22.2%
		Rank 2	7	13.0%
		Rank 3	4	7.4%
		Rank 4	5	9.3%
		Rank 5	1	1.9%
		Rank 6	4	7.4%
		Rank 7	9	16.7%
		Rank 8	3	5.6%
		Rank 9	5	9.3%
		Rank 10	4	7.4%
RM10	Importance of Flexibility in Procedures	Rank 1	1	1.9%
		Rank 2	1	1.9%
		Rank 3	9	16.7%
		Rank 4	7	13.0%
		Rank 5	6	11.1%
		Rank 6	3	5.6%
		Rank 7	8	14.8%
		Rank 8	10	18.5%
		Rank 9	4	7.4%
		Rank 10	5	9.3%
RM11	Importance of Control Over the Process	Rank 1	0	0.0%
		Rank 2	2	3.7%
		Rank 3	4	7.4%
		Rank 4	2	3.7%
		Rank 5	8	14.8%
		Rank 6	9	16.7%
		Rank 7	7	13.0%
		Rank 8	9	16.7%
		Rank 9	7	13.0%
		Rank 10	6	11.1%

Table 3.7. Resolution Method Characteristics Attributes – Categorical Labels & Frequencies (Continued)

Attr. ID	Attribute	Categorical Label (Level or Rank)	Frequency in the Dataset	
			Resolution Method	
RM12	Importance of Reaching Creative or Remedying Solutions	Rank 1	2	3.7%
		Rank 2	9	16.7%
		Rank 3	1	1.9%
		Rank 4	8	14.8%
		Rank 5	11	20.4%
		Rank 6	6	11.1%
		Rank 7	5	9.3%
		Rank 8	5	9.3%
		Rank 9	6	11.1%
		Rank 10	1	1.9%
RM13	Importance of Willingness in Reaching Solutions	Rank 1	3	5.6%
		Rank 2	6	11.1%
		Rank 3	11	20.4%
		Rank 4	3	5.6%
		Rank 5	5	9.3%
		Rank 6	3	5.6%
		Rank 7	1	1.9%
		Rank 8	6	11.1%
		Rank 9	8	14.8%
		Rank 10	8	14.8%

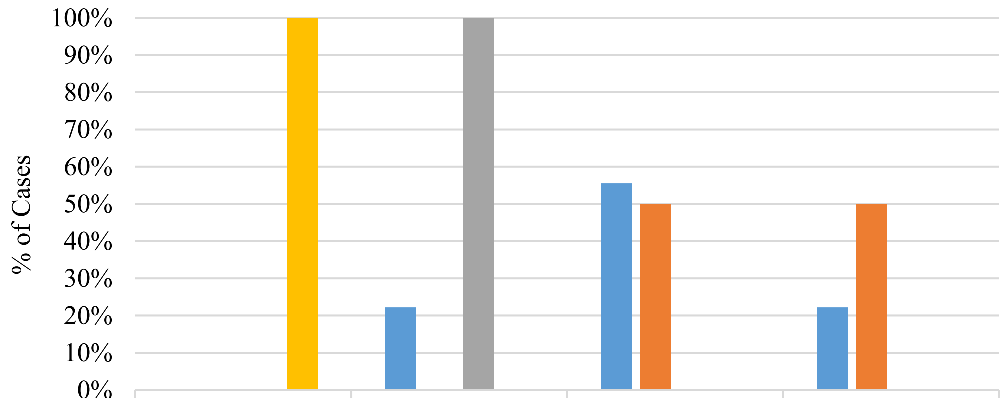
Besides the SEA and negotiation techniques that do not have additional resolution costs, other methods are further examined in terms of resolution costs and details are given in Figure 3.14.

In the dataset, there are 9 litigated cases with an average resolution cost of 1 million U.S. Dollars approximately. The maximum litigation cost was as high as 3 million U.S. Dollars approximately. The resolution cost is higher than 350,000 U.S. Dollars in 78% of litigated cases.

Although litigation processes commonly associated with high costs, their costs are lower compared to arbitration for this dataset. Arbitration is the most expensive method of resolution in the dataset with all arbitrated cases having a cost greater than 350,000 U.S. Dollars. The minimum arbitration cost was 0.5 million U.S. Dollars and the maximum was 27.7 million U.S. Dollars.

The average resolution cost associated with DRB is 252,000 U.S. Dollars that makes the technique the most expensive ADR method behind the traditional ones. DRB costs are variable depending on factors such as the size of the board, the degree of involvement, the number of disputed issues the board listened, etc. In the dataset, all cases resolved by DRB technique have costs between 100,000 and 350,000 U.S. Dollars.

Mediation seems to be an economic ADR method with an average resolution cost of 18,652 U.S. Dollars. The maximum amount associated with a mediated case is 50,000 U.S. Dollars.

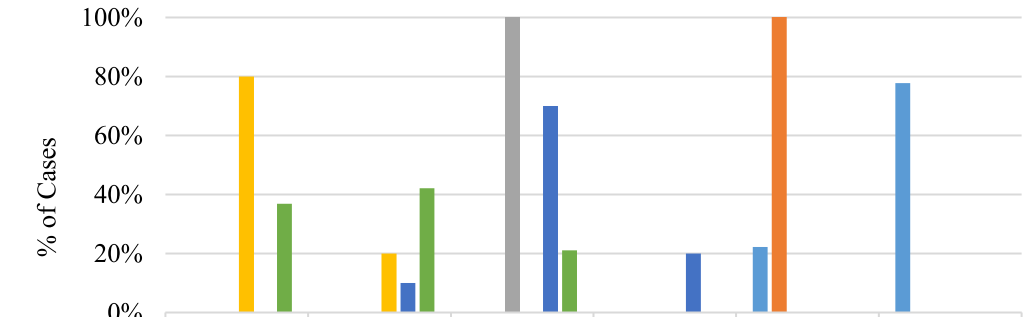


	0 - 100.000$	100.000 - 350.000$	350.000 - 1.000.000$	> 1.000.000$
Litigation	0%	22%	56%	22%
Arbitration	0%	0%	50%	50%
DRB	0%	100%	0%	0%
Mediation	100%	0%	0%	0%

Figure 3.14. Resolution Methods with respect to Resolution Costs

All resolution methods are also examined in terms of resolution duration and details are given in Figure 3.15. For this dataset, litigated cases have an average resolution duration of 1620 days, where the shortest litigated case is 720 days and the longest one is 3650 days. The resolution duration is longer than 2.5 years in 78% of litigated cases. Meanwhile, arbitrated cases have an average resolution duration of 637 days.

The shortest arbitrated case lasted for 365 days and the longest one for 900 days. The average resolution duration is 55 days for DRB cases and 13 days for mediated cases. None of the DRB cases have lasted longer than 3 months and 80% of the mediated cases are resolved in less than 2 weeks. The average resolution duration associated with cases resolved by SEA technique is 78 days with the longest case settled in 180 days. Meanwhile, the average resolution duration associated with negotiated cases is 23 days with the longest case settled in 45 days. Thus, it can be claimed that this dataset proves the importance of ADR techniques in terms of resolution costs and durations. ADR techniques have significantly lower resolution cost and shorter resolution duration compared to traditional methods of litigation and arbitration.



	< 2 weeks	2 - 4 weeks	1 - 3 months	3 - 6 months	6 months - 2.5 years	> 2.5 years
Litigation	0%	0%	0%	0%	22%	78%
Arbitration	0%	0%	0%	0%	100%	0%
DRB	0%	0%	100%	0%	0%	0%
Mediation	80%	20%	0%	0%	0%	0%
SEA	0%	10%	70%	20%	0%	0%
Negotiation	37%	42%	21%	0%	0%	0%

Figure 3.15. Resolution Methods with respect to Resolution Durations

Participants rated their own level of knowledge using 5-point Likert scale based on their theoretical and practical level of knowledge on processes related to resolution methods in the dataset. Table 3.8 shows levels and frequencies of resolution method knowledge of participants in the dataset. Level of resolution method knowledge is not taken as input in dispute occurrence and potential compensation models. Therefore, frequencies are given for resolution method selection model only. For each attribute, Level 1 corresponds to the weakest (worst) level and Level 5 corresponds to the strongest (best) level.

Table 3.8. Level of Resolution Method Knowledge –Levels & Frequencies

Attr. ID	**Attribute**	**Levels**	**Frequency in the Dataset**	
			Resolution Method	
K1	Level of Knowledge on Litigation	Level 1	4	7.4%
		Level 2	6	11.1%
		Level 3	9	16.7%
		Level 4	18	33.3%
		Level 5	17	31.5%
K2	Level of Knowledge on Arbitration	Level 1	6	11.1%
		Level 2	7	13.0%
		Level 3	10	18.5%
		Level 4	21	38.9%
		Level 5	10	18.5%
K3	Level of Knowledge on DRB	Level 1	14	25.9%
		Level 2	3	5.6%
		Level 3	11	20.4%
		Level 4	11	20.4%
		Level 5	15	27.8%
K4	Level of Knowledge on Mediation	Level 1	3	5.6%
		Level 2	3	5.6%
		Level 3	10	18.5%
		Level 4	17	31.5%
		Level 5	21	38.9%
K5	Level of Knowledge on SEA	Level 1	2	3.7%
		Level 2	3	5.6%
		Level 3	6	11.1%
		Level 4	11	20.4%
		Level 5	32	59.3%
K6	Level of Knowledge on Negotiation	Level 1	0	0.0%
		Level 2	1	1.9%
		Level 3	3	5.6%
		Level 4	22	40.7%
		Level 5	28	51.8%

The average level of knowledge of participants on resolution methods are given in Figure 3.16. According to average values, participants are least familiar with the DRB method and most familiar with the negotiation technique. As they tend to perceive

SEA technique as a form of negotiation that is performed with the top-level management and owners, the second most familiar technique is SEA. Participants' familiarity to mediation can said to be moderate. Leaving ADR techniques aside, although traditional methods (litigation and arbitration) exist in the construction industry for quite long time, the level of knowledge about them can said to be low.

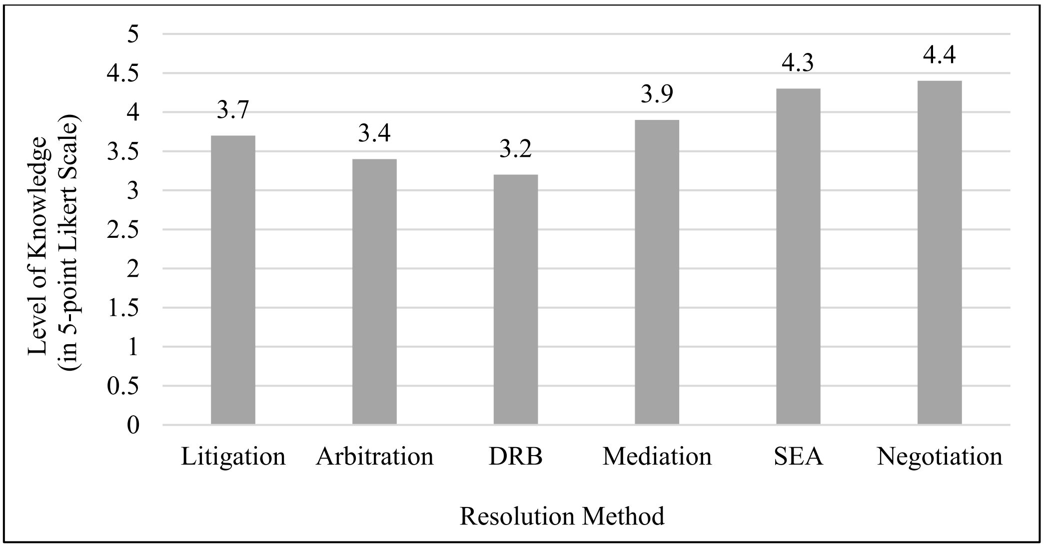


Figure 3.16. Participants' Average Knowledge Levels on Resolution Methods

Subsequent to collecting empirical data for development of prediction models, the next objective is to determine which input variables affect output variables of prediction models. For this purpose, the data is analyzed in terms of significance. In other words, the significance of associations between input and output variables are analyzed for all prediction models. The insignificant input variables are removed from the original conceptual model and three different prediction models are established with fewer input variables. In this research, the Chi-Square tests of association is utilized for attribute elimination and details will be given in the next section.

3.3. THE CHI-SQUARE TESTS ON DATASETS

The performance of ML algorithms is generally affected negatively by irrelevant or insignificant attributes (Pulket and Arditi, 2009b). Therefore, elimination of insignificant attributes and selection of the ones impacting the model outcomes improve generalization performance of ML algorithms (Arditi and Pulket, 2009; Sönmez and Sözgen, 2017). Such an elimination is known as attribute/variable/feature selection/elimination in practice and it helps achieving better algorithm generalization (Drucker et al., 1999). Thus, attribute selection has become an important research area with the following benefits; (1) enhanced prediction capability, (2) faster solutions, (3) improved data visualization and understanding, (4) reduced measurement, data collection, and storage requirements, (5) lower computational costs via reduced training and utilization times, and (6) avoiding problems caused by data dimensionality (Guyon and Elisseeff, 2003).

Among numerous attribute selection methods in the literature, the Chi-Square statistics is preferred in this research. In the literature, the Chi-Square test of independence (Pearson Chi-Square test) is known as one of the most effective methods in testing the hypothesis between two categorical variables (McHugh, 2013). In other words, the Chi-Square statistics is a useful way of testing the existence of a statistically significant relationship between categorical variables. Moreover, the Chi-Square statistics is a non-parametric method that is robust to distribution of the data and unequal variances among study groups (Weisburd and Britt, 2007). This means the Chi-Square results can compensate the problematic issues due to data distribution (i.e. skewed data) unlike many other methods that require data with almost normal distribution and equality of variances. In addition, the method can handle both dichotomous variables and variables with multiple categories (McHugh, 2013). Considering that, the dataset in this research is composed of dichotomous and multiple categorical input and output variables with various distributions, the Chi-Square statistics is an appropriate evaluation technique for attribute elimination. As previously given in Section 3.2.3, input variables in this research are either categorical (nominal or ordinal) or numeric. However, numeric attributes are further processed according to certain predefined threshold values and they are converted to ordinal categorical variables so that the dataset can be analyzed by the Chi-Square tests. In addition, dispute occurrence as an output variable is a categorical one (0 = undisputed projects and 1 = disputed projects). Similarly, potential compensation (1 = no compensation, 2 = cost compensation only, 3 = time compensation only, 4 = both cost and time compensation) and resolution method as output variables are also categorical (1 = litigation, 2 = arbitration, 3 = DRB, 4 = mediation, 5 = SEA, 6 = negotiation).

At this point, it might be convenient to explain categorical variables in detail. Categorical variables can be divided into two groups based on their measurement scales as (1) nominal variables and (2) ordinal variables. Categorical variables with an ordered scale is called ordinal variables and results of statistical analysis depend on that order. However, for nominal variables, the order of categories is not important and it does not change results of statistical analysis. Similarly, statistical methods for nominal variables can be used with nominal and ordinal variables as the only requirement is to have categories. However, when used with ordinal variables, the information related to the ordering is lost. Thus, it is better to apply nominal statistical methods to nominal variables only. In addition, methods for ordinal variables can only be used with ordinal variables as they require an ordered scale (Agresti, 2007). Therefore, in the Chi-Square statistical tests, nominal and ordinal variables should be analyzed accordingly.

Besides mentioned advantages of the Chi-Square tests, there are also limitations and assumptions of this statistics. Firstly, as mentioned earlier, the Chi-Square statistics can only be used with categorical (nominal and ordinal) variables. Secondly, the Chi-Square tests are based on data counts or frequencies only and they cannot work with percentages or other forms of data. Thirdly, categories of variables should be mutually exclusive with each instance contributing to one category only. In addition, large number of categories for a variable should be avoided due to difficulties in interpretation of results. Large number of categories corresponds to 20 or more generally (McHugh, 2013). Finally, as a rule of thumb, the expected value in a category cannot be less than ‘5’ in more than 80% of the categories. In addition, an expected value cannot be less than ‘1’. Consequently, the sample size should be compatible with these limitations (Mehta and Patel, 2012).

In short, advantages of utilizing the Chi-Square tests can be listed as (1) robustness against data distribution, unequal sizes and variances, (2) capability of handling dichotomous and multiple category variables, and (3) computational ease (McHugh, 2013). Therefore, the Chi-Square tests are used for evaluating the statistical significance in this research. According to results of the Chi-Square tests, the insignificant attributes on outcomes will be eliminated and prediction models will be developed using the remaining significant attributes. Instead of explaining technical details and equations related to the Chi-Square statistics, the attention will be on performing the analysis appropriately by focusing on important points such as how to use the technique appropriately for nominal and ordinal variables, how to determine the strength of association between variables, etc.

3.3.1. Strength of Relationships in the Chi-Square Tests

The Chi-Square test of independence reveals the significance of dependence between categorical independent (input) and dependent (output) variables. However, it does not reveal the strength of the relationship (Agresti, 2007). Consequently, it should be followed with a statistic showing the strength of the relationship (McHugh, 2013). For this purpose, numerous measures of strength are available in the literature to help evaluating the strength of the relationship between independent and dependent variables. Although there are strength measures based on the Chi-Square value directly, many other measures do not use this value and just transform results and frequencies to interpret the strength (Weisburd and Britt, 2007). Regardless of the measure, for nominal variables, the strength measures output a value between “0” and “1”, where “0” represents no relationship and “1” represents perfect relationship. The higher this value is, the stronger the relationship between two nominal variables. For ordinal variables, the strength measure is between “-1” and “+1”, where “-1” represents the perfect negative relationship, “0” represents no relationship, and “+1” represents the perfect positive relationship (Mehta and Patel, 2012).

Phi and Cramer’s V are two of the measures for evaluating the strength of association between nominal variables. Phi measure transforms the Chi-Square value according to the sample size to calculate a strength value only for 2x2 tables (Weisburd and Britt, 2007). In a 2x2 table, both independent and dependent variables on rows and columns of a table should have two possible categories. However, the real world data and variables generally do not suit this kind of limitation. For this reason, there is another measure called Cramer’s V, which has a capability of handling tables with varying number of rows and columns (Weisburd and Britt, 2007). Therefore, although Cramer’s V can calculate low correlation values for highly significant results, it became the most commonly preferred strength test for nominal variables (McHugh, 2013). Consequently, in order to handle the changing number of rows and columns between variables of the dataset in this thesis study, Cramer’s V is preferred.

As stated earlier, methods for nominal and ordinal variables should not be applied to each other. Therefore, besides Cramer’s V that is used in strength tests for nominal attributes, measures for ordinal attributes should also be investigated. IBM SPSS Statistics version 22.0 that is used to perform the Chi-Square tests presents four measures of association for ordinal variables. These are Gamma, Somers’ d, Kendall’s tau-b, and Kendall’s tau-c.

Strength values obtained from Gamma, Somers’ d, and Kendall’s tau measures are generally different from each other due to differences in handling tied pairs of observations. In Gamma measure, tied pairs of observations are not considered and consequently, there is a problem of overestimating the strength of the relationship between two ordinal variables. On the other hand, Somers’ d takes tied pairs of observations into account for independent variables only and Kendall’s tau measures do it for both independent and dependent variables. Thus, Somers’ d and Kendall’s measures are superior to Gamma measure. However, Kendall's tau-b is more adequate when the number of columns and rows, or in other words, the number of categories of independent and dependent variables are equal. For unequal number of rows and columns, Kendall's tau-c is more adequate. When the number of rows and columns are not equal, Somers' d can be considered as a better measure than Kendall's tau-c. Somers' d is superior when the independent and dependent variables are clearly defined (Weisburd and Britt, 2007). In the light of these, Somers' d will be preferred as a strength measure of association for ordinal variables.

Table 3.9. Summary of Reviewed Measures of Strength

Measure of Association	Scale of Measurement	Dimensions (Rows x Columns)	Preferred Measure in this Research
Phi	Nominal	2 x 2	No
Cramer's V	Nominal	Any Size	**Yes**
Gamma	Ordinal	Any Size	No
Kendall's tau b	Ordinal	Number of rows = Number of columns	No
Kendall's tau c	Ordinal	Number of rows ≠ Number of columns	No
Somer's d	Ordinal	Any Size	**Yes**

3.3.2. Results of the Chi-Square Tests

The Chi-Square tests are performed in IBM SPSS Statistics version 22.0 using the Crosstabs menu under the Analyze section.

Before giving results of the Chi-Square tests, basic examples will be given and explained in detail to highlight what have been done for various cases during the tests. The first example is based on the attribute "PC1 – Project Location" as input and "O3 – Resolution Method" as output. Both attributes are nominal categorical attributes. The ordering of categories does not change results of the analysis. Thus, the impact of "PC1 – Project Location" on "O3 – Resolution Method" is tested via nominal Chi-Square methods. The SPSS output is given in Table 3.10.

The upper section of the SPSS output in Table 3.10 is a cross-tabulation of frequencies tabulating the outcome of resolution method by project location, which is called a contingency table.

Table 3.10. SPSS Output for the Chi-Square Test of “Project Location” and “Resolution Method” Attributes

ProjectLocation * ResolutionMethod Crosstabulation

Count

		ResolutionMethod						Total
		Litigation	Arbitration	DRB	Mediation	SEA	Negotiation	
ProjectLocation	Domestic	7	2	3	4	3	11	30
	International	2	4	2	1	7	8	24
Total		9	6	5	5	10	19	54

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	6,937[a]	5	,225	,236		
Likelihood Ratio	7,204	5	,206	,270		
Fisher's Exact Test	6,644			,251		
Linear-by-Linear Association	,554[b]	1	,457	,479	,252	,043
N of Valid Cases	54					

a. 8 cells (66.7%) have expected count less than 5. The minimum expected count is 2.22.

b. The standardized statistic is .744.

The lower section of the SPSS output shows results of the Chi-Square tests. The null hypothesis in this test is that there is no relationship between project location and resolution method selection. On the other hand, the alternative hypothesis is that there is an association between project location and resolution method selection. Below the results of the Chi-Square tests, there are two explanations. The first explanation states, “8 cells (66.7%) have expected count less than ‘5’. The minimum expected count is 2.22”. Remembering limitations of the Chi-Square tests, there were two rules of thumb; one stating that the minimum expected value in a cell should be at least ‘5’ in 80% of the categories (cells) and the other one stating that the expected count cannot be equal to ‘0’ in any category (Mehta and Patel, 2012). Thus, there is a contradiction in the SPSS output as in 66.7% (>20%) of the categories, the expected count is below ‘5’. Normally, in the Chi-Square tests, asymptotic probability values are the considered values. However, the expected count assumption is violated and asymptotic results cannot be used. In such cases, by using the true sampling distribution of the Chi-Square, an exact probability value should be calculated (Mehta and Patel, 2012). SPSS contains several exact tests such as the exact Pearson Chi-Square statistics and the Fisher's Exact Test. However, the Fisher's Exact Test is designed for 2x2 contingency tables only (Mehta and Patel, 2012). Therefore, in cases that violates the minimum expected value assumption with a contingency table larger than 2x2, the exact Pearson Chi-Square statistic should be calculated (Bal et al., 2009). For this example, the probability value obtained from the exact Pearson Chi-Square test is equal to '0.236'. This value is higher than the alpha level 0.05 for 95% confidence level and it is not statistically significant. Therefore, we should accept the null hypothesis stating that there is no relationship between project location and resolution method selection for this dataset. Consequently, "PC1 – Project Location" attribute is eliminated from the resolution method selection model. If project location were a significant attribute, the strength of relationship would be calculated by the nominal strength measure of Cramer's V.

The second example is based on "PC2 – Project or Contract Value" as input and "O1 – Dispute Occurrence" as output. "PC2 – Project or Contract Value" is a numeric attribute normally. However, it is converted to a categorical variable using predefined threshold values. The ordering of categories for this attribute is important as they indicate an order. Simply, the contract value is an ordinal attribute. The ordering changes results of the analysis and consequently, it should be considered. For ordinal variables, SPSS performs the Mantel-Haenszel Linear-by-Linear Association Chi-Square Test, which calculates a probability value with one degree of freedom. The null hypothesis here states that there is no relationship between contract value and dispute occurrence, while the alternative hypothesis is that these two variables are associated. Thus, the impact of "PC2 – Project or Contract Value" on "O1 – Dispute Occurrence" is tested via ordinal Chi-Square methods. The SPSS output is given in Table 3.11.

Table 3.11. SPSS Output for the Chi-Square Test of "Project or Contract Value" and "Dispute Occurrence" Attributes

ContractValue * DisputeOccurrence Crosstabulation

Count

		DisputeOccurrence		Total
		Undisputed	Disputed	
ContractValue	Less than 10m	23	21	44
	10-100m	11	24	35
	More than 100m	4	25	29
Total		38	70	108

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	11,669[a]	2	,003	,003		
Likelihood Ratio	12,345	2	,002	,003		
Fisher's Exact Test	11,793			,003		
Linear-by-Linear Association	11,535[b]	1	,001	,001	,000	,000
N of Valid Cases	108					

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.20.

b. The standardized statistic is 3.396.

Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.	Exact Sig.
Ordinal by Ordinal	Somers' d	Symmetric	,305	,079	3,761	,000	,001
		ContractValue Dependent	,373	,097	3,761	,000	,001
		DisputeOccurrence Dependent	,259	,069	3,761	,000	,001

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The upper section of the SPSS output is the contingency table for the tested attributes and the middle section shows results of the Chi-Square tests. This time "0 cells (0%) have expected count less than '5'. The minimum expected count is 10.20". Therefore, the assumption is not violated and asymptotic results can be used. However, in order to achieve more accurate results, this research will use exact results whenever they can be calculated. The probability value obtained from the Mantel-Haenszel Linear-by-Linear Association Test is equal to '0.001'. This value is smaller than the alpha level at 0.05 for 95% confidence interval and it is statistically significant. Therefore, null hypothesis is rejected and alternative hypothesis that associates contract value with dispute occurrence is accepted. Therefore, "PC2 – Project or Contract Value" is a significant attribute that should be kept in the final prediction model for dispute occurrence.

The analysis for these two attributes has not been completed yet. Since there is a relationship between contract value and dispute occurrence, the strength of this relationship should be determined. As discussed in Section 3.3.1, Somers' d measure of association will be preferred for ordinal variables. The lower section in Table 3.11 shows the Somers' d value, which is equal to '0.259' when the dependent variable is dispute occurrence. This value indicates a moderately strong relationship between contract value and dispute occurrence in a positive direction.

So far, an example on the Chi-Square test for nominal variables that violates the expected count assumption and an example for an ordinal variable that does not violate the expected count assumption have been reviewed. In both examples, it was possible to perform exact tests and reasons of using exact results are explained. In addition, the attribute in nominal case was insignificant that required no more action, while the attribute in ordinal case was significant that required a strength of relationship test. Thus, calculation of the strength of relationship in an ordinal variable is explained.

When contingency tables have too many rows and columns, calculations become costly, if not impossible, with current technologies. In such cases, SPSS resorts to the Monte Carlo method that calculates an unbiased estimate of the exact probability value (the Monte Carlo probability value) in a very short duration. Although SPSS calculates 99% accurate estimates by default, the user can increase or decrease the accuracy by sampling more or less outcomes from the reference set. The Monte Carlo algorithm is useful when exact results cannot be computed due to size of the dataset or when asymptotic results are not reliable. The Monte Carlo estimates are calculated within a specified confidence interval guaranteeing to contain the exact probability value at that level (Mehta and Patel, 2012). The Monte Carlo estimates are within 99% confidence interval based on 10,000 sampled tables with starting seed 2,000,000. In this research, the Monte Carlo method is always used with this configuration.

3.3.2.1. Chi-Square Results for Dispute Occurrence Prediction Model

Following details of the Chi-Square calculations, results of the tests can now be given. Firstly, input variables will be analyzed for dispute occurrence (as the output). In other words, attributes identified in Table 2.10 (attributes in the conceptual model) are tested for dispute occurrence prediction model. The Chi-Square test results of project characteristics attributes with respect to dispute occurrence are given in Table 3.12. In this table, exact probability values are obtained from the exact Pearson Chi-Square statistics for nominal attributes and the Mantel-Haenszel Linear-by-Linear Association Test for ordinal ones. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. In addition, if a statistically significant association is discovered, the strength of this association is determined by the Cramer's V measure for nominal attributes and the Somers' d measure for ordinal attributes.

Table 3.12. The Chi-Square Test Results of Project Characteristics Attributes for Dispute Occurrence Prediction Model

Attributes & Categories	p-value	Dispute Occurred (%)	Selected for Final Model	Strength of Association
PC1 – Project Location	0.037		**YES**	Cramer's V
Domestic		67.1%		0.215
International		32.9%		
PC2 – Project or Contract Value	0.003		**YES**	Somers' d
< 10 million $		30.0%		0.259
10 – 100 million $		34.3%		
> 100 million $		35.7%		
PC3 – Planned Project Duration	0.000		**YES**	Somers' d
< 1 year		18.6%		0.286
1 - 2 years		32.9%		
2 - 3 years		21.4%		
> 3 years		27.1%		
PC4 – Type of Construction	0.157		NO	-
Housing		21.4%		
Commercial		11.4%		
Industrial		10.0%		
Transportation		18.6%		
Pow.Plants&Lines		4.3%		
WaterSupp.&Reser.		10.0%		
Sport&Cult.&Edu.		10.0%		
Medical		4.3%		
Public		5.7%		
Soil Works		4.3%		

Table 3.12. The Chi-Square Test Results of Project Characteristics Attributes for Dispute Occurrence Prediction Model (Continued)

Attributes & Categories	p-value	Dispute Occurred (%)	Selected for Final Model	Strength of Association
PC5 – Type of Contractor	0.749		NO	-
Single		81.4%		
Joint Venture		11.4%		
Consortium		7.1%		
PC6 – Type of Employer	0.961		NO	-
Public		47.1%		
Private		40.0%		
PPP		12.9%		
PC7 – Type of Contract	0.074		NO	-
Private Contracts		52.8%		
Public Procurement		25.7%		
FIDIC Red		12.9%		
FIDIC Silver/Yellow		8.6%		
PC8 – Payment Method	0.842		NO	-
Fixed (Lump-Sum)		52.9%		
Unit Price		47.1%		
PC9 – Project Delivery System	0.957		NO	-
DBB		62.9%		
DB		22.9%		
EPC		14.3%		
PC10 – Level of Design Comp.	0.938		NO	-
Very Low		14.3%		
Low		11.4%		
Moderate		18.6%		
High		34.3%		
Very High		21.4%		
PC11 – Level of Constr. Comp.	1.000		NO	-
Very Low		11.4%		
Low		10.0%		
Moderate		17.1%		
High		35.7%		
Very High		25.7%		

The Chi-Square test results of skills attributes with respect to dispute occurrence are given in Table 3.13. All attributes in skills category are ordinal categorical variables. Therefore, exact probability values are obtained from the Mantel-Haenszel Linear-by-Linear Association Test and they are compared to the alpha level at 0.05 for 95% confidence interval. The statistically significant attributes (attributes that are associated with each other) are selected for the final dispute occurrence prediction model. Since skills attributes are ordinal variables, the Somers' d measure is used for determining the strength of association.

Table 3.13. The Chi-Square Test Results of Skills Attributes for Dispute Occurrence Prediction Model

Attributes	p-value	Selected for Final Model	Strength of Association
S1 – Relationship between Parties / Individuals	0.000	**YES**	Somers' d -0.406
S2 – Previous Experience with Each Other or Reputation	0.007	**YES**	Somers' d -0.185
S3 – Dispute Avoidance Incentives	0.158	NO	-
S4 – Communication between Parties	0.000	**YES**	Somers' d -0.370
S5-1 – Working Culture & Skills of Represented Party	0.012	**YES**	Somers' d -0.162
S5-2 – Working Culture & Skills of Counter Party	0.000	**YES**	Somers' d -0.303
S6-1 – Response Rate & Communication Skills of Represented Party	0.228	NO	-
S6-2 – Response Rate & Communication Skills of Counter Party	0.000	**YES**	Somers' d -0.280
S7-1 – Experience of Represented Party	0.085	NO	-
S7-2 – Experience of Counter Party	0.001	**YES**	Somers' d -0.233
S8-1 – Project Management & Coordination Skills of Represented Party	0.006	**YES**	Somers' d -0.199
S8-2 – Project Management & Coordination Skills of Counter Party	0.000	**YES**	Somers' d -0.321

The Chi-Square test result of changes attribute with respect to dispute occurrence is given in Table 3.14. Both attributes are nominal categorical attributes. Therefore, the exact probability value is obtained from the exact Pearson Chi-Square statistics and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between changes and dispute occurrence is statistically significant (p-value = 0.000 < 0.05). As both attributes are nominal categorical variables, the Cramer's V measure is used for determining the strength of association.

Table 3.14. The Chi-Square Test Result of Changes Attribute for Dispute Occurrence Prediction Model

Attributes & Categories	p-value	Dispute Occurred (%)	Selected for Final Model	Strength of Association
C1 – Changes	0.000		**YES**	Cramer's V
Yes		58.6%		0.576
No		41.4%		

The Chi-Square test result of delays attribute with respect to dispute occurrence is given in Table 3.15. Although the delays attribute is a numeric attribute normally, it is converted into an ordinal categorical variable using predefined threshold values. Therefore, the exact probability value is obtained from the Mantel-Haenszel Linear-by-Linear Association Test and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between delays and dispute occurrence is statistically significant (p-value = 0.002 < 0.05). The Somers' d measure is used for determining the strength of association.

Table 3.15. The Chi-Square Test Result of Delays Attribute for Dispute Occurrence Prediction Model

Attributes & Categories	p-value	Dispute Occurred (%)	Selected for Final Model	Strength of Association
D1 – Delays	0.002		**YES**	Somers' d
Ratio = 0%		30.0%		0.232
Ratio 0% - 20%		21.4%		
Ratio 20% 40%		22.9%		
Ratio > 40%		25.7%		

3.3.2.2. Chi-Square Results for Potential Compensation Model

Following the analysis for dispute occurrence prediction model, input variables will be analyzed for potential compensation type (as the output) this time. In other words, attributes identified in Table 2.11 (attributes in the conceptual model) are tested for potential compensation prediction model. The Chi-Square test results of project characteristics attributes with respect to potential compensation type are given in Table 3.16. The exact probability values and the strength of association are determined using the same methodology in dispute occurrence prediction model and values are compared to the alpha level 0.05 for 95% confidence interval again.

Table 3.16. The Chi-Square Test Results of Project Characteristics Attributes for Potential Compensation Prediction Model

		Compensation Type					
Attributes & Categories	**p-value**	**No Comp.**	**Cost Only**	**Time Only**	**Cost & Time**	**Selected in Final Model**	**Str. of Assoc.**
PC1 – Project Location	0.068					NO	-
Domestic		41.7%	73.7%	80.0%	48.1%		
International		58.3%	26.3%	20.0%	51.9%		
PC2 – Project or Contract Value	0.291					NO	-
< 10 million $		25.0%	31.6%	20.0%	18.5%		
10 – 100 million $		16.7%	39.5%	60.0%	25.9%		
> 100 million $		58.3%	28.9%	20.0%	55.6%		
PC3 – Planned Project Duration	0.716					NO	-
< 1 year		8.3%	18.4%	0.0%	18.5%		
1 - 2 years		25.0%	34.2%	60.0%	22.2%		
2 - 3 years		8.3%	31.6%	0.0%	11.1%		
> 3 years		58.3%	15.8%	40.0%	48.2%		
PC4 – Type of Construction	0.237					NO	-
Housing		0.0%	34.2%	20.0%	7.4%		
Commercial		33.3%	10.5%	0.0%	18.5%		
Industrial		8.3%	10.5%	0.0%	7.4%		
Transportation		16.7%	15.8%	0.0%	29.6%		
Pow.Plants & Lines		0.0%	5.3%	0.0%	3.7%		
WaterSupp. & Reser.		25.0%	5.3%	40.0%	7.4%		
Sport&Cult. & Edu.		16.7%	5.3%	20.0%	11.1%		
Medical		0.0%	5.3%	0.0%	3.7%		
Public		0.0%	2.6%	20.0%	7.4%		
Soil Works		0.0%	5.3%	0.0%	3.7%		

Table 3.16. The Chi-Square Test Results of Project Characteristics Attributes for Potential Compensation Prediction Model (Continued)

Attributes & Categories	p-value	Compensation Type: No Comp.	Cost Only	Time Only	Cost & Time	Selected in Final Model	Str. of Assoc.
PC5 – Type of Contractor	0.010					**YES**	Cramer'sV
Single		100.0%	84.2%	60.0%	66.7%		0.327
Joint Venture		0.0%	15.8%	0.0%	18.5%		
Consortium		0.0%	0.0%	40.0%	14.8%		
PC6 – Type of Employer	0.075					NO	-
Public		58.3%	39.5%	100.0%	55.6%		
Private		41.7%	50.0%	0.0%	25.9%		
PPP		0.0%	10.5%	0.0%	18.5%		
PC7 – Type of Contract	0.010					**YES**	Cramer'sV
Private Contracts		50.0%	68.4%	0.0%	40.7%		0.297
Public Procurement		25.0%	18.4%	80.0%	14.8%		
FIDIC Red		25.0%	7.9%	20.0%	29.6%		
FIDIC Silv./Yellow		0.0%	5.3%	0.0%	14.8%		
PC8 – Payment Method	0.335					NO	-
Fixed (Lump-Sum)		66.7%	44.7%	80.0%	55.6%		
Unit Price		33.3%	55.3%	20.0%	44.4%		
PC9 – Project Delivery Syst.	0.680					NO	-
DBB		58.3%	57.9%	100.0%	55.6%		
DB		33.3%	26.3%	0.0%	29.6%		
EPC		8.3%	15.8%	0.0%	14.8%		
PC10 – Lvl of Dsgn. Comp.	0.689					NO	-
Very Low		16.7%	18.4%	20.0%	14.8%		
Low		8.3%	13.2%	0.0%	7.4%		
Moderate		8.3%	18.4%	40.0%	14.8%		
High		25.0%	39.5%	0.0%	37.0%		
Very High		41.7%	10.5%	40.0%	25.9%		
PC11 – Lvl of Constr Comp.	0.275					NO	-
Very Low		16.7%	13.2%	0.0%	7.4%		
Low		8.3%	13.2%	0.0%	11.1%		
Moderate		0.0%	23.7%	20.0%	11.1%		
High		33.3%	39.5%	20.0%	37.0%		
Very High		41.7%	10.5%	60.0%	33.3%		

The Chi-Square test results of skills attributes with respect to potential compensation type are given in Table 3.17. Since attributes in skills category are ordinal categorical variables, exact probability values are obtained from the Mantel-Haenszel Linear-by-Linear Association Test. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. However, none of the skills attributes managed to has a place in the final prediction model for potential compensation type.

Table 3.17. The Chi-Square Test Results of Skills Attributes for Potential Compensation Prediction Model

Attributes	p-value	Selected for Final Model	Strength of Association
S1 – Relationship between Parties / Individuals	0.082	NO	-
S2 – Previous Experience with Each Other or Reputation	0.647	NO	-
S3 – Dispute Avoidance Incentives	0.417	NO	-
S4 – Communication between Parties	0.390	NO	-
S5-1 – Working Culture & Skills of Represented Party	0.280	NO	-
S5-2 – Working Culture & Skills of Counter Party	0.443	NO	-
S6-1 – Response Rate & Communication Skills of Represented Party	0.193	NO	-
S6-2 – Response Rate & Communication Skills of Counter Party	0.105	NO	-
S7-1 – Experience of Represented Party	0.061	NO	-
S7-2 – Experience of Counter Party	0.562	NO	-
S8-1 – Project Management & Coordination Skills of Represented Party	0.160	NO	-
S8-2 – Project Management & Coordination Skills of Counter Party	0.795	NO	-

The Chi-Square test result of changes attribute with respect to potential compensation type is given in Table 3.18. Both attributes are nominal categorical attributes. Therefore, the exact probability value is obtained from the exact Pearson Chi-Square statistics and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between changes and potential compensation type is statistically significant (p-value = 0.000 < 0.05). Thus, changes attribute should be kept as an input variable in the final prediction model for potential compensation type. As both attributes are nominal categorical variables, the Cramer's V measure is used for determining the strength of association.

Table 3.18. The Chi-Square Test Result of Changes Attribute for Potential Compensation Prediction Model

		Compensation Type					
Attributes & Categories	**p-value**	**No Comp.**	**Cost Only**	**Time Only**	**Cost & Time**	**Selected in Final Model**	**Str. of Assoc.**
C1 – Changes	0.000					**YES**	Cramer'sV
Yes		91.7%	65.8%	60.0%	92.6%		0.585
No		8.3%	34.2%	40.0%	7.4%		

The Chi-Square test result of delays attribute with respect to potential compensation type is given in Table 3.19. Although the delays attribute is a numeric attribute normally, it is converted into an ordinal categorical variable using predefined threshold values. Therefore, the exact probability value is obtained from the Mantel-Haenszel Linear-by-Linear Association Test and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between delays and potential compensation type is statistically significant (p-value = 0.000 < 0.05). Thus, delays attribute should be kept as an input variable in the final prediction model. The Somers' d measure is used for determining the strength of association.

Table 3.19. The Chi-Square Test Result of Delays Attribute for Potential Compensation Prediction Model

		Compensation Type					
Attributes & Categories	**p-value**	**No Comp.**	**Cost Only**	**Time Only**	**Cost & Time**	**Selected in Final Model**	**Str. of Assoc.**
D1 – Delays	0.000					**YES**	Somers'd
Ratio = 0%		16.7%	50.0%	0.0%	0.0%		0.232
Ratio 0% - 20%		58.3%	7.9%	40.0%	33.3%		
Ratio 20% 40%		16.7%	21.1%	40.0%	18.5%		
Ratio > 40%		8.3%	21.1%	20.0%	48.2%		

The Chi-Square test results of dispute characteristics attributes with respect to potential compensation type are given in Table 3.20. In this table, exact probability values are obtained from the exact Pearson Chi-Square statistics for nominal attributes and the Mantel-Haenszel Linear-by-Linear Association Test for ordinal ones. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. In addition, if a statistically significant association is discovered, the strength of this association is determined by the Cramer's V measure for nominal attributes and the Somers' d measure for ordinal attributes.

Table 3.20. The Chi-Square Test Results of Dispute Characteristics Attributes for Potential Compensation Prediction Model

Attributes & Categories	p-value	Compensation Type: No Comp.	Cost Only	Time Only	Cost & Time	Selected in Final Model	Str. of Assoc.
DC1 – Disputant Party	0.017					**YES**	Cramer'sV
Client		91.7%	73.7%	100.0%	100.0%		0.361
Contractor		8.3%	26.3%	0.0%	0.0%		
DC2 – Phase of Occurrence	0.069					NO	-
Plan. & Design		0.0%	2.6%	20.0%	0.0%		
Construction		83.3%	76.3%	80.0%	96.3%		
Transfer & Repair		16.7%	21.1%	0.0%	3.7%		
DC3 – Dispute Sources	0.000					**YES**	Cramer'sV
Cost compensation of change orders		25.0%	10.5%	0.0%	0.0%		0.584
Time & Cost compensation of change orders		8.3%	2.6%	40.0%	63.0%		
Measurement & valuation of contracted works		0.0%	15.8%	0.0%	0.0%		
Extended overhead due to extensions		0.0%	2.6%	0.0%	0.0%		
Delay in site handover & possession		0.0%	0.0%	0.0%	14.8%		
Defects, errors & quality		8.3%	15.8%	20.0%	0.0%		
Contractor fails to act as a prudent merchant		0.0%	15.8%	20.0%	3.7%		
Delays in payments		0.0%	18.4%	0.0%	0.0%		
Errors or substantial changes in BoQ		0.0%	10.5%	0.0%	0.0%		
Inadequate site or soil investigation		25.0%	2.6%	0.0%	14.8%		
Interpretation of contract clauses		33.3%	5.3%	20.0%	3.7%		
DC4 – Suspension of works	0.622					NO	-
Yes		41.7%	28.9%	40.0%	44.4%		
No		58.3%	71.1%	60.0%	55.6%		

Table 3.20. The Chi-Square Test Results of Dispute Characteristics Attributes for Potential Compensation Prediction Model (Continued)

Attributes & Categories	p-value	Compensation Type: No Comp.	Cost Only	Time Only	Cost & Time	Selected in Final Model	Str. of Assoc.
DC5 – Disputed Amount	0.019					**YES**	Somers' d
< 5 million $		50.0%	65.8%	40.0%	22.2%		0.248
5 - 25 million $		16.7%	21.1%	60.0%	33.3%		
25 - 75 million $		0.0%	10.5%	0.0%	18.5%		
> 75 million $		33.3%	2.6%	0.0%	25.9%		
DC8 – Presen. of EoT Claim	0.000					**YES**	Cramer'sV
Yes		25.0%	2.6%	0.0%	0.0%		0.917
No		75.0%	97.4	100.0%	100.0%		
DC9 – Disp. EoT Amount	0.000					**YES**	Somers' d
0 days		75.0%	97.4%	0.0%	0.0%		0.659
0 - 6 months		0.0%	2.6%	40.0%	37.0%		
6 months - 1 year		8.3%	0.0%	60.0%	25.9%		
> 1 year		16.7%	0.0%	0.0%	37.0%		

3.3.2.3. Chi-Square Results for Resolution Method Selection Model

Following the analysis for dispute occurrence and potential compensation type prediction models, input variables will be analyzed for resolution method (as the output) finally. In other words, attributes identified in Table 2.12 (attributes in the conceptual model) are tested for resolution method selection model.

The Chi-Square test results of project characteristics attributes with respect to resolution methods are given in Table 3.21. The exact probability values and the strength of association are determined using the same methodology as in previous prediction models and values are compared to the alpha level 0.05 for 95% confidence interval again. In other words, exact probability values are obtained from the exact Pearson Chi-Square statistics for nominal attributes and the Mantel-Haenszel Linear-by-Linear Association Test for ordinal attributes. The only project characteristics related attribute that is found to be associated with resolution method selection is "PC5 – Type of Contractor", which is a nominal categorical attribute. Thus, the Cramer's V measure is used for determining the strength of association.

Table 3.21. The Chi-Square Test Results of Project Characteristics Attributes for Resolution Method Selection Model

Attributes & Categories	p-value	Resolution Method						Selected in Final Model	Str. of Assoc.
		LIT	ARB	DRB	MED	SEA	NEG		
PC1 - Prj. Location	0.236							NO	-
Domestic		77.8%	33.3%	60.0%	80.0%	30.0%	57.9%		
International		22.2%	66.7%	40.0%	20.0%	70.0%	42.1%		
PC2 - Contr. Value	0.349							NO	-
< 10 million $		33.3%	0.0%	0.0%	100.0%	10.0%	21.1%		
10-100 million $		22.2%	16.7%	40.0%	0.0%	40.0%	52.6%		
> 100 million $		44.4%	83.3%	60.0%	0.0%	50.0%	26.3%		
PC3 - Plan.Prj.Dur.	0.221							NO	-
< 1 year		33.3%	0.0%	0.0%	40.0%	20.0%	5.3%		
1 - 2 years		33.3%	16.7%	20.0%	60.0%	30.0%	31.6%		
2 - 3 years		33.3%	16.7%	80.0%	0.0%	10.0%	26.3%		
> 3 years		0.0%	66.7%	0.0%	0.0%	40.0%	36.8%		
PC4 - Const. Type	0.131							NO	-
Housing		44.4%	16.7%	20.0%	0.0%	0.0%	26.3%		
Commercial		0.0%	16.7%	0.0%	0.0%	20.0%	15.8%		
Industrial		22.2%	16.7%	0.0%	20.0%	20.0%	0.0%		
Transportation		11.1%	33.3%	60.0%	0.0%	40.0%	15.8%		
Pow.Plants&Line		0.0%	0.0%	0.0%	40.0%	0.0%	0.0%		
WaterSup.&Res.		0.0%	16.7%	0.0%	0.0%	0.0%	15.8%		
Sprt&Cult.&Edu.		11.1%	0.0%	0.0%	20.0%	10.0%	10.5%		
Medical		0.0%	0.0%	20.0%	0.0%	10.0%	5.3%		
Public		11.1%	0.0%	0.0%	20.0%	0.0%	5.3%		
Soil Works		0.0%	0.0%	0.0%	0.0%	0.0%	5.3%		
PC5 - Contractor	0.003							**YES**	Cramer'sV
Single		88.9%	83.3%	20.0%	80.0%	80.0%	89.5%		0.514
Joint Venture		0.0%	0.0%	80.0%	20.0%	0.0%	10.5%		
Consortium		11.1%	16.7%	0.0%	0.0%	20.0%	0.0%		
PC6 - Employ.Typ.	0.581							NO	-
Public		44.4%	50.0%	80.0%	20.0%	40.0%	47.4%		
Private		55.6%	16.7%	20.0%	60.0%	50.0%	36.8%		
PPP		0.0%	33.3%	0.0%	20.0%	10.0%	15.8%		
PC7 - ContractTyp.	0.540							NO	-
Private Contracts		66.7%	33.3%	40.0%	60.0%	50.0%	57.9%		
Public Procure.		33.3%	0.0%	0.0%	20.0%	10.0%	15.8%		
FIDIC Red		0.0%	33.3%	40.0%	20.0%	20.0%	21.1%		
FIDICSilv./Yel.		0.0%	33.3%	20.0%	0.0%	20.0%	5.3%		
PC8 - Pay. Method	0.354							NO	-
Fixed Price		44.4%	66.7%	0.0%	60.0%	50.0%	47.4%		
Unit Price		55.6%	33.3%	100.0%	40.0%	50.0%	52.6%		
PC9 - Dlvry. Syst.	0.172							NO	-
DBB		66.7%	50.0%	20.0%	60.0%	60.0%	79.0%		
DB		22.2%	16.7%	20.0%	0.0%	20.0%	21.0%		
EPC		11.1%	33.3%	60.0%	40.0%	20.0%	0.0%		
PC10 - Des.Comp.	0.601							NO	-
Very Low		22.2%	16.7%	20.0%	0.0%	20.0%	15.8%		
Low		11.1%	16.7%	0.0%	0.0%	10.0%	15.8%		
Moderate		11.1%	0.0%	20.0%	40.0%	30.0%	15.8%		
High		33.3%	16.7%	60.0%	40.0%	10.0%	47.4%		
Very High		22.2%	50.0%	0.0%	20.0%	30.0%	5.3%		
PC11 -Cons.Comp.	0.342							NO	-
Very Low		11.1%	0.0%	20.0%	0.0%	0.0%	15.8%		
Low		0.0%	16.7%	0.0%	0.0%	30.0%	15.8%		
Moderate		11.1%	16.7%	40.0%	20.0%	10.0%	10.5%		
High		55.6%	16.7%	40.0%	40.0%	30.0%	36.8%		
Very High		22.2%	50.0%	0.0%	40.0%	30.0%	21.1%		

The Chi-Square test results of skills attributes with respect to resolution methods are given in Table 3.22. Since attributes in skills category are ordinal categorical variables, exact probability values are obtained from the Mantel-Haenszel Linear-by-Linear Association Test. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. However, none of the skills attributes managed to have a place in the final model for resolution method selection.

Table 3.22. The Chi-Square Test Results of Skills Attributes for Resolution Method Selection Model

Attributes	p-value	Selected for Final Model	Strength of Association
S1 – Relationship between Parties / Individuals	0.356	NO	-
S2 – Previous Experience with Each Other or Reputation	0.445	NO	-
S3 – Dispute Avoidance Incentives	0.321	NO	-
S4 – Communication between Parties	0.799	NO	-
S5-1 – Working Culture & Skills of Represented Party	0.862	NO	-
S5-2 – Working Culture & Skills of Counter Party	0.577	NO	-
S6-1 – Response Rate & Communication Skills of Represented Party	0.526	NO	-
S6-2 – Response Rate & Communication Skills of Counter Party	0.144	NO	-
S7-1 – Experience of Represented Party	0.520	NO	-
S7-2 – Experience of Counter Party	0.954	NO	-
S8-1 – Project Management & Coordination Skills of Represented Party	0.735	NO	-
S8-2 – Project Management & Coordination Skills of Counter Party	0.547	NO	-

The Chi-Square test result of changes attribute with respect to resolution methods is given in Table 3.23. Both attributes are nominal categorical attributes. Therefore, the exact probability value is obtained from the exact Pearson Chi-Square statistics and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between changes and resolution method selection is statistically significant (p-value = 0.018 < 0.05). Thus, changes attribute should be kept as an input variable in the final model for resolution method selection. As both attributes are nominal categorical variables, the Cramer's V measure is used for determining the strength of association.

Table 3.23. The Chi-Square Test Result of Changes Attribute for Resolution Method Selection Model

Attributes & Categories	p-value	Resolution Method						Selected in Final Model	Str. of Assoc.
		LIT	ARB	DRB	MED	SEA	NEG		
C1 – Changes	0.018							**YES**	Cramer'sV
Yes		22.2%	100.0%	60.0%	80.0%	80.0%	47.4%		0.491
No		77.8%	0.0%	40.0%	20.0%	20.0%	52.6%		

The Chi-Square test result of delays attribute with respect to resolution methods is given in Table 3.24. Although the delays attribute is a numeric attribute normally, it is converted into an ordinal categorical variable using predefined threshold values. Therefore, the exact probability value is obtained from the Mantel-Haenszel Linear-by-Linear Association Test and it is compared to the alpha level at 0.05 for 95% confidence interval. The association between delays and resolution method selection is not statistically significant (p-value = 0.088 > 0.05). Therefore, it is eliminated from the final resolution method selection model.

Table 3.24. The Chi-Square Test Result of Delays Attribute for Resolution Method Selection Model

Attributes & Categories	p-value	Resolution Method						Selected in Final Model	Str. of Assoc.
		LIT	ARB	DRB	MED	SEA	NEG		
D1 – Delays	0.088							NO	-
Ratio = 0%		66.7%	16.7%	40.0%	20.0%	0.0%	26.3%		
Ratio 0% - 20%		0.0%	50.0%	0.0%	20.0%	30.0%	21.1%		
Ratio 20% 40%		22.2%	0.0%	20.0%	20.0%	10.0%	21.1%		
Ratio > 40%		11.1%	33.3%	40.0%	40.0%	60.0%	31.6%		

The Chi-Square test results of dispute characteristics attributes with respect to resolution methods are given in Table 3.25. In this table, exact probability values are obtained from the exact Pearson Chi-Square statistics for nominal attributes and the Mantel-Haenszel Linear-by-Linear Association Test for ordinal ones. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. The only dispute characteristics related attribute that is found to be associated with resolution method selection is "DC3 – Dispute Sources", which is a nominal categorical attribute. Thus, the Cramer's V measure is used for determining the strength of association.

Table 3.25. The Chi-Square Test Results of Dispute Characteristics Attributes for Resolution Method Selection Model

Attributes & Categories	p-value	Resolution Method						Selected in Final Model	Str. of Assoc.
		LIT	**ARB**	**DRB**	**MED**	**SEA**	**NEG**		
DC1 - Disputant	0.390							NO	-
Client		33.3%	0.0%	20.0%	20.0%	0.0%	21.1%		
Contractor		66.7%	100.0%	80.0%	80.0%	100.0%	78.9%		
DC2 -Phase Occur.	0.406							NO	-
Plan.&Design		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		
Construction		66.7%	100.0%	100.0%	80.0%	90.0%	89.5%		
Transfer&Repair		33.3%	0.0%	0.0%	20.0%	10.0%	10.5%		
DC3 – Disp.Source	0.014							**YES**	Cramer'sV 0.498
Cost compens. of change orders		0.0%	0.0%	0.0%	0.0%	20.0%	5.3%		
Time&Cost comp. of change orders		11.1%	66.7%	20.0%	60.0%	40.0%	21.1%		
Measurement & valuation of contracted works		33.3%	0.0%	20.0%	0.0%	20.0%	0.0%		
Extended overheads due to extensions		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		
Delay in site handover & possession		0.0%	0.0%	0.0%	0.0%	0.0%	15.8%		
Defects, errors and poor quality		22.2%	0.0%	0.0%	20.0%	0.0%	15.8%		
Contractor fails to act as a prudent merchant		11.1%	0.0%	0.0%	0.0%	0.0%	26.3%		
Delays in payments		11.1%	0.0%	0.0%	0.0%	10.0%	5.3%		
Errors or substantial changes in BoQ		0.0%	16.7%	60.0%	0.0%	0.0%	0.0%		
Inadequate site or soil investigation		11.1%	16.7%	0.0%	20.0%	0.0%	10.5%		
Interpretation of contract clauses		0.0%	0.0%	0.0%	0.0%	10.0%	0.0%		

Table 3.25. The Chi-Square Test Results of Dispute Characteristics Attributes for Resolution Method Selection Model (Continued)

Attributes & Categories	p-value	Resolution Method LIT	ARB	DRB	MED	SEA	NEG	Selected in Final Model	Str. of Assoc.
DC4 - Suspension	0.778							NO	-
Yes		44.4%	16.7%	40.0%	20.0%	40.0%	47.4%		
No		55.6%	83.3%	60.0%	80.0%	60.0%	52.6%		
DC5-Disp. Amount	0.485							NO	-
< 5 million $		44.4%	0.0%	40.0%	100.0%	30.0%	52.6%		
5 - 25 million $		33.3%	16.7%	20.0%	0.0%	50.0%	26.3%		
25 - 75 million $		22.2%	66.7%	20.0%	0.0%	10.0%	0.0%		
> 75 million $		0.0%	16.7%	20.0%	0.0%	10.0%	21.1%		
DC6-Sett. Amount	0.668							NO	-
0 $		11.1%	0.0%	0.0%	20.0%	0.0%	5.3%		
< 1 million $		22.2%	0.0%	40.0%	80.0%	20.0%	26.3%		
1 - 5 million $		33.3%	16.7%	20.0%	0.0%	30.0%	36.8%		
5 - 25 million $		22.2%	33.3%	0.0%	0.0%	30.0%	10.5%		
> 25 million $		11.1%	50.0%	40.0%	0.0%	20.0%	21.1%		
DC7-Success(Fnc.)	0.910							NO	-
0%		0.0%	0.0%	0.0%	0.0%	0.0%	5.3%		
0% - 25%		0.0%	0.0%	0.0%	0.0%	10.0%	0.0%		
25% - 50%		22.2%	50.0%	40.0%	20.0%	30.0%	15.8%		
50% - 75%		22.2%	50.0%	40.0%	20.0%	30.0%	42.1%		
> 75%		55.6%	0.0%	20.0%	60.0%	30.0%	36.8%		
DC8 – Presen. EoT	0.202							NO	-
Yes		22.2%	83.3%	20.0%	60.0%	50.0%	52.6%		
No		77.8%	16.7%	80.0%	40.0%	50.0%	47.4%		
DC9 -EoT Amount	0.976							NO	-
0 days		77.8%	16.7%	80.0%	40.0%	50.0%	47.4%		
0 - 6 months		11.1%	0.0%	0.0%	40.0%	30.0%	21.1%		
6 months - 1 year		0.0%	16.7%	0.0%	20.0%	10.0%	15.8%		
> 1 year		11.1%	66.7%	20.0%	0.0%	10.0%	15.8%		
DC10-Sett. EoT	0.709							NO	-
0 days		88.9%	16.7%	80.0%	40.0%	50.0%	47.4%		
0 - 6 months		0.0%	0.0%	0.0%	40.0%	30.0%	21.1%		
6 months - 1 year		0.0%	50.0%	0.0%	20.0%	10.0%	15.8%		
> 1 year		11.1%	33.3%	20.0%	0.0%	10.0%	15.8%		
DC11-Succes(EoT)	0.129							NO	-
0%		88.9%	16.7%	80.0%	40.0%	50.0%	47.4%		
0% - 25%		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		
25% - 50%		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		
50% - 75%		0.0%	33.3%	20.0%	20.0%	0.0%	0.0%		
> 75%		11.1%	50.0%	0.0%	40.0%	50.0%	52.6%		

The Chi-Square test results of resolution method characteristics attributes with respect to resolution methods are given in Table 3.26. Since all attributes in resolution method characteristics category are ordinal categorical variables, the Mantel-Haenszel Linear-by-Linear Association Test would have been used to obtain exact probability values. However, it was not possible to obtain the exact values due to computational limitations (except for RM1 and RM2). Therefore, the Monte Carlo method is used to calculate an unbiased estimate of exact probability values. The Monte Carlo estimates are within 99% confidence interval based on 10,000 sampled tables with starting seed 2,000,000. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. "RM3 – Level of Satisfaction with the Resolution Method" is not tested since this attribute is only used for determining which cases will be in the dataset for resolution method selection model. The only resolution method characteristics related attributes that are found to be associated with resolution method selection are "RM1 – Resolution Cost" and "RM2 – Resolution Duration", which are ordinal categorical attributes. Thus, the Somers' d measure is used for determining the strength of association.

Table 3.26. The Chi-Square Test Results of Resolution Method Characteristics Attributes for Resolution Method Selection Model

Attributes	p-value	Selected for Final Model	Strength of Association
RM1 – Resolution Cost	0.000	**YES**	Somers' d -0.909
RM2 – Resolution Duration	0.000	**YES**	Somers' d -0.667
RM4 – Importance of Preserving Relation. btw. Parties	0.943	NO	-
RM5 – Importance of Speed of Resolution	0.823	NO	-
RM6 – Importance of Cost of Resolution	0.687	NO	-
RM7 – Importance of Bindingness of the Process	0.571	NO	-
RM8 – Importance of Confidentiality of the Process	0.521	NO	-
RM9 – Importance of Fairness in the Process	0.069	NO	-
RM10 – Importance of Flexibility in Procedures	0.308	NO	-
RM11 – Importance of Control Over the Process	0.468	NO	-
RM12 – Importance of Reaching Creative or Remedying Solutions	0.387	NO	-
RM13 – Importance of Willingness in Reaching Soln.	0.759	NO	-

Finally, the Chi-Square test results of attributes that measure the level of knowledge on resolution methods with respect to resolution method selection are given in Table 3.27. Since all level of knowledge related attributes are ordinal categorical variables, exact probability values are obtained from the Mantel-Haenszel Linear-by-Linear Association Test. Probability values are compared to the alpha level at 0.05 for 95% confidence interval. The only knowledge level related attributes that are found to be associated with resolution method selection are “K1 – Level of Knowledge on Litigation” and “K2 – Level of Knowledge on Arbitration”, which are ordinal categorical attributes. Thus, the Somers’ d measure is used for determining the strength of association.

Table 3.27. The Chi-Square Test Results of Level of Knowledge on Resolution Methods Attributes for Resolution Method Selection Model

Attributes	p-value	Selected for Final Model	Strength of Association
K1 – Level of Knowledge on Litigation	0.005	**YES**	Somers’ d -0.309
K2 – Level of Knowledge on Arbitration	0.016	**YES**	Somers’ d -0.283
K3 – Level of Knowledge on DRB	0.699	NO	-
K4 – Level of Knowledge on Mediation	0.480		
K5 – Level of Knowledge on SEA	0.899	NO	-
K6 – Level of Knowledge on Negotiation	0.879	NO	-

In the light of test results, all three models will be finalized. In these finalized models, insignificant attributes will be removed and classifications will be based on the models that are composed of only the significant input variables determined in the Chi-Square tests. Thus, the finalized models will have less input variables that will enhance the classification performance of ML algorithms later. The details related to the finalization of prediction models will be given in the next section.

3.4. **FINALIZATION OF PREDICTION MODELS**

According to the Chi-Square results, the insignificant attributes (alpha level at 0.05) are eliminated. All significant attributes are in either moderately strong or strong relationship with the outputs. Therefore, all significant attributes are kept for the final prediction models.

In the conceptual model for dispute occurrence prediction, there were 25 input variables associated with the dispute occurrence. However, according to results of the Chi-Square tests, 14 of them are found to be associated with dispute occurrence in a statistically significant manner. These remaining attributes for dispute occurrence prediction model are:

1) PC1 – Project Location
2) PC2 – Project or Contract Value
3) PC3 – Planned Project Duration
4) S1 – Relationship between Parties / Individuals
5) S2 – Previous Experience with Each Other or Reputation
6) S4 – Communication between Parties
7) S5-1 – Working Culture & Skills of Represented Party
8) S5-2 – Working Culture & Skills of Counter Party
9) S6-2 – Response Rate & Communication Skills of Counter Party
10) S7-2 – Experience of Counter Party
11) S8-1 – Project Management & Coordination Skills of Represented Party
12) S8-2 – Project Management & Coordination Skills of Counter Party
13) C1- Changes
14) D1 – Delays

Figure 3.17 shows the finalized dispute occurrence prediction model with all attribute categories and significant input variables.

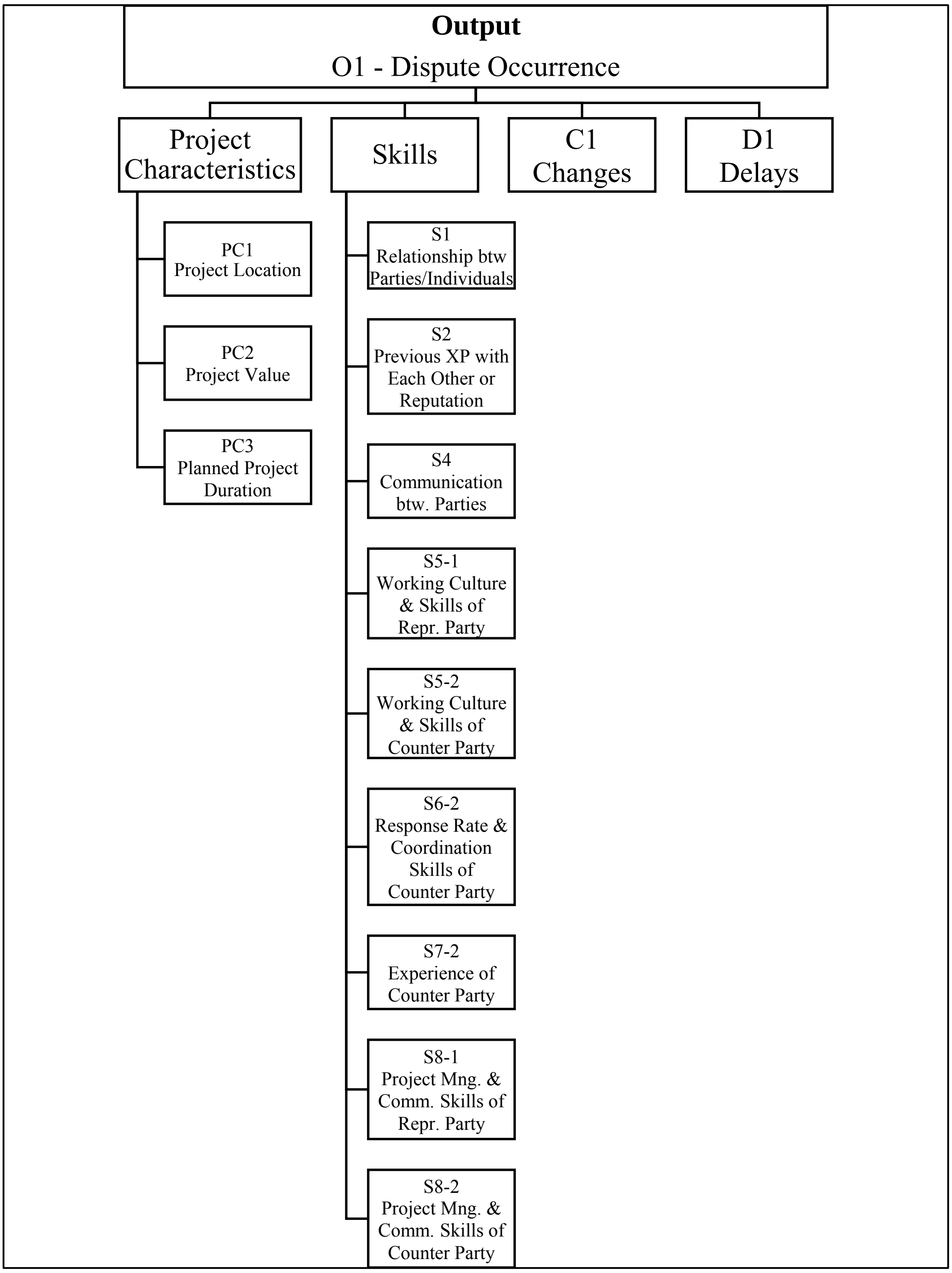


Figure 3.17. Finalized Dispute Occurrence Prediction Model

In the conceptual model for potential compensation prediction model, there were 32 input variables associated with potential compensation type. However, according to results of the Chi-Square tests, 9 of them are found to be associated with potential compensation in a statistically significant manner. The remaining attributes for potential compensation prediction model are:

1) PC5 – Type of Contractor
2) PC7 – Type of Contract
3) C1 – Changes
4) D1 – Delays
5) DC1 – Disputant Party
6) DC3 – Dispute Sources
7) DC5 – Disputed Amount (Financially)
8) DC8 – Presence of EoT Claim
9) DC9 – Disputed EoT Amount

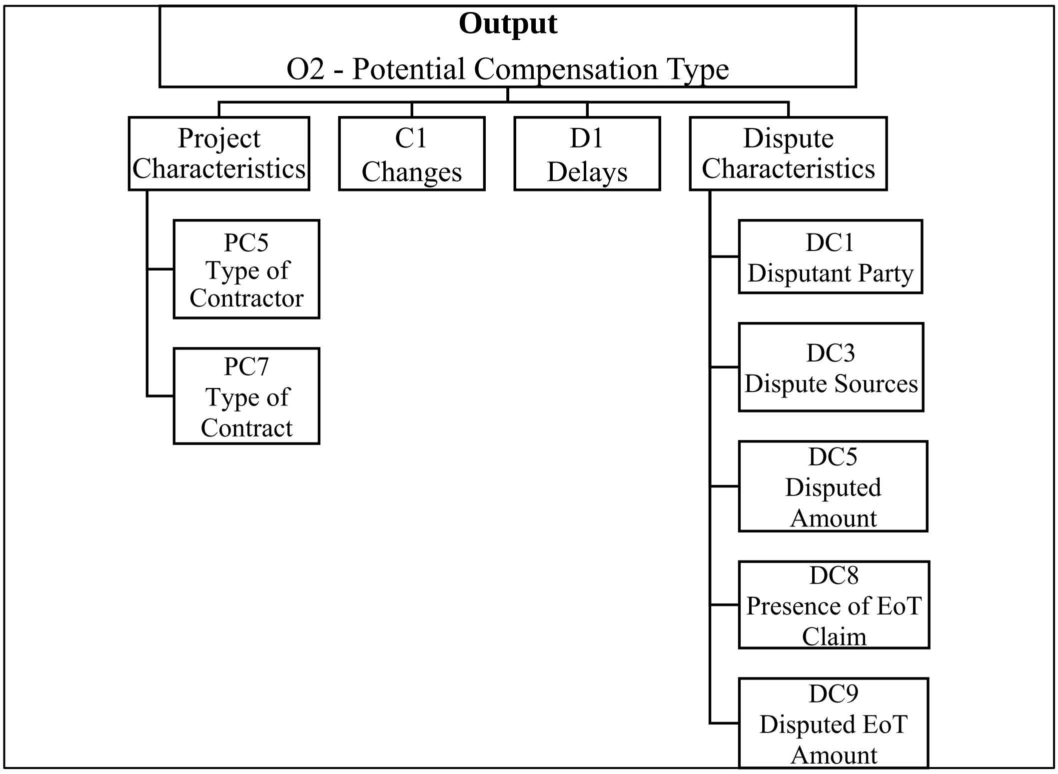


Figure 3.18. Finalized Potential Compensation Prediction Model

Figure 3.18 shows the finalized potential compensation prediction model with all attribute categories and significant input variables.

In the conceptual model for resolution method selection model, there were 55 input variables associated with resolution method selection. However, according to results of the Chi-Square tests, only 7 of them are found to be associated with resolution method selection in a statistically significant manner. The remaining attributes for resolution method selection model are:

1) PC5 – Type of Contractor
2) C1 – Changes
3) DC3 – Dispute Sources
4) RM1 – Resolution Cost
5) RM2 – Resolution Duration
6) K1 – Level of Knowledge on Litigation
7) K2 – Level of Knowledge on Arbitration

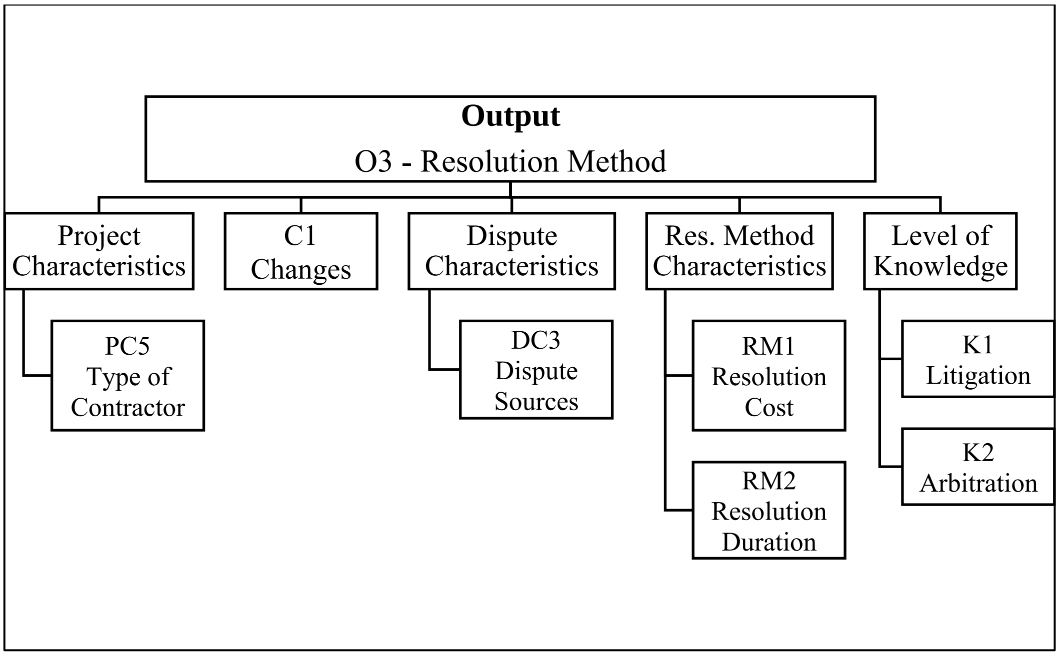


Figure 3.19. Finalized Resolution Method Selection Model

Figure 3.19 shows the finalized resolution method selection model with all attribute categories and significant input variables.

In short, this chapter started with the design of a questionnaire based on the conceptual model (Chapter 2) to collect empirical data on past construction projects. Initially, the collected dataset is analyzed to understand the nature of the data and to reveal initial findings. Then, the dataset is processed (i.e. data type conversions) to be prepared for attribute elimination and ML classification tasks. Remembering that the objective in this research is to develop three distinct prediction models (Chapter 2), there are three prediction problems. Prediction problems become data classification problems when the output variable is a categorical variable, which is the case in all prediction problems that are considered in this thesis study. In this research, data classification problems are handled by ML algorithms. In addition, attribute elimination is required to decrease the number of attributes used in ML algorithms and to enhance the generalization performance of these algorithms. For this purpose, attribute elimination on prediction models are performed based on the results from Chi-Square tests of association on attributes. In other words, insignificant attributes are eliminated from prediction models developed in Chapter 2 and all three models are finalized having significant attributes only. Subsequent to the mentioned efforts in Chapter 3 and finalization of three prediction models, in the fifth chapter, data classification via alternative ML algorithms will be performed on the finalized prediction models. In other words, finalized dispute occurrence prediction, potential compensation prediction, and resolution method selection models will be experimented by alternative ML algorithms in order to reveal the best classification performance for the corresponding dataset. Prior to this, the concept of data classification using ML algorithms and utilized algorithms in this thesis study will be explained in the next (fourth) chapter.

CHAPTER 4

DATA CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Construction professionals have to make challenging decisions while trying to achieve goals of the project. The current tendency in the construction industry is to make these challenging decisions intuitively based on the experience of the decision-maker with limited available information of questionable quality (Chou et al., 2013b). Therefore, current decision-making practices in dispute management domain are subjective instead of relying on systematical processes. AI applications, on the other hand, have the potential to minimize this subjectivity. In addition, construction professionals may benefit from decision-support systems in making informed decisions (Cheung et al., 2004a). Such systems can be developed by AI techniques that already yielded promising results in the literature and as a result, they are being soundly used in establishing decision-support systems (Pulket and Arditi, 2009a).

The superiority of AI techniques results from relying on empirical data via models and systems to justify the theories, test the insights, and interpret the results (İlter and Dikbaş, 2008). However, many researchers stated that there is a lack of empirical studies in disputes literature (Fenn et al., 1997; Love et al., 2010; İlter, 2012). In other words, despite the popularity of AI applications in construction, the attention to these applications is not reflected to construction dispute resolution domain. Leaving limited number of studies aside, more attention should be paid to utilization of AI techniques in construction dispute domain (Cheung et al., 2004a). Although AI techniques are not common in legal field (Chau, 2007), utilization of these techniques would have several benefits such as systematical selection of the resolution strategy (Cheung and Suen, 2002).

Among various AI applications, data mining via ML techniques form an important research branch since 1960's as they enable gathering valuable information from large volumes of data that is difficult to understand and interpret (Liao et al., 2012). Simply, data mining is the process of discovering useful structural patterns automatically or semi-automatically in large volumes of data with the aim of explaining the data and making predictions from it about new examples (Witten et al., 2016). Data mining is the application of ML methods to large databases and ML can be simply defined as programming of computers to optimize a performance criterion using past example data (Alpaydın, 2010). Proactive modeling by supervised ML as in classification and regression problems, clustering and association by unsupervised learning, evolution, pattern matching, data visualization, rule guided mining, etc. are among the data mining methods and the use of ML techniques is a potential data mining tool to deal with classification problems in construction management domain (Chou et al., 2014).

The intent of this thesis study is not to focus on all areas in data mining and ML but instead, the primary concern is data classification via ML techniques. For this purpose, general concepts related to ML techniques will be explained starting with the following section.

4.1. CONCEPTS RELATED TO MACHINE LEARNING

This section will introduce the necessary concepts in ML domain starting with the basic terms.

The primary goal of this research in utilizing ML techniques is to determine the patterns in the collected data so that prediction models can be developed, which can classify future cases. Each case in the dataset is called an instance and instances are defined by values of their attributes. These attributes are also called features and they are the observed variables. Each instance may have several attributes however; they can only have one target value that is called the class value or the class of the instance (Hsu et al., 2003).

4.1.1. Data Classification Problems

Data classification problems are problems of associating an instance (a case), which is defined by values of its attributes, with a class among predefined classes (Pulket and Arditi, 2009a). When the output variable is a categorical variable, prediction problems become data classification problems (Chou and Lin, 2012). In the case of dispute occurrence prediction, the output variable is dispute occurrence, where 'undisputed projects' can be categorized as '0' and 'disputed projects' as '1'. This kind of classification is known as 'binary data classification' since there are only two classes that instances can be assigned. ML algorithms are well equipped to solve binary data classification problems.

A similar situation is also present in potential compensation prediction and resolution method selection, where the potential compensation type and the resolution method as output variables can again be categorical. For example, in potential compensation prediction, no compensation cases can be categorized as one group, cost compensation only cases as another, time compensation only cases as another, and both cost and time compensation cases as another. Similarly in resolution method selection, litigation can be categorized as one group, arbitration as another, mediation as another, and so on. This kind of classification is known as 'multiclass data classification' since there are more than two classes (multiple classes) that instances can be assigned. Although not all ML techniques can handle multiclass classification problems, several techniques with multiclass classification capabilities are available in the literature. These techniques are solving multiclass classification problems either naturally by extending their binary classification capabilities or artificially by decomposing the problem into several binary classification tasks (i.e. one-versus-all, one-versus-one, error correcting output coding, etc.) (Aly, 2005). Thus, it can be said that ML techniques are well equipped to solve multiclass data classification problems observed in potential compensation type prediction and resolution method selection.

In short, data classification tasks are handled by data mining and ML techniques that is used to predict class labels (categories, group memberships) of instances in a dataset (Patel et al., 2014). In other words, classification involves the process of developing a model depending on training set instances with known class labels in pursue of predicting unknown class labels of each instance in the test set (Sobhana, 2014).

4.1.2. Training, Testing and Validation Sets

In classification tasks, data is usually separated into training and testing sets. The training set is composed of instances with several attributes and one assigned class. The goal of the ML algorithm is to search for structural patterns in the training set using these attributes and develop a model that links attributes of instances to their assigned classes. On the other hand, classes of instances in the test set are unknown to the algorithm. The algorithm tries to predict the class of the test set instance based on test instance's attributes using the developed model (Hsu et al., 2003).

ML algorithms generate classifiers resulting from the learning process using the training data. The performance of the classifier is measured in terms of error rate, where correct predictions of classes of instances are counted as success and incorrect ones as errors. The error rate of the classifier during prediction of classes of training set instances is the training set error rate. However, researchers are generally interested in future performance of classifiers on new data, not on the training data. Therefore, the error rate on a dataset that played no part in the learning process is required, which corresponds to the test set. In other words, the performance of the classifier is tested using error rate in the test set, which is called the test set error rate. It is important that both training and test sets should be capable of representing the dataset adequately and test set is not used in any way during establishment of the classifier (Witten et al., 2016).

The true error rate of a classifier that is determined by the test set depends on the size of the test set. Therefore, to eliminate the effect of the test set size, confidence intervals are used. Confidence intervals indicate the range of true error rate or true success rate

(Witten et al., 2016). This thesis study utilized 95% confidence level in the assessment of classification results.

If a classifier can neither perform well on training data nor generalize to new data, the classifier is under-fitting. Such models are not suitable. On the other hand, there is a problem of overfitting a model. Overfitting refers to a model that generalizes data too well, especially on the training set, so that the model's generalization capabilities are negatively affected. Overfitting results in poor classification performance during classification of new data. Overfitting is a common problem in applied ML domain. To avoid overfitting, one can hold back a validation dataset or use a resampling technique to estimate model accuracy such as cross-validation (Brownlee, 2018a).

In some cases, a third independent set can be used, which is the validation set. In such applications, training data is used by one or more algorithms to develop classifiers; and then, validation set is used to optimize parameters of these classifiers or to select a specific classifier among others. Finally, test set is used to determine the performance of the optimized classifier. Training, validation, and test sets should be completely different from each other with no common instances (Witten et al., 2016). Validation sets are good for checking the generalization ability of the classifier (Alpaydın, 2010).

4.1.3. Cross-Validation and Stratification

When limited amount of data is available, splitting the dataset into training and test sets may cause loss of information. Instead, researchers may prefer using all the data for knowledge extraction. However, this leaves no unseen instances, or in other words, test set. In such cases, cross-validation technique can be used, which is a procedure using all the data for learning and estimating the accuracy of the classifier by resampling the dataset (Vanwinckelen and Blockeel, 2012). K-fold cross-validation is the most common resampling technique. This technique is based on training and testing the model k-times on different subsets of training data to generate an estimate of the performance of a classifier on new data (Brownlee, 2018a).

In k-fold cross-validation, the dataset is divided randomly into k equally sized parts that are called folds (Alpaydın, 2010). One of the k folds is selected as a test set to test the trained classifier and the remaining k-1 folds are defined as training sets. This enables predicting each instance once and the cross-validation accuracy is directly equal to the correctly classified data percentage (Hsu et al., 2003). In short, in k-fold cross-validation, data is divided into k equal parts with each part being used for testing in turn, while the remaining parts are used for training.

In addition to the advantage of benefiting from all instances in the dataset, another advantage of cross-validation technique is that it can avoid overfitting (Hsu et al., 2003; Brownlee, 2018a). Besides the advantages, k-fold cross-validation has two primary shortcomings that should be considered. Firstly, the random sampling of k equally sized folds causes a risk related to data representation. Remembering training and test sets should be representative of the dataset; random sampling may cause uneven representation in training and test sets. In order to overcome such a problem, there is a procedure called stratification and cross-validation using this procedure is called stratified cross-validation. Stratification guarantees that during random sampling, each class is properly represented in both training and test sets (Witten et al., 2016). Thus, resorting to stratified cross-validation technique is beneficial in achieving representative training and test sets. Secondly, if two different k-fold cross-validations are performed using the same algorithm and dataset, but with different random sampling, there will most likely be two quite different classification performances. This is due to the high variance associated with results obtained from k-fold cross-validation. Results of a single k-fold cross-validation with high variance can be restored by repeating the cross-validation several times with different random samples of the same dataset (repeated cross-validation) and taking the average of results obtained from each cross-validation (Vanwinckelen and Blockeel, 2012). As a result, the variance associated with cross-validation technique will be decreased.

The k number in k-fold cross-validation is typically 10 (Alpaydın, 2010). Although this number can be adjusted depending on the size of the dataset and the desired level of analysis, literature has proven that 10-folds is the right number of folds based on experiments using various datasets and algorithms (Witten et al., 2016).

In the light of these, this thesis study will utilize ‘10 times repeated stratified 10-fold cross-validation’ in evaluating classifier performances.

4.1.4. Measures of Classifier Performance

Numerous measures that can be used when evaluating an ML algorithm are developed to serve for various domains with changing goals and considerations (Brownlee, 2018b). Therefore, measures to evaluate a classifier’s performance should be determined depending on the reviewed problem and desired outcomes. This is because usable evaluation measures change according to the type of data classification problem (i.e. binary or multiclass classification) and distribution of classes (i.e. balanced or unbalanced), (Sokolova and Lapalme, 2009).

In order to understand the performance of classification algorithms, (1) classification accuracy, (2) accuracy by class labels, and (3) confusion matrix should be considered. Other measures can be derived from these three items (Brownlee, 2018b). The classification accuracy, which is also known as the prediction accuracy, is the percentage of instances predicted correctly by a classifier divided by the total number of instances in the dataset and it is the primary evaluation criterion (Chou et al., 2013a). As this percentage increases, the success of the classifier increases. Although prediction accuracy is a primary evaluation criterion, making decisions solely by looking at this value can be misleading. If class distributions are unbalanced in reviewed datasets, other evaluation measures should also be checked such as Kappa statistic, precision, recall, etc. to take the class balance into account (Brownlee, 2018b). Therefore, consideration of accuracies by class labels can be beneficial in understanding the class breakdown for uneven datasets and multiclass classification problems. The confusion matrix can summarize such information and it is an appropriate tool to reveal the performance of the classifier for different classes. Thus, in ML domain, confusion matrices are commonly used to evaluate the performance of classifiers (Sönmez and Sözgen, 2017). This matrix is useful in calculating the accuracy (or error rates), especially in binary classification problems (Chou et al., 2013a). Basically, confusion matrix is a table that contains the number of predictions for each class compared to the actual number of instances that belongs to each class (Brownlee, 2018b). Table 4.1 is a typical confusion matrix for a binary classification problem.

Table 4.1. Confusion Matrix for Binary Classification

Class	Predicted Class: Positive	Predicted Class: Negative
Actual Class: Positive	True Positive (TP)	False Negative (FN)
Actual Class: Negative	False Positive (FP)	True Negative (TN)

In Table 4.1, it can be observed that the confusion matrix is composed of four counts as (1) true positive (TP), (2) true negative (TN), (3) false positive (FP), and (4) false negative (FN). If the prediction of an actually positive instance is positive, it is called a true positive classification. The TP rate is calculated according to Eq. [1]. If the prediction of an actually negative instance is negative, it is called a true negative classification. The TN rate is calculated according to Eq. [2]. If the prediction of an actually negative instance is positive, it is called a false positive classification. The FP rate is calculated according to Eq. [3]. If the prediction of an actually positive instance is negative, it is called a false negative classification. The FN rate is calculated according to Eq. [4]. (Alpaydın, 2010). The correctness of a classifier can be evaluated using these four counts (Sokolova and Lapalme, 2009).

$$TP\ Rate = \frac{TP}{actually\ positives} = \frac{TP}{TP + FN} \qquad [1]$$

$$TN\ Rate = \frac{TN}{actually\ negatives} = \frac{TN}{FP + TN} \qquad [2]$$

$$FP\ Rate = \frac{FP}{FP + TN} = 1 - TN\ Rate \qquad [3]$$

$$FN\ Rate = \frac{FN}{TP + FN} = 1 - TP\ Rate \qquad [4]$$

In the confusion matrix, the total of TP and TN classifications is equal to the correctly classified instances. In other words, the diagonal of the matrix represents correct classifications. Consequently, the classification (or prediction) accuracy is equal to the number of correctly classified instances divided by the total number of instances in the dataset as stated in Eq. [5] (Witten et al., 2016):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = (1 - error) \qquad [5]$$

In the ideal case, diagonal elements of the confusion matrix should be large and off-diagonal elements should be low (preferably ‘0’) for achieving high accuracy values (Witten et al., 2016). Besides correct classifications, remaining terms, which are FP and FN classifications, represent the two types of errors in the confusion matrix. (Alpaydın, 2010).

As mentioned earlier, the classification accuracy can be a misleading criterion in case of unbalanced datasets and there are other measures that can be resorted for this purpose. One of these measures is the precision measure that gives the positive predictive power of a classifier. Precision is the number of correctly classified positive instances (TP) divided by the number of instances predicted as positive by the classifier (TP + FP) and it is given by Eq. [6]. (Sokolova and Lapalme, 2009):

$$Precision = \frac{TP}{TP + FP} \qquad [6]$$

Another widely used performance measure is the recall, which is also called the sensitivity. Recall is the number of correctly classified positive instances (TP) divided by the number of actually positive instances in the dataset (TP+FN) and it shows the power of a classifier in identifying positive labeled instances (Sokolova and Lapalme, 2009). Thus, recall (sensitivity) is equal to the TP rate and it is given by Eq. [7] as follows:

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN} = TP\ Rate \qquad [7]$$

Similar to the recall measure, a classifier's power in identifying negative labeled instances can be determined by a measure called the specificity. It is the number of correctly classified negative instances (TN) divided by the number of actually negative instances (FP+TN) as given in Eq. [8] (Sokolova and Lapalme, 2009):

$$Specificity = \frac{TN}{FP + TN} = 1 - FP\ Rate \qquad [8]$$

There is a combined measure called the receiver operating characteristic (ROC) curve that characterizes the trade-off between the TP rate (or recall or sensitivity) and the FP rate (or '1 – specificity') by visualization. Thus, it is a combined measure of sensitivity and specificity (Park et al., 2004). The ROC curve depicts the performance of a classifier regardless of class distributions and error costs (Witten et al., 2016). The ROC curve is plotted with the TP rate (or precision or sensitivity) on the vertical axis and the FP rate on the horizontal axis. In ideal case, a TP rate of '1' and an FP rate of '0' is desired. Meanwhile, the reference case is the diagonal line representing the worst possible case where the number of correct classifications is equal to the number of incorrect ones. Consequently, the closer the ROC curve is to the upper-left corner, the better the performance of a classifier (Alpaydın, 2010). In other words, as the distance between the ROC curve and the reference line increases, the test accuracy increases (Chou et al., 2013a). Figure 4.1 represents sample ROC curves where curve A is the ideal form and Curve D is the diagonal (worst case) (Park et al., 2004). As curves move towards Curve A, better classifier performances are achieved.

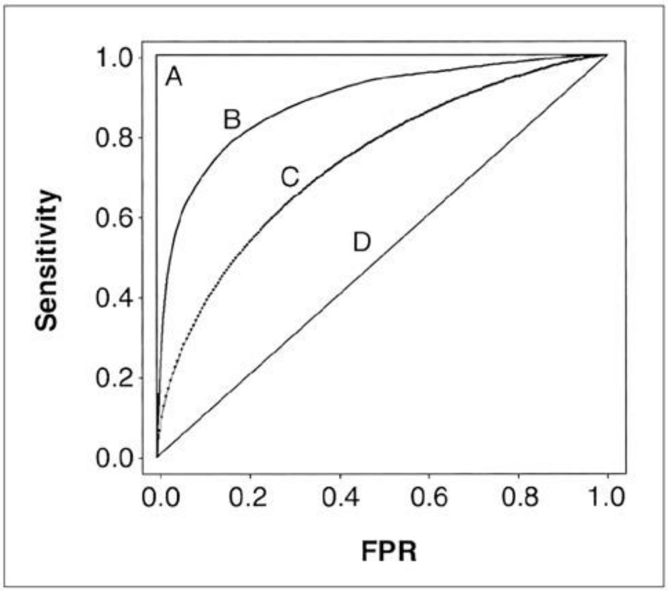


Figure 4.1. Sample ROC curves (Park et al., 2004)

Basically, the ROC curve indicates the ability of a classifier to avoid misclassifications (Chou and Lin, 2012). These curves are valuable as they enable visual analysis (Alpaydın, 2010). Moreover, ROC curves can be summarized in a single quantity that is called the area under the ROC curve (AUROC) and it can be said that the larger the area, the better the model (Witten et al., 2016). The ideal AUROC value is '1' and AUROC values can be used to compare a general performance averaged over different loss conditions (Alpaydın, 2010). The AUROC values can be simply calculated using the following formula (Sokolova and Lapalme, 2009):

$$AUROC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN}\right) = \frac{1}{2}(Sensitivity + Specificity) \qquad [9]$$

The final measure to be mentioned is the Cohen's Kappa coefficient, or in short, the Kappa statistic. It is a value that is used to measure the agreement between predicted and actual values in a dataset with a correction for agreements by chance (Witten et al., 2016). The Kappa statistic takes values between '-1' and '+1', where '+1' represents the perfect agreement, '0' represents the agreement is equal to chance, and '-1' represents the perfect disagreement. Kappa statistic can be calculated using Eq. [10], where P(A) is the observed agreement between actual and predicted values that is equal to the accuracy and P(E) is the expected agreement that is equal to the probability of agreement by chance (Eugenio and Glass, 2004):

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)} \qquad [10]$$

In ML terms, the Kappa statistic is another view of classification accuracy that is rescaled by comparing to the accuracy of a random classifier. Let C be the classification accuracy of the tested ML classifier and R be the classification accuracy of a classifier that randomly assigns instances to classes. Here, both C and R should assign the same number of instances to each class. In such a case, the Kappa statistic will be equal to:

$$Kappa = \frac{C - R}{1 - R} \qquad [11]$$

Thus, if the classification accuracy of the tested classifier (C) does not improve on the classification accuracy of the random classifier (R), the Kappa value will be equal to '0'. If C improves on R perfectly, the Kappa value will be equal to '1'. In other words, the Kappa value is '1' when the tested classifier is 100% accurate, which is the ideal case. Thus, the higher the Kappa statistic value, the better the performance of a classifier.

Among positive values of the Kappa statistic, Cohen suggested that results can be interpreted as: (1) values between '0.01-0.20' indicate none to slight agreement, (2) values between '0.21-0.40' indicate fair agreement, (3) values between '0.41-0.60' indicate moderate agreement, (4) values between '0.61-0.80' indicate substantial agreement, and (5) values between '0.81-1.00' indicate almost perfect agreement (Mchugh, 2012).

For multiclass classification problems, the class distribution dependent performance evaluation measures, which are precision, recall (TP rate or sensitivity), specificity, and AUROC, are calculated for each class separately. Then, weighted average values for these measures are taken as the final value. Moreover, in multiclass classification problems, the confusion matrix should be adjusted so that there is a row and a column for each class (Witten et al., 2016). Thus, when there are K classes with ($K > 2$), there should be a KxK confusion matrix. In such a matrix, an entry of {i,j} will represent the number of instances that belong to class i (C_i) but assigned to class j (C_j). Similar to the binary case, correct classifications are located on the diagonal of the matrix and in the ideal case, off-diagonal elements should be equal to '0' for no misclassifications (Alpaydın, 2010).

Table 4.2 shows a sample confusion matrix for multiclass classification with 3 classes. In this table, true positives are on the diagonal and remaining elements are misclassifications.

Table 4.2. Confusion Matrix for Multiclass Classification

Class	Predicted Class: Ci	Predicted Class: Cj	Predicted Class: Ck
Actual Class: Ci	True Positive Cii	Misclassification Cij	Misclassification Cik
Actual Class: Cj	Misclassification Cji	True Positive Cjj	Misclassification Cjk
Actual Class: Ck	Misclassification Cki	Misclassification Ckj	True Positive Ckk

As mentioned earlier, this thesis study will utilize stratified 10-fold cross-validation by repeating the process 10 times in order to mitigate impacts resulting from variance in cross-validation process. As a result, each experiment with a specific configuration of an algorithm will generate 10 different cross-validation results. These results will be averaged to identify the final performance of the experimented classifier. For example, consider experimenting the Naïve Bayes algorithm for the binary classification problem of dispute occurrence prediction. The Naïve Bayes algorithm will be run 10 times using stratified 10-fold cross-validation. Consequently, there will be 10 different Naïve Bayes classifiers with 10 different accuracy values. The final accuracy for the Naïve Bayes algorithm on this problem is the average of these 10 accuracy values. Therefore, a 95% confidence interval (CI) is constructed around average accuracy value so that the consistency of classifiers among cross-validation sets can be revealed.

A summarizing table for utilized measures of classifier performance can be seen in Table 4.3.

Table 4.3. Measures of Classifier Performance

Measure	Formula	Evaluation Focus
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$	Primary evaluation criterion
Kappa Statistic	$\frac{P(A) - P(E)}{1 - P(E)}$	Agreement between predicted and actual values, Correction for agreements by chance
Precision	$\frac{TP}{TP + FP}$	Positive predictive power of a classifier.
Recall Sensitivity TP Rate	$\frac{TP}{TP + FN}$	Power of a classifier in identifying positive labeled instances
Specificity	$\frac{TN}{FP + TN}$	Power of a classifier in identifying negative labeled instances
AUROC	$\frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN}\right)$	Ability of a classifier to avoid misclassification

4.1.5. Attribute Elimination in Machine Learning

As stated in Section 3.3, the performance of ML algorithms is generally affected negatively by the irrelevant or insignificant attributes (Pulket and Arditi, 2009b). As a rule of thumb, the number of attributes used in ML algorithms should be low (Arditi and Pulket, 2009). Therefore, for datasets containing large number of attributes, a process called attribute elimination (selection) should be performed so that insignificant or irrelevant attributes can be eliminated and the ones impacting the model performance are selected (Arditi and Pulket, 2009; Sönmez and Sözgen, 2017). Attribute elimination maintains better algorithm generalization (Drucker et al., 1999).

There are numerous techniques in the ML literature that can be used in attribute elimination such as correlation-based attribute subset selection, evaluating the worth of an attribute depending on the information gain with respect to class labels, etc. However, they require additional training time for determination of best attributes. Moreover, the level of complexity of an attribute elimination algorithm is at least quadratic times the level of complexity of the utilized ML algorithm (Drucker et al., 1999). Therefore, it can be beneficial if the attribute elimination is performed automatically or previously by external methods so that the complexity and the computation time would be lower. In the light of these observations, the Chi-Square statistics is preferred in this research among various alternatives. Details of the Chi-Square analysis has already given in Section 3.3. In short, the Chi-Square test is used as an attribute elimination tool before ML algorithms start to perform. In other words, insignificant attributes are eliminated according to results obtained from the Chi-Square tests.

4.2. SINGLE MACHINE LEARNING ALGORTIHMS

It is difficult to select the best performing ML algorithm that suit the prediction problem at hand. The literature has proven that it is not possible to solve all data mining problems using a single ML technique because of the varying characteristics of real world datasets. Instead, in order to obtain accurate results, the bias due to learning technique should be compatible with the dynamics of the problem domain, which makes data mining an experimental process (Witten et al., 2016). An accurate model that is proven to be well performing on a certain dataset does not necessarily have to perform well on another (Pulket and Arditi, 2009b). The conventional approach in ML domain is to experimentally compare classification performances of promising single ML algorithms with each other as base classifiers and select the best performing one in that dataset. (Arditi and Pulket, 2009).

In the light of these, it can be claimed that classification problems of dispute occurrence in construction projects, potential compensation type prediction, and resolution method selection require an experimental data mining approach. In other words, the ML technique that generates the best classification performance with available datasets should be experimentally determined. For this reason, several single ML algorithms will be considered as potential tools in this thesis study. In this section, only the single classifiers that are used in this thesis study will be introduced.

The reviewed single ML algorithms for data classification are taken from Witten et al. (2016) that lists the top 10 data mining algorithms based on results of a poll (Table 4.4). Among these algorithms, the ones that can be used in data classification are experimentally tested on collected dispute datasets. These algorithms are C4.5 (J48 in WEKA tool), SVM, kNN, and Naïve Bayes. The only data classification algorithm that is not experimentally tested from Table 4.4 is the CART algorithm. This is because both C4.5 and CART algorithms are decision tree based algorithms and only one of them is selected. C4.5 algorithm is preferred in this research as it is more popular in data mining domain. Besides selected algorithms, the MLP algorithm is also included in experiments since it is a commonly used technique in construction research. In short, the final 5 single data classification algorithms used in this thesis study are; (1) Naïve Bayes, (2) kNN, (3) C4.5 (or J48), (4) MLP, and (5) SVM.

In the following sections, selected ML algorithms will be introduced one by one.

Table 4.4. The Top 10 Algorithms in Data Mining (Witten et al., 2016)

No.	Algorithm	Category
1	C4.5 (J48 in WEKA)	Classification
2	K-means	Clustering
3	SVM	Statistical Learning (both binary classification and regression)
4	Apriori	Association Analysis
5	Expected Maximization	Statistical Learning
6	PageRank	Link Mining
7	AdaBoost	Ensemble Learning
8	KNN	Classification
9	Naïve Bayes	Classification
10	CART	Classification

4.2.1. Naïve Bayes Algorithm

The Naïve Bayes algorithm, which is a simple but powerful technique with comparable performance to many sophisticated algorithms for predictive modeling (Patel et al., 2014), is introduced in this section.

The Naïve Bayes is a simple probabilistic method that predicts the class label of an instance based on probabilistic calculations (Farid et al., 2014). In order to do this, the Naïve Bayes algorithm estimates conditional probabilities of classes given an observation by using joint probabilities of sample observations and classes based on the Bayes Theorem assuming conditional independence between classes (Li et al., 2004).

The Bayes Theorem can be applied when an observed event E occurs with any one of k mutually exclusive and exhaustive events such as $A = \{A_1, A_2,\ldots, A_k\}$. The formula for finding conditional probabilities $P(A_i \mid E)$, for $i = \{1, 2,\ldots,k\}$, is (Mendenhall and Sinchic, 2016):

$$P(A_i|E) = \frac{P(A_i \cap E)}{P(E)} = \frac{P(A_i)\,P(E|A_i)}{P(A_1)\,P(E|A_1) + P(A_2)\,P(E|A_2) + \cdots + P(A_k)\,P(E|A_k)} \qquad [12]$$

In ML point of view, given data (d), the best hypothesis (h) should be selected and in a classification problem, hypothesis (h) is the class label to assign for a new data instance (d). Thus, Eq. [12] can be converted to Eq. [13] from the ML point of view (Brownlee, 2018a):

$$P(h|d) = \frac{P(h)P(d|h)}{P(d)} \qquad [13]$$

In Eq. [13], P (h | d) is the probability of hypothesis (h) given the data (d) (posterior probability), P (d | h) is the probability of data (d) given the hypothesis (h) was true, and P (h) is the probability of (h) being true regardless of the data (prior probability of h). Finally, P (d) is the probability of the data regardless of the hypothesis (prior probability of d).

In order to calculate the posterior probability of P (h | d), the posterior probability for various hypothesis P (d | h) are calculated and the hypothesis with the highest probability is selected as the maximum probable hypothesis, which is called the 'maximum a posteriori (MAP) hypothesis' (Brownlee, 2018a).

$$MAP\ (h) = \max(P(h|d)) = \max\left(\frac{P(h)P(d|h)}{P(d)}\right) = \max\left(P(d|h)\ \times P(h)\right) \qquad [14]$$

In Eq. [14], P (d) can be omitted since it is a constant value that is only used to normalize the term for calculating the probability.

To explain these in more details, equations given above will be further reviewed. Firstly, the Naïve Bayes algorithm calculates class probabilities for each class in the training set. For example, the class probability of an instance belonging to 'class 1' among k classes can be calculated by Eq. [15]:

$$P\ (class = 1) = \frac{frequency\ (class = 1)}{freq.(class = 1) + freq.(class = 2) + \cdots + freq.(class = k)} \qquad [15]$$

Secondly, conditional probabilities of each attribute value (frequencies of each attribute value), given the class labels are calculated. For example, given a categorical attribute with 2 categories (values) and 2 categorical class labels (binary), conditional probabilities are calculated by operations in Eq. [16]:

$$P\ (attribute = value\ 1 \mid class = 1) = \frac{freq.(attr. = value\ 1\ \cap class = 1)}{freq.(class = 1)}$$

$$P\ (attribute = value\ 2 \mid class = 1) = \frac{freq.(attr. = value\ 2 \cap class = 1)}{freq.(class = 1)}$$

$$P\ (attribute = value\ 1 \mid class = 2) = \frac{freq.(attr. = value\ 1 \cap class = 2)}{freq.(class = 2)}$$

$$P\ (attribute = value\ 2 \mid class = 2) = \frac{freq.(attr. = value\ 2 \cap class = 2)}{freq.(class = 2)} \qquad [16]$$

Remembering Eq. [14], in order to make a prediction for the class of a new instance, choose the largest value obtained from Eq. [17]:

$$(class = 1) = P(attr. = value\ 1 \mid class = 1) \times P(class = 1)$$

$$(class = 2) = P(attr. = value\ 1 \mid class = 2) \times P(class = 2) \qquad [17]$$

If the value obtained for ‘class = 1’ is greater than the value obtained for ‘class = 2’, the new instance belongs to ‘class = 1’, and otherwise, it belongs to ‘class = 2’. Thus, the new instance belongs to the class with the highest posterior probability.

In the light of the given simple probabilistic calculation process, the Naïve Bayes classifier is commonly used for data classification problems due to its simplicity and high performance classification accuracy. Moreover, the classifier is easy to use and probability values are generated in one iteration through the training set (Farid et al., 2014). Although the algorithm is very simple, it can outperform many sophisticated algorithms in terms of classification performance (Witten et al., 2016). In addition, the Naïve Bayes algorithm can be naturally extended to solve multiclass data classification problems (Thakkar et al., 2011). Therefore, it can be used both for the binary data classification problem in dispute occurrence prediction and for multiclass classification problems in potential compensation type prediction and resolution method selection. Another advantage of the Naïve Bayes algorithm is its capability of handling missing values easily by simply omitting corresponding probabilities for missing attributes during calculation of class probabilities (Farid et al., 2014).

The Naïve Bayes algorithm has some limitations. To start with, the algorithm can use categorical variables only. Thus, numeric input variables can only be used if they are converted to categorical variables or an adjusted version of the algorithm, which is called the Gaussian Naïve Bayes, is utilized. However, in the Gaussian Naïve Bayes, numeric attributes are assumed to have Gaussian distributions. Therefore, the data distribution should be processed by removing the outliers, etc. (Brownlee, 2018a). In addition, the Naïve Bayes algorithm can use kernel density functions to handle numeric attributes and upon satisfying the normality assumption, better performance can be achieved using kernel estimators (Amin and Habib, 2015). The most crucial assumptions in the Naïve Bayes technique is that it assumes attributes do not interact and they are independent in terms of probability. In other words, the impact of an attribute value on a given class is independent of values of other attributes, which is called conditional independence (Patel et al., 2014). This is a very simplifying assumption that is unlikely to happen in real datasets, but the classification accuracy of the algorithm can still has comparable performance to other algorithms (Thakkar et al., 2011).

Considering the simplicity and mentioned advantages, the Naïve Bayes algorithm is considered as a potential ML algorithm to be experimentally tested in this research.

4.2.2. K-Nearest Neighbor (KNN) Algorithm

In this section, the kNN algorithm for data classification will be introduced. The kNN is a conventional non-parametric classification technique that classifies an unknown instance represented with some feature vectors as a point in a feature space by calculating distances between the point and other points in the training set (Chou et al., 2013a). Here, the term non-parametric refers to methods that do not involve strong assumptions about functions, which are used to map inputs to outputs. For example, in the Linear Regression method, a line is assumed as a function to map input variables to the output. On the other hand, non-parametric techniques are capable of learning any functional form from training instances. As non-parametric methods depend on training sets, they have high variance compared to parametric ones. However, the bias in non-parametric methods are lower as assumptions made by the model about the target function are less strong (Brownlee, 2018a).

The kNN algorithm has been widely used in various problems of information retrieval (Li et al., 2004). The algorithm is considered as an inductive method as it performs a search through all instances of a training set to classify a new instance (Chou et al., 2013a). It does not learn a model and instead, predictions are made by calculating a similarity distance between the new instance and every training instance (Brownlee, 2018a). Thus, it is categorized as an instance-based (case-based) learning technique and sometimes called a lazy-learner.

Basically, the kNN classifier assigns the new instance, which is a point in the feature space, to the class with the most instances among k neighbors in the feature space. The k value is an integer determining the number of neighboring instances to be evaluated. All neighbors have equal vote in determination of the class of the new instance and upon ties, an arbitrary selection is made for the class label or a weighted voting is performed. The k value can be taken as an odd number to avoid ties (Alpaydın, 2010). In short, the algorithm finds the k number of closest training data points to the new instance and predicts the class label of the new instance based on the majority of class labels of neighboring points (instances) (Li et al., 2004).

While considering the closest training instances, the algorithm utilizes a distance measure. The distance between the new instance and every other training instances are calculated, k smallest distances are identified, and the most occurring class label in these k instances are assigned to the new instance as the class label (Thakkar et al., 2011). One of the most common distance measures utilized in the kNN algorithms is the Euclidean distance function that calculates the distance as the square root of the sum of the squared differences between two points across all attributes (Altun and Polat, 2008). There are several other distance measures in the literature, which involves the Hamming distance functions for calculating distances between binary vectors, the Manhattan distance functions for measurements using the sum of absolute differences, and the Minkowski distance function that combines the Euclidean and the Manhattan measures. However, the most suitable distance measure should be selected according to the dataset. The Euclidean distance measure is more appropriate when attributes are similar in types (i.e. all numeric) and the Manhattan distance measure is more suitable when attributes are not similar in types (i.e. nominal, ordinal, and numeric attributes together) (Brownlee, 2018a). Considering that datasets in this research have varying attribute types, the Manhattan distance measure seems to be a more appropriate measure. Indeed, the best classification performance is obtained from the kNN classifier when the Manhattan distance measure is utilized as proven in Chapter 5.

Another important parameter that should be considered is the k value in kNN classifiers. The k value that gives the best performance on a test set by matching characteristics of the data should be determined. Although there are studies suggesting to take the k value as ‘1’ will be enough to achieve considerable classification performance in most applications (Jain et al., 2000), the value of k should normally be determined using a validation set for parameter optimization or using cross-validation (Thakkar et al., 2011). As a rule of thumb, if the number of training instances were large, it would be more accurate to use more than one nearest neighbor. However, for datasets with few instances, using a large k value might be problematic. When the ratio of the k value to the number of instances approaches ‘0’, the probability of error approaches the theoretical minimum for the dataset (Witten et al., 2016). Therefore, selection of the k value should be carefully performed.

The kNN algorithm can handle multiclass data classification problems in addition to binary classification problems (Altun and Polat, 2008). Thus, kNN classifier is eligible for use in all problems reviewed in this thesis study.

Besides advantages, the kNN algorithm contains limitations. One major problem is that the kNN is more suitable for classification problems with lower dimensionality. This means less dimensionality in the feature space, or in other words, less number of attributes (input variables) should be used. The increasing number of attributes causes an exponential volume increase in the feature space and as a result, the distance between similar instances may be calculated as if they have large distances between each other. In addition, the algorithm performs poorly in case of missing values (Brownlee, 2018a). Finally, the user of this algorithm has to identify the best distance measure suitable for the dataset and the optimum k value for the best classification performance (Brownlee, 2018b). This requires extra effort in computations unlike the case in the Naïve Bayes algorithm.

Considering the popularity of the kNN algorithm in the data mining domain and mentioned advantages, the algorithm is worthy of a try for classification problems reviewed in this thesis study.

4.2.3. J48 Decision Tree Algorithm

In this section, the C4.5 decision tree algorithm (Quinlan, 1993) for data classification will be introduced. The WEKA version of the C4.5 algorithm is called the J48 decision tree algorithm. The decision tree is a powerful classification technique and the most commonly used versions are the CART and the C4.5 decision trees (Thakkar et al., 2011).

Decision trees are algorithms that display the classification process of instances graphically using a tree-like structure (Drazin and Montag, 2012). The tree structure of decision trees is top-to-down structure, where internal nodes represent a test of an attribute, branches represent a test outcome, and leaf nodes represent classes or class distributions. The top-most node is the root node and it is the node with the highest information gain. After partitioning the samples using the root node, the second attribute with the highest information gain is used as an internal node to partition the data further. Partitioning is terminated when all attributes are used and data cannot be partitioned anymore (Chou et al., 2013a). The representation of a typical decision tree structure can be seen from Figure 4.2.

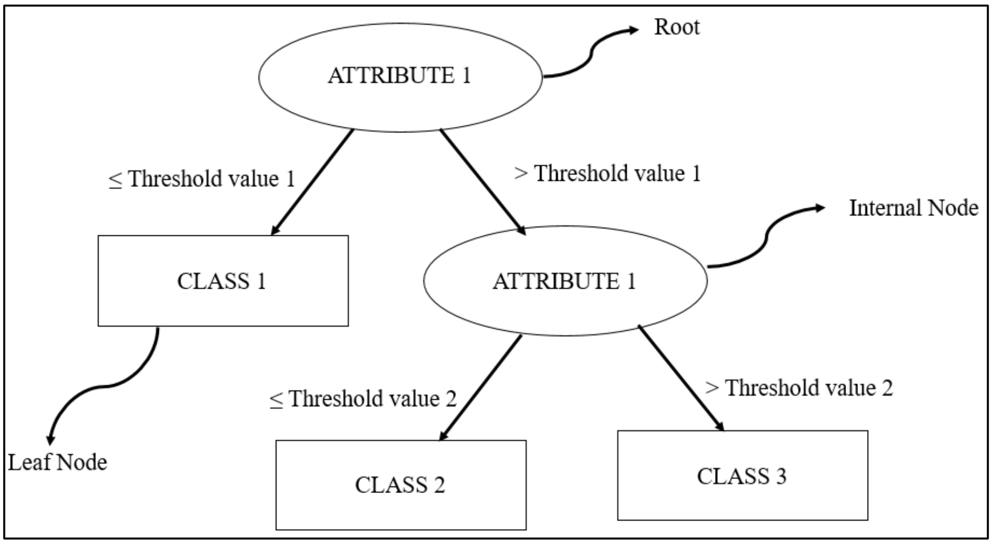


Figure 4.2. Representation of a Typical Decision Tree Structure

The information gain process will be briefly explained. Consider a training dataset with a definite number of instances and these instances can belong to several classes with labels $\{C_1, C_2, \ldots, C_n\}$. For example, the probability of an instance belonging to a certain class (let's say class label '1' for this case), which is denoted as P_1, is:

$$P_1 = \frac{freq.of\ all\ class\ 1\ instances\ in\ training\ set}{number\ of\ all\ instances\ in\ the\ training\ set} \quad [18]$$

Consequently, the probability of an instance belonging to class '2' is denoted by P_2, and so on. Thus, there is a probability distribution such that $P = \{P_1, P_2, \ldots, P_n\}$. The information associated with these probabilities are called entropy of P, which is equal to (Chou et al., 2013a):

$$Info\ (P) = \sum_{i=1}^{n} -p_i log_2 p_i \quad [19]$$

Consider a set A composed of instances partitioned according to an attribute X to generate sets such that {A_1, A_2, …, A_m}. The information required to determine the class label of an instance belonging to set A will be the weighted average of the information needed to identify the class of an element A_i, that is, the weighted average of Info (A_i) (Chou et al., 2013a):

$$Info\ (X, A) = \sum_{i=1}^{m} \frac{|A_i|}{A} \times Info(A_i) \quad [20]$$

The information gain, which is denoted as Gain (X, A), represents the difference between the information required to identify an element of A and the information required to identify an element of A after the threshold value of attribute X is determined (Chou et al., 2013a). In other words, it is the information gain resulting from the attribute X. The equation is:

$$Gain\ (X, A) = Info\ (A) - Info\ (X, A) \quad [21]$$

Returning back to the philosophy behind decision trees, the tree splits the training dataset based on a threshold value of an attribute at a node. The attribute selected for splitting the data is the attribute that gives the highest information gain. When a new instance should be classified, a test is performed according to attribute values of the instance following a path starting from the root node and ending at a leaf node. The final leaf node that is reached depending on test results will be the class label of the tested instance (Thakkar et al., 2011).

Decision trees are usually followed by a process called pruning. Pruning is a process that optimizes the computational efficiency and the classification accuracy of decision trees. Moreover, the size of the pruned tree is reduced so that the complexity of generated results are also reduced (Drazin and Montag, 2012). Decision tree algorithms are generally associated with overfitting and the resulting tree can be subjected to a post-pruning process in order to avoid overfitting. Post-pruning is performed on a generated tree by removing statistically insignificant nodes and consequently, reducing the size of the tree (Li et al., 2004). There are various types of pruning strategies utilized in decision trees and subtree raising is one of the most common pruning techniques available. Subtree raising is a post-pruning process that raises a subtree out of the most popular branch, which is the branch with the highest number of training instances, in a decision tree (Lavesson and Davidsson, 2006).

There are many advantages associated with decision trees that can be listed as (1) resulting trees are easy to understand, (2) the algorithm is capable of handling various attribute types such as categorical and numeric, (3) the algorithm is capable of handling missing values, (4) the algorithm has high performance compared to the number of iterations, (5) it is easy to implement, and (6) it is easy to generate rules of classification (Girja Sharma et al., 2013). Besides these advantages, the algorithm is a powerful technique that can naturally handle both binary and multiclass data classification problems (Thakkar et al., 2011). Thus, the decision tree algorithm, specifically the C4.5 (J48 in WEKA tool) is considered as a candidate technique for this research. However, disadvantages should not be underestimated. The size of trees increases linearly with the number of instances in the dataset and consequently, the algorithm performs poorly in large and noisy datasets. In addition, decision trees require large storage spaces as attribute values are stored repeatedly in arrays (Girja Sharma et al., 2013).

4.2.4. Multilayer Perceptron (MLP)

In this section, an ANN algorithm that is called the MLP neural network will be introduced. The ANN is composed of information-processing units that mimic synaptic processes in biological neurons of the brain in order to reveal relationships in input datasets through iterative operations so that new data can be generalized. The resulting network is a collection of interconnected adaptive processing elements (Cheung et al., 2002). The primary goal of the ANN is to establish a brain-like computational system that performs various tasks (i.e. classification, optimization, clustering, etc.) using parallel processing elements to achieve faster solutions than competing systems (Sobhana, 2014). In a typical ANN, brain neurons are represented by a group of neural and weighted nodes, while synapses between brain neurons are represented by interconnections between these nodes (Chou et al., 2013a). The connection strength between nodes (or processing elements) is network weights that can be adjusted in order to achieve an output that matches a desired response such that; each input is multiplied by a weight and the sum of all weighted inputs reveal the degree of activation level, which is processed further by an activation function to produce an output (Cheung et al., 2002).

The rationale behind ANNs can be more simply explained as follows: Each node in the input layer represents an attribute and is accompanied by an additional constant bias unit (Witten et al., 2016). All nodes in the network are fully connected to every node in the following layer via connection weights and an operation is performed on the inputs, the connection weights, and the bias term all together to calculate values for the next layer. The activation function propagates in the forward direction from input layer to reach the hidden layer. (Alpaydın, 2010). There are several activation functions in the literature including binary step, bipolar step, identity, and sigmoid functions (Sobhana, 2014). The MLP uses a nonlinear (sigmoid) function as the activation function to calculate values of the hidden layer (Alpaydın, 2010). If there are more than one hidden layers, values of the predecessor hidden layers are considered as inputs and values for the successor hidden layers are calculated with a similar approach. This process is repeated until the output layer is reached. Each node in the output layer represents a class and values of the output nodes are calculated as the weighted sum of its inputs (from the hidden layer) through an activation function (Sobhana, 2014).

The MLP neural networks are the most common form of feed forward neural networks that are trained by back propagation algorithm (Sobhana, 2014). They are now considered as the standard ANN models in the literature that is composed of an input layer containing a set of sensory input nodes, one or more hidden layers and an output layer containing computational nodes (Chou et al., 2013a). Other than input nodes, all nodes in the MLP network are neurons, or processing elements, with a nonlinear activation function (Patel et al., 2014). Indeed, the name perceptron refers to the basic processing element that has inputs coming from a dataset or from other perceptrons' (i.e. hidden layers) outputs (Alpaydın, 2010). The MLP generates output values at the end and the difference between these calculated outputs and target outputs are defined as the mean-squared error function. In MLP networks, the aim is to minimize this error function. In order to minimize the error function, weighted connections between neurons are optimized. The optimization in the MLP can be performed by back propagation algorithm (Altun and Polat, 2008).

A typical MLP network structure with one hidden layer can be seen in Figure 4.3 that shows the nonlinear mapping from inputs to an output (Witten et al., 2016). However, this representation omits the bias terms in the input layer. In this representation, inputs $\{a_o, a_1, \ldots, a_k\}$ represent attributes in the input layer, weights $\{w_{00}, w_{10}, \ldots, w_{ik}\}$ represent connection weights between the input layer and the hidden layer, hidden units $\{0, 1, \ldots, l\}$ represent hidden neurons, functions $\{f(x_1), f(x_2), \ldots, f(x_n)\}$ represent nonlinear activation functions, weights $\{w_1, w_2, \ldots, w_l\}$ represent connection weights between the hidden layer and the output layer, and finally, f(x) represents the output activation function.

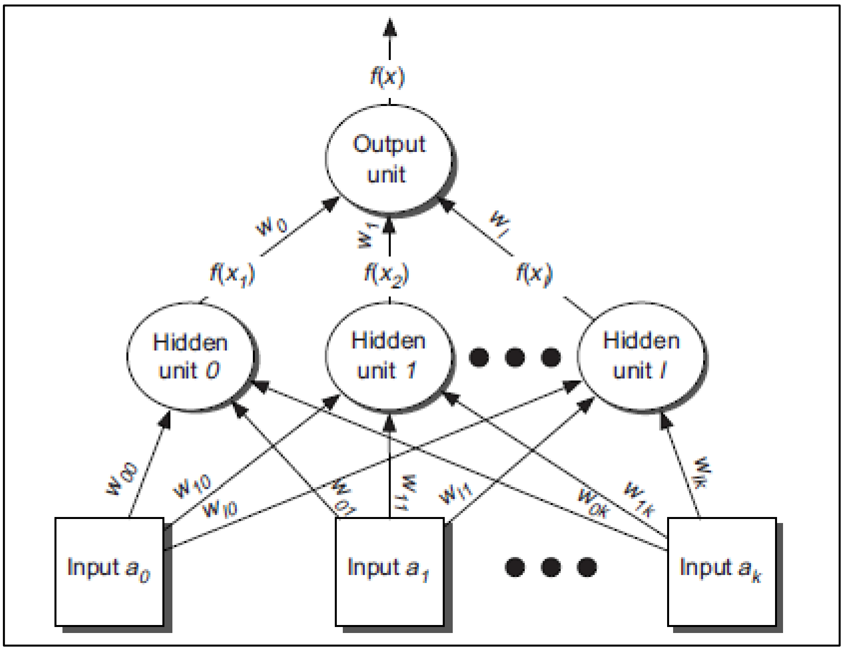


Figure 4.3. A Typical MLP Network with One Hidden Layer Omitting the Bias Terms (Witten et al., 2016)

The MLP can distinguish data that are not linearly separable (Patel et al., 2014) and it is a non-parametric estimator that can be used for classification problems (Alpaydın, 2010). This non-parametric nature of the MLP is associated by an advantage and a disadvantage at the same time. The advantage is that there are no prior assumptions about the distribution of the data so that, the bias due to the algorithm decreases (Sobhana, 2014). However, when the error function is expressed as the mean-squared error, which is the case in the MLP, the distribution of the dataset cannot be controlled by the algorithm and the literature has proven that the distribution of the data significantly affects the training performance of MLP networks (Altun and Polat, 2008). Other determinants of performance of the MLP are the structure of the network and the utilized training algorithm (Sobhana, 2014). As mentioned earlier, a supervised learning technique that is called back propagation is used for training the MLP network (Patel et al., 2014). In back propagation, the error propagates from outputs to inputs (Alpaydın, 2010). Thus, connection weights in MLP networks are continuously modified, or optimized, to reduce errors until the total error from all training instances is lower than a predefined threshold value (Chou et al., 2013a).

One final advantage of the MLP is that it can be naturally adapted to multiclass data classification problems. In binary classification problems, the MLP network will only have one neuron in the output layer. On the other hand, in multiclass classification problems, if there are ‘n’ classes, there will be ‘n’ output neurons (Thakkar et al., 2011). Therefore, the MLP technique can be used appropriately for the binary data classification problem in dispute occurrence prediction as well as for multiclass data classification problems in potential compensation prediction and resolution method selection.

4.2.5. Support Vector Machines (SVM)

In this section, the SVM algorithm, which was developed by Cortes and Vapnik (1995) for binary classification problems, will be introduced. The SVM algorithm uses linear models to implement nonlinear class boundaries by transforming the input space into a higher-dimensional feature space using a nonlinear mapping that generates a linear model in the new space to represent a nonlinear decision boundary in the original one (Witten et al., 2016). The theoretical background of the SVM algorithm will be explained in Section 4.2.5.1. During transformation of input space to a higher-dimensional space, the SVM utilizes kernel functions, which will be explained in Section 4.2.5.2.

The SVM is a popular supervised learning technique containing many favorable properties (Moraes et al., 2013). The algorithm has gained attention due to its outstanding generalization capability in cases with limited training data, which is the situation in most real world applications (Belousov et al., 2002a). It is now among the most powerful tools for classification and regression problems (Cheng and Wu, 2009).

It is claimed that the generalization performance of the SVM algorithm is either matching or significantly better than competing methods (Burges, 1998). Resulting from its solid theoretical background, the SVM classifies instances more accurately than most of the other algorithms in the literature (Moraes et al., 2013). Moreover, experimental results prove that the SVM algorithm achieves good performance on classification problems and outperforms competing methods substantially (Joachims, 1998). Consequently, the SVM algorithm has been successfully applied to many real world classification problems including the ones in the construction management domain such as contractor qualification decision, project success prediction, contractor default prediction, cash flow prediction, conceptual cost estimation, bid-no bid decision-making, litigation outcome prediction, dispute prediction, etc. (Sönmez and Sözgen, 2017).

The SVM algorithm learns a classifier from attributes and class labels of the training data in order to predict unknown class labels of the test data (new data) using attributes of the test data only (Hsu et al., 2003). During classification, a subset of training instances, which are called the support vectors, are used to derive a hyperplane that separates instances from each other according to their class labels in a feature space (Cheng and Wu, 2009). In other words, the SVM algorithm utilizes support vectors to parametrize functions during derivation of a separating hyperplane in a higher-dimensional feature space with the aim of achieving linear classification (Belousov et al., 2002b). This separating hyperplane forms the decision boundary in class assignments and maximizes the margin between the two classes. For linearly non-separable cases and cases with nonlinear decision boundaries, the SVM resorts to kernel functions that map the input space into a higher-dimensional feature space during the training process (Cheng and Wu, 2009). An extensive explanation of the theoretical background will be given in Section 4.2.5.1. Instead, advantages and favorable features of the SVM algorithm will be highlighted at this point.

As mentioned earlier, the generalization performance of the SVM is outstanding in most applications. While the generalization performance is maximized, training errors are minimized (An et al., 2007). Moreover, considering the current nature of data mining applications in which it is required to measure large number of variables simultaneously with limited amount of training samples due to time and cost constraints on data gathering processes, the importance of the SVM algorithm can be better realized. This is because the SVM algorithm can produce flexible classifiers automatically and systematically to achieve outstanding generalization performance on datasets with numerous attributes and limited amount of instances (Belousov et al., 2002a). Therefore, the SVM can solve classification problems in datasets with high input dimensionality (too many attributes). In addition, the SVM algorithm does not require the estimation of parameters of class distributions and consequently, all instances can be assigned to a class including the clear outliers (Belousov et al., 2002b). Thus, it can be said that the generalization ability of the SVM is robust against datasets that have high input dimensionality with problematic distributions.

Similar to all ML techniques, the SVM has its own drawbacks. Firtsly, the algorithm is designed for binary classification problems (Cortes and Vapnik, 1995). For multiclass classification problems, methods such as "one-versus-one (OVO)", "one-versus-all (OVA)", etc. should be externally used (An et al., 2007). Secondly, the utilized kernel function should be carefully selected as the learning performance of the SVM depends on the choice of kernel function (Friedrichs and Igel, 2005). Thirdly, besides SVM parameters, kernel function parameters should also be optimized. In other words, the success of the SVM classification depends on selected parameters of the algorithm and for this purpose, external methods should be used for parameter optimization (Chou et al., 2014). This process is known as hyperparameter optimization in practice.

In order to give a better insight on the SVM rationale, the theoretical background of the algorithm is explained in the next section.

4.2.5.1. Theoretical Background of SVM Algorithm

Basically, the SVM is a supervised classification technique originated from statistical learning theory (Sönmez and Sözgen, 2017). The main principle is to find an optimal hyperplane that separates two classes and for this purpose, the SVM algorithm searches for the hyperplane that maximizes the distance to the closest training instance from either class in order to achieve better classification performance on test instances (Moraes et al., 2013). This hyperplane is the optimal separating hyperplane with the maximal margin, where the margin is the distance from the hyperplane to closest instances on both sides of the hyperplane (Alpaydın, 2010). Both the optimal hyperplane and the associated optimal margin are determined by a small portion of the training data that is called the support vectors. These support vectors impact the generalization ability of the SVM. If the number of support vectors needed to derive the optimal hyperplane is small, the resulting SVM will have better generalization (classification) performance (Sönmez and Sözgen, 2017).

In more simple terms, the SVM finds the optimal hyperplane, which separates training instances with the maximum margin according to their class labels, by using closest instances to the hyperplane as support vectors (Joachims, 1998). An example linearly separable problem can be seen in Figure 4.4 (Cortes and Vapnik, 1995). In this representation, there is a linearly separable problem in a 2-dimensional space with instances of one class is shown by crosses and the other with circles. Instances in grey squares represent the support vectors that define the maximal margin, which is the largest separation between these two classes.

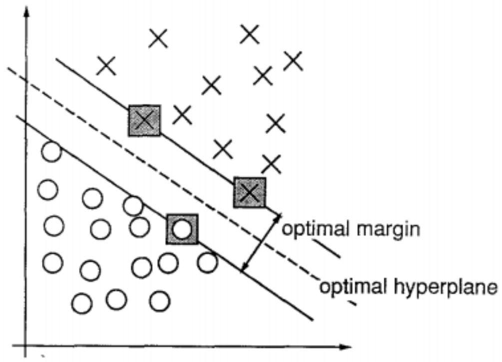


Figure 4.4. Representation of a Separable Binary Classification Problem in 2-Dimensional Space by the SVM (Cortes and Vapnik, 1995)

The SVM rationale is explained following the notation of Burges (1998) based on Cortes and Vapnik (1995). Suppose there is a binary classification problem with a training set containing L elements belonging to two separate classes. Each element consists of a pair as x_i and y_i such as $\{x_i, y_i\}$ and $y_i \in \{-1, 1\}$ with i = {1, …, L}. Here, x_i represents input vectors and y_i represents the associated class label of x_i. In the light of these, the training set can be expressed as follows:

$$(y_1, x_1), \dots, (y_L, x_L), \quad y_i \in \{-1, 1\} \qquad [22]$$

The equation of a hyperplane is defined by:

$$w.x + b = 0 \qquad [23]$$

In the hyperplane equation, w is the normal vector to the hyperplane (weight vector), x is the input vector, and b is the bias term.

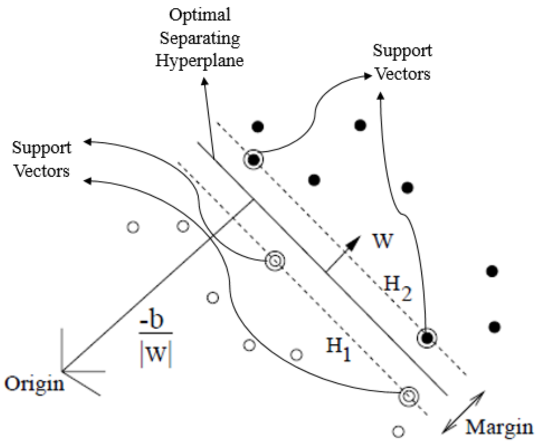


Figure 4.5. Linear Separating Hyperplanes for Separable Case (Burges, 1998)

The representation of linear separating hyperplanes for a separable case can be seen in Figure 4.5 (Burges, 1998). In this case, there are two hyperplanes H_1 and H_2 such as:

$$H_1 : x_i.w + b = 1 \qquad [24]$$

$$H_2 : x_i.w + b = -1 \qquad [25]$$

Points on the hyperplanes H_1 and H_2 are the support vectors. There is another hyperplane that is in the middle of H_1 and H_2. This hyperplane is defined as:

$$H_0 : x_i.w + b = 0 \qquad [26]$$

The hyperplane H_0 represents the decision boundary that separates the two classes. The distance between H_0 and H_1 is given by (Burges, 1998):

$$\frac{|x_i.w+b|}{||w||}=\frac{1}{||w||} \quad [27]$$

In Eq. [27], the term ||w|| represents the Euclidean norm of vector w. Since H_0 is placed at the same distance from both H_1 and H_2, the total distance is (2 / ||w||). This is the margin between separating hyperplanes (Burges, 1998). In other words, this is the shortest distance between the optimal separating hyperplane H_0 and the closest positive and negative instances.

For a linearly separable binary classification case, the SVM finds the optimal separating hyperplane with the largest margin according to the following constraints (Burges, 1998):

$$x_i.w+b \geq +1 \qquad when\ y_i=+1 \qquad [28]$$

$$x_i.w+b \leq -1 \qquad when\ y_i=-1 \qquad [29]$$

The constraints in Eq. [28] and Eq. [29] can be combined in one constraint as follows:

$$y_i(x_i.w+b) \geq 1 \qquad \forall i \qquad [30]$$

To summarize the procedure up to this point, the summation of the shortest distances from the separating hyperplane to the closest positive and negative instances will be equal to the margin. Hyperplanes H_1 and H_2 are parallel to each other with no training instances between them. In this case, there are instances at a distance (1 / ||w||) on both sides of the separating hyperplane according to Eq. [27]. Therefore, the summation giving the margin is (2 / ||w||). Consequently, the hyperplane that gives the maximal margin can be determined by maximizing (2 / ||w||), which is the same thing as minimizing the Euclidean norm ||w||, and it is subject to the constraint given in Eq. [30] (Burges, 1998).

In more simple terms, the task for determining the maximal margin can be expressed by Eq.[31] as a standard quadratic optimization problem (Cortes and Vapnik, 1995):

$$\min \frac{1}{2}\|w\|^2 \qquad \textit{subject to} \qquad y_i(x_i.w+b) \geq 1 \qquad \forall i \qquad [31]$$

The optimization problem in Eq. [31] is known as a constrained optimization problem that can be solved by introducing positive Lagrangian multipliers such as ($\alpha_i \geq 0$) with i = {1, …, L}. Therefore, the problem can be reformulated into an equivalent unconstrained optimization problem (Burges, 1998):

$$L_P = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i \, y_i(x_i.w+b) + \sum_{i=1}^{L} \alpha_i \qquad [32]$$

According to Eq. [32], the task is to minimize the primal Lagrangian problem L_P and it can be achieved by minimizing w and b such that (Dibike et al., 2001):

$$\frac{\partial L_P}{\partial b} = 0 \rightarrow \qquad \sum_{i=1}^{L} \alpha_i \, y_i = 0 \qquad [33]$$

$$\frac{\partial L_P}{\partial w} = 0 \rightarrow \qquad w = \sum_{i=1}^{L} \alpha_i \, y_i x_i \qquad [34]$$

Substituting Eq. [33] and Eq. [34] into Eq. [32], the following dual Lagrangian formula L_D is obtained:

$$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{L} \alpha_i \alpha_j y_i y_j (x_i . x_j) \qquad [35]$$

According to Eq. [35], the task is to maximize the dual Lagrangian problem L_D and it can be achieved by maximizing α_i subject to the constraint of positive Lagrangian multipliers ($\alpha_i \geq 0$) and the constraint in Eq. [33] with the solution obtained from Eq. [34] (Burges, 1998).

Points where the Lagrangian multiplier are greater than '0' ($\alpha_i \geq 0$), are the support vectors of the solution (Alpaydın, 2010). These are the vectors on hyperplanes H_1 and H_2 in Figure 4.5. Remaining points have Lagrangian multipliers equal to '0' ($\alpha_i = 0$) and they are located on either side of hyperplanes H_1 and H_2.

The solution can be found in the form of $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_L\}$ by maximizing the L_D and weights are given by (Burges, 1998):

$$w = \sum_{support\ vectors} y_i \alpha_i x_i \qquad [36]$$

In the light of all these, the class label of a new instance x is determined using the following decision function (Burges, 1998):

$$f(x) = sign.(w.x + b) = sign.\left(\sum_{support\ vectors} y_i \alpha_i (x_i . x) + b \right) [37]$$

The mentioned rationale of the SVM algorithm is valid for the linear and separable cases. However, real data is generally more complex and cannot be separated perfectly with a hyperplane (Brownlee, 2018a). When the training data is not separable without errors, the SVM algorithm needs some adjustments to find a solution. It is known as the soft margin hyperplane for non-separable cases (Alpaydın, 2010). In such cases, the training set should be separated with the minimum number of misclassified instances, or errors (Cortes and Vapnik, 1995). Figure 4.6 demonstrates the non-separable linear SVM classifier in 2-dimensional feature space (Burges, 1998) with the slack variable and the constraint violation (soft margin classifier).

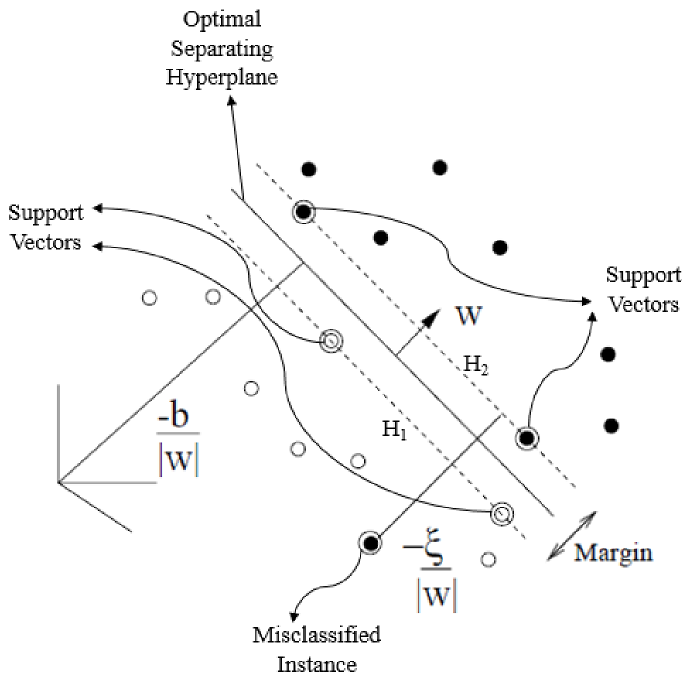


Figure 4.6. Linear Separating Hyperplanes for Non-Separable Case (Burges, 1998)

The soft margin classifier allows some training instances to violate the separation between the two classes. For this reason, an additional set of coefficients are utilized that are called the slack variables (Brownlee, 2018a). In other words, the constraint given in Eq. [30] should be violated for non-separable cases using the slack variables ξ_i with i = {1, …, L} that determine the amount of violation of the constraint (Alpaydın, 2010). Then, the constraint becomes (Cortes and Vapnik, 1995):

$$y_i(x_i.w + b) \geq 1 - \xi_i \qquad for\ \forall i \qquad [38]$$

$$\xi_i \geq 0 \qquad for\ \forall i \qquad [39]$$

Then, a penalty parameter C is introduced to the optimization problem and the problem becomes (still subject to constraints in [Eq. 38] and [Eq. 39]) (Cortes and Vapnik, 1995):

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{L} \xi_i \qquad [40]$$

The penalty parameter C is selected beforehand and it determines the cost of violating the constraint (Dibike et al., 2001). This time, the primal Lagrangian formulation is given by:

$$L_P = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{L} \xi_i - \sum_{i=1}^{L} \alpha_i \left(y_i(x_i.w + b) - 1 + \xi_i\right) - \sum_{i=1}^{L} \mu_i \xi_i \qquad [41]$$

In Eq. [41], the term μ_i represents the Lagrangian multiplier enforcing the positivity of ξ_i. Similar to the linearly separable case, this primal Lagrangian problem is converted to the dual form such that (Alpaydın, 2010):

$$\frac{\partial L_P}{\partial b} = 0 \rightarrow \qquad \sum_{i=1}^{L} \alpha_i \, y_i = 0 \qquad [42]$$

$$\frac{\partial L_P}{\partial w} = 0 \rightarrow \qquad w = \sum_{i=1}^{L} \alpha_i \, y_i x_i \qquad [43]$$

$$\frac{\partial L_P}{\partial \xi} = 0 \rightarrow \qquad C - \alpha_i - \mu_i = 0 \qquad [44]$$

As the term ($\mu_i \geq 0$), the Eq. [44] implies that:

$$0 \leq \alpha_i \leq C \qquad [45]$$

According to these, the dual Lagrangian problem, which is subject to Eq. [42] and Eq. [45] (different than the separable case), becomes:

$$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{L} \alpha_i \alpha_j y_i y_j (x_i . x_j) \qquad [46]$$

Up to this point, the linear SVM for separable and non-separable cases are considered. One final case is the nonlinear SVMs, which are mentioned in the next section.

4.2.5.2. Kernel Functions in SVM

Besides the abilities to classify instances in linearly separable and non-separable cases, the SVM algorithm is also capable of classifying instances with classes that cannot be linearly separated. In order to do this, the input data is transformed into a higher-dimensional space, where the data is linearly separable in this new space. The nonlinear decision boundary in the original feature space (original input space) can be determined easily because it is linear in the higher dimensional feature space. Moreover, there is no need to compute the parameters of the optimal hyperplane in the feature space with high dimensionality. Instead, the solution is computed as a weighted sum of values obtained from utilized kernel function that is evaluated at the support vectors only. The transformation of a 2-dimensional input space, which cannot be linearly separated, into a 3-dimensional space is illustrated in Figure 4.7 (Moraes et al., 2013). The 3-dimensional version can be linearly separated as shown by the shaded hyperplane.

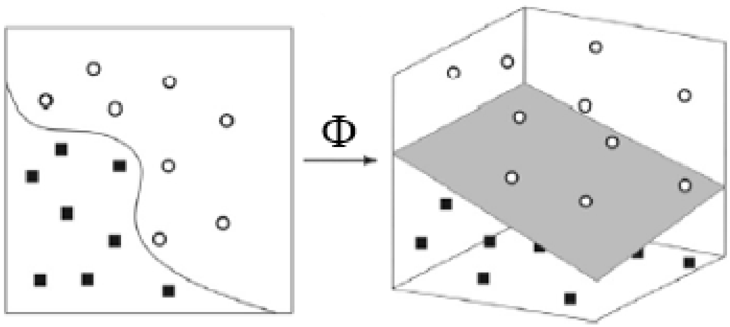


Figure 4.7. Representation of the Transformation of Input Space to a Higher-Dimensional Feature Space using Kernel Functions (Moraes et al., 2013)

The mapping of input space to a higher-dimensional feature space by a kernel function K is expressed as follows (Cortes and Vapnik, 1995):

$$K(x_i, x_j) = \Phi(x_i) \,.\, \Phi(x_j) \qquad [47]$$

The goal of the SVM in cases such as the one in Figure 4.7 is to map the nonlinear problem in input space into a linear problem in higher-dimensional feature space using a nonlinear mapping Φ, which is an unknown term that computes the inner product of input data points in a feature space created by Φ (Wang et al., 2003). Due to Eq. [47], the dual Lagrangian problem in Eq. [46] becomes (still subject to constraints in Eq. [42] and Eq. [45]):

$$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{L} \alpha_i \alpha_j y_i y_j K(x_i . x_j) \qquad [48]$$

The SVM algorithm is characterized by the utilized kernel function (Burges, 1998). Therefore, in order to achieve reasonable performance from SVM classifiers, it is necessary to select the appropriate kernel function that suits to the reviewed problem and dataset (Dibike et al., 2001). The learning performance of the SVM is dependent to the choice of kernel function (Friedrichs and Igel, 2005). Although there are various kernel functions in the literature, the three commonly used kernel functions are the polynomial kernel, the Gaussian radial basis function (RBF) kernel, and the sigmoid kernel (An et al., 2007).

4.2.5.3. Polynomial Kernel SVM

The polynomial kernel is expressed as follows (Cortes and Vapnik, 1995):

$$Polynomial\ Kernel: \ K(x_i, x_j) = (x_i . x_j + 1)^d \qquad [49]$$

In Eq. [49], the term ‘d’ of the polynomial kernel function determines the degree (order) of the polynomial with ‘d’ always taking positive integer values. As can be seen from the equation, the polynomial kernel computes dot products of two vectors (x_i and x_j) and raises the result to the power ‘d’. While using polynomial kernel functions, the task is to use the best value for parameter ‘d’ that gives the best generalization performance. For this purpose, it is suggested to start with a linear model, which means the ‘d’ parameter is equal to ‘1’, and increase it until the performance criteria is met (Witten et al., 2016).

4.2.5.4. Gaussian Radial Basis Function (RBF) Kernel SVM

The RBF kernel, which is also known as the Gaussian kernel, can be expressed as follows (Burges, 1998):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^{\;2}}{2\sigma^2}\right) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \qquad for\, \gamma > 0 \qquad [50]$$

Among all kernel functions for the SVM algorithm, the most commonly used function is the Gaussian RBF kernel, where the parameter ‘σ’, which is always positive, is the spread parameter that affects the generalization performance of the algorithm by determining the characteristic width of the function in Eq. [50] (Belousov et al., 2002b). In RBF kernel SVM algorithm, only two parameters are required as ‘C’ and ‘γ’, which should be optimized (Hsu et al., 2003).

The literature has proven that the Gaussian kernel SVM algorithm exhibits good features and strong learning capability in classification problems (Wang et al., 2003). Indeed, the RBF kernel is suggested by researchers as a reasonable first choice in attempt to classify instances using the SVM algorithm (Hsu et al., 2003; Cheng and Wu, 2009). The RBF kernel is claimed to be superior to linear, polynomial, and sigmoid kernels. Unlike the linear kernel, the RBF kernel can handle a dataset that is composed of attributes and associated class labels with nonlinear relationships. In fact, the linear kernel is a special case of RBF kernel where the performance of the linear kernel SVM with a parameter ‘C’ is the same as the performance of the RBF kernel SVM with a parameter pair as ‘C’ and ‘γ’ (Hsu et al., 2003). Therefore, compared to linear kernel, there is an observation that RBF kernel produces better accuracy for nonlinear cases (Hsu and Lin, 2002). Similarly, the sigmoid kernel mimics the RBF kernel for certain parameters and moreover, in a study comparing the sigmoid and RBF kernels, it is revealed that the sigmoid kernel is not better than the RBF (Lin and Lin, 2003). In addition, the sigmoid kernel function is not valid under certain parameters (Cheng and Wu, 2009). Finally, the RBF has fewer parameters than the polynomial kernel that makes it numerically and computationally less complex (Sönmez and Sözgen, 2017).

Although RBF kernels are superior to linear, polynomial, and sigmoid counterparts, it might not be appropriate to use them in some cases, especially when the number of features (attributes) are very large compared to available instances in the dataset (Hsu et al., 2003). In addition, the number of support vectors can be relatively high in the RBF kernel SVM algorithm. Considering that computations are made only for support vectors during determination of the optimal hyperplane, the large number of support vectors in the RBF kernel SVM may result in a higher computational time while developing the classifier (Dibike et al., 2001).

In the light of all, the RBF kernel SVM is experimented in datasets of this research. On the other hand, considering the amount of instances and attributes in reviewed datasets, the polynomial kernel SVM is also experimented.

4.3. ENSEMBLE MACHINE LEARNING ALGORITHMS

The classification performances (prediction accuracies) of single ML algorithms (base classifiers) can be enhanced further by creating ensemble classification schemes systematically (Arditi and Pulket, 2009). In a series of studies on dispute prediction and resolution method selection, it is highlighted that the prediction performance of ensemble models can outperform the classification performance of single classifiers (Chou, 2012; Chou and Lin, 2012; Chou et al., 2013a; Chou et al., 2013b; Chou et al., 2014). This is mainly because misclassified instances by various single ML algorithms do not overlap (Kittler et al., 1996). Therefore, ensemble approaches, which are simply adding or combining base classifiers, can compensate errors of base classifiers and improve the classification accuracy.

There are various approaches to develop ensemble models. One approach is to combine classification results of two or more classifiers into a single ensemble score using voting techniques such as majority voting, average of probabilities, etc. (Chou and Lin, 2012). The voting technique is explained in Section 4.3.1. Another approach is based on combining single classifiers sequentially so that the first classifier can be used to reduce the amount of data for following classifiers (Chou et al., 2013a). This thesis study utilized the stacked generalization, in which different classifiers are combined, and the AdaBoost algorithm, in which the performance of a weak classifier is enhanced. The stacked generalization is explained in Section 4.3.2 and the AdaBoost algorithm is in Section 4.3.3.

4.3.1. Voting Technique

The voting technique is utilized in generation of ensemble models and Figure 4.8 shows the rationale behind the technique. Firstly, several base classifiers classify instances on their own. Then, results of the models are combined by voting techniques to develop the ensemble model that will make the final classification decision of instances.

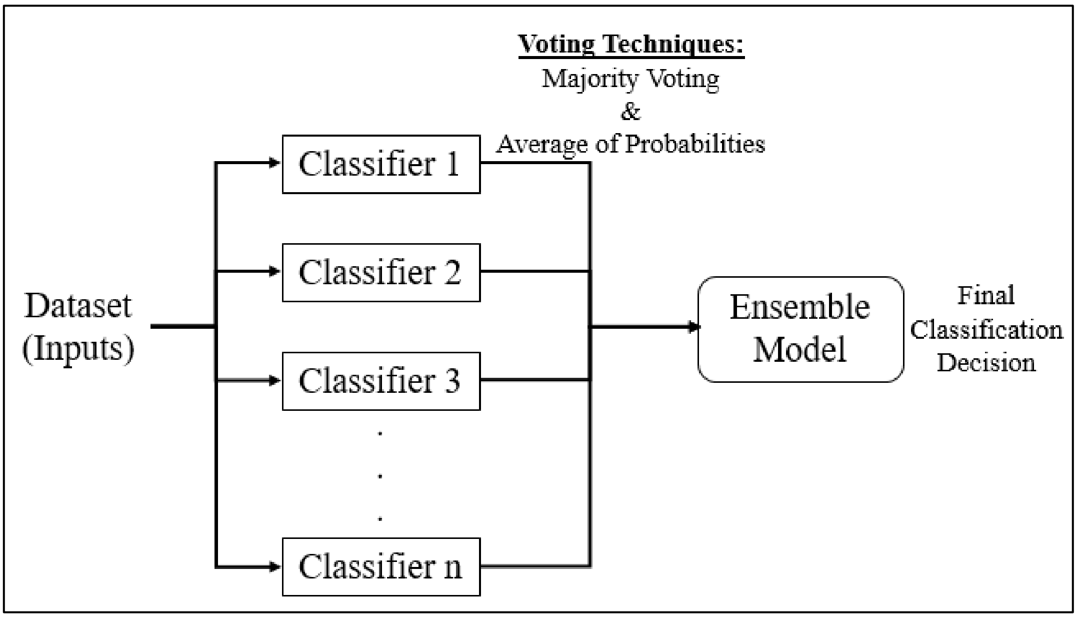


Figure 4.8. Generation of Ensemble Models using the Voting Technique

Among various voting strategies in the literature, the majority voting and the average of probabilities techniques are experimented. In majority voting, the class assigned to an instance by the majority of single classifiers is accepted as the final class label decision (Kittler et al., 1996). When base classifiers have comparably well classification performances, equal (unweighted) voting would be the sensible choice (Witten et al., 2016). For this reason, this research considered votes as equally important and they are not weighted.

Classification models obtained from single ML algorithms can output probability estimates of class labels of instances. In other words, they do not only generate class predictions and instead, they can generate probabilistic outcomes on class assignments. The average of probabilities voting technique uses advantage of this by taking probabilistic outcomes of each single classifier and averaging them to make the final class assignment decision. Such an approach may improve the classification accuracy (Witten et al., 2016).

In this research, results of the top three base classifiers in terms of classification accuracy out of six experimented single algorithms are considered during voting. For dispute occurrence prediction, these algorithms are (1) Gaussian RBF kernel SVM, (2) polynomial kernel SVM, and (3) C4.5 (J48). For potential compensation prediction, these algorithms are (1) Naïve Bayes, (2) kNN, and (3) C4.5 (J48). For resolution method selection, these algorithms are (1) C4.5 (J48), (2) Naïve Bayes, and (3) MLP.

4.3.2. Stacked Generalization

Stacked generalization (or in short, stacking) is a technique for minimizing the generalization error by combining two or more classifiers and the error reduction is achieved by reducing biases of classifiers with respect to a provided learning set (Wolpert, 1992).

In more simple terms, a classifier is trained and tested using a dataset that contains L instances with some classification performance; inevitably, there will be classification errors and incorrectly classified instances are removed from the original dataset to obtain a smaller dataset with L' instances (L' < L), which is the subset of the first dataset with L elements. These, L' instances are correctly classified instances by the first classifier. Then, the second classifier is trained and tested using this new dataset. The resulting classifier, which combines performances of two classifiers, will be the ensemble classifier that might achieve better classification performance than each single classifier (Chou et al., 2013a). In other words, correct predictions of an initial base classifier are used as inputs in a secondary classifier to form the combined ensemble model. Here, the initial base classifier is called the base-learner and the secondary classifier is called the meta-learner, where most of the work is done by the base-learner and the meta-learner is like an arbiter. Therefore, it would be better to select a simpler algorithm as a meta-learner considering the computational complexity associated with stacking (Witten et al., 2016). Figure 4.9 shows the rationale in ensemble model generation using stacked generalization.

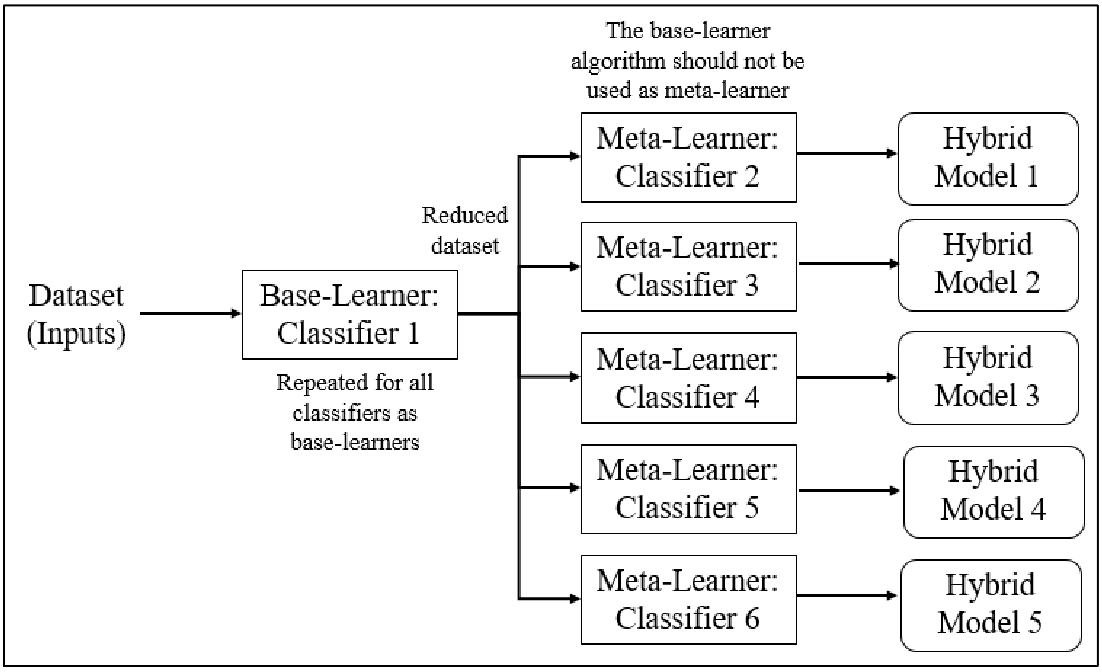


Figure 4.9. Generation of Ensemble Models using Stacked Generalization

Upon using multiple classifiers, the strategy of the stacking can said to be more sophisticated and sound compared to the voting (Wolpert, 1992). In voting, it is not clear which classifiers to trust during the voting process. This is a major drawback as some algorithms may be more suitable to certain datasets than others, but they still might be voted out. For example, suppose two of the three classifiers in voting make incorrect classifications. In such a case, the final classification would be incorrect. In stacking, the meta-learner replaces the voting mechanism and it identifies base classifiers that can be trusted (Witten et al., 2016). As a result, the bias is reduced; however, stacking adds extra parameters to be dealt with, extra variance due to having more than one training set (reduced training sets), and consequently, extra computational time (Alpaydın, 2010). However, in general, the stacked generalization is claimed to be a more suitable approach than voting (Witten et al., 2016).

When the classification accuracy of single classifiers contained in the ensemble model is as high as possible and classifiers are selected as diverse as possible, the ensemble model can outperform performances of single classifiers it contains. As it can be seen from Figure 4.9, the stacked generalization should not be used to combine classifiers of the same type (i.e. if the base-learner is Naïve Bayes, the meta-learner should not be Naïve Bayes) and instead, it should be applied to combine different classification techniques so that different classifiers can complement to each other (Alpaydın, 2010). Considering this, the best performing three single ML techniques are used as base-learners one by one and remaining algorithms are used as meta-learners in turns while same classifiers cannot be base-learner and meta-learner at the same time; such that, the algorithm with the best performance will be the base-learner and remaining algorithms (other than the base-learner) will be meta-learners one by one to generate five models. Then, the second best performing algorithm will be the base-learner and remaining algorithms will again be meta-learners one by one to generate five more models. This procedure will continue until all combinations are experimented, which means 15 ensemble models are experimented for dispute occurrence prediction, 15 ensemble models are experimented for potential compensation prediction, and 15 ensemble models are experimented for resolution method selection using the stacked generalization.

4.3.3. AdaBoost Algorithm

Boosting is an ensemble method that is based on generating a strong classifier from a number of sequentially iterated weak classifiers (Brownlee, 2018a). Boosting is a general method that can be used to improve the classification performance of any ML algorithm by reducing the error of the weak algorithm that generates classifiers with a performance little better than random chance (Freund and Schapire, 1996). This is achieved by generating base-learners that complement each other by training the algorithm with a special emphasis on misclassified instances of the previous learner (Alpaydın, 2010). More simply, an initial model is trained with the unweighted (initial) training set; then, a second model is trained in attempt to correct misclassifications of the first model with assigning weights to instances on the initial training set. This process continues until the training set has no errors or a predefined value for maximum number of classifiers is reached (Brownlee, 2018a). During this process, the weak learner is repeatedly trained on various distributions of the training data and resulting classifiers are combined to generate a single and stronger classifier (Freund and Schapire, 1996).

The AdaBoost, which is the short form of adaptive boosting, is a boosting algorithm that can combine an arbitrary number of base-learners by using the same training set repeatedly (Alpaydın, 2010). In original version, 'AdaBoost.M1' algorithm, which is developed by Freund and Schapire (1996), is the one for classification tasks specifically.

The AdaBoost algorithm assigns equal weights to all instances in the training set at the beginning. Then, an ML algorithm is used as a classifier (base-learner) for this training set with some classification errors. The classification error is calculated using instance weights such that the sum of weights of misclassified instances is divided by the total weight of all instances (Witten et al., 2016).

The AdaBoost algorithm adjusts weights of instances according to classification errors such that weights are increased for misclassified instances and decreased for correctly classified ones. As a result, instances that are correctly classified by the classifier are assigned lower weights and considered as easy instances. Meanwhile, misclassified instances by the classifier are assigned higher weights and considered as hard instances (Freund and Schapire, 1996).

Resulting from the weighting rationale, the AdaBoost algorithm can focus on the correct classification of misclassified instances (Freund and Schapire, 1996). In weight assignments, the AdaBoost algorithm uses the following formula that generates a value between '0' and infinity (Witten et al., 2016):

$$weight = -log\frac{(classifier\ error)}{(1 - classifier\ error)} \qquad [51]$$

After adjusting the weights, a new classifier is developed for this new weighted training set. Similarly, there will be classification errors again. Then, the weighting process is repeated and the new training set is classified again. After each iteration, weights are changed for achieving correct classifications and each developed classifier complements to each other for this purpose (Witten et al., 2016).

At the end of training process, the AdaBoost acts like a voting method using a weighted voting strategy with assigned weights being proportional to the classification accuracy of base-learners on the training set (Alpaydın, 2010). In other words, the AdaBoost algorithm combines weak classifiers by summing their probabilistic predictions (Freund and Schapire, 1997).

In short, the advantage of the AdaBoost algorithm is to derive a strong classifier out of several weak classifiers (Freund and Schapire, 1996). However, the boosting process might perform poorly on test set (new instances) if individual classifiers are too complex for the amount of available training data (Witten et al., 2016). An experimental study proved that when the weak algorithm creates simple classifiers, the boosting method performs significantly better than bagging (Freund and Schapire, 1996). Unfortunately, if classifiers are complex, the AdaBoost model may have an overfitting problem (Alpaydın, 2010) and an overfitted AdaBoost model may generate a classification performance worse than the single classifier built from the same training set (Witten et al., 2016). In addition, boosting is vulnerable to noise and outliers (Alpaydın, 2010).

In the light of these, this research utilized the AdaBoost algorithm to generate ensemble models with the aim of enhancing classification performances of the single algorithms used.

In short, with an aim to generate an understanding, this chapter reviewed binary and multiclass data classification problems, basic concepts in ML domain, theoretical background related to utilized single and ensemble ML techniques in this research along with reasons of selecting these algorithms, measures of classification performance, and importance of attribute elimination in ML domain. Following this explanatory chapter, in the next (fifth) chapter, data classification via alternative ML algorithms will be performed on the finalized prediction models developed in Chapter 3. In other words, finalized dispute occurrence prediction, potential compensation prediction, and resolution method selection models will be experimented by alternative ML algorithms in order to reveal the best classification performance for the corresponding dataset.

CHAPTER 5

RESULTS OF DATA CLASSIFICATION EXPERIMENTS

Results obtained from binary data classification for dispute occurrence prediction as well as multiclass data classification for potential compensation prediction and resolution method selection are given in this chapter. The analysis are performed by using the WEKA workbench version 3.8.3 and the tool is introduced in the next section. Then, configurations in WEKA to test utilized single ML algorithms and detailed binary classification results for dispute occurrence prediction are given. It will be followed by configurations and results of ensemble ML algorithms for the same dataset. Similarly, configurations of utilized single and ensemble ML algorithms and detailed multiclass classification results are given for potential compensation prediction and resolution method selection respectively. Moreover, adjustments to algorithms to enable multiclass classification is explained. Finally, best classifiers will be selected among numerous tested algorithms by comparing them with respect to various measures of classifier performance. The selected classifier for binary classification will be the final model for dispute occurrence prediction. Similarly, classifiers with the best performance will be determined as final models for potential compensation prediction and resolution method selection based on results of multiclass classification experiments.

5.1. THE WEKA WORKBENCH

The WEKA workbench is an open-source Java based application for data mining that is produced by the University of Waikato in New Zealand (Drazin and Montag, 2012). The name stands for Waikato Environment for Knowledge Analysis and the software is under the GNU General Public License as freely available (Girja Sharma et al., 2013).

The WEKA workbench is a collection of ML algorithms for data mining tasks that contain tools for data preprocessing, classification, regression, clustering, association rule mining, and visualization (Frank et al., 2016). WEKA is the state-of-the-art tool for applying ML algorithms to real world problems. Besides numerous algorithms it contains, it is also possible to develop new ML schemes using this tool (Patel et al., 2014). Moreover, it can be used in various operating systems such as Windows, Linux, and Mac (Sobhana, 2014).

WEKA can be used to apply an ML algorithm to a dataset for analyzing the data to obtain more insight. It can also be used for making predictions using generated classifiers. Moreover, it is possible to apply various algorithms on a dataset to compare performances and to select the most appropriate one for predictions (Witten et al., 2016).

Algorithms available in WEKA can be applied directly to datasets (Patel et al., 2014). This can be achieved by using a graphical-user-interface (GUI) or by calling the Java code using the Java library for WEKA (Sobhana, 2014). The WEKA GUI presents several applications; (1) the Explorer, (2) the Experimenter, (3) the Knowledge Flow, (4) the Workbench, and (5) the Simple Command Line Interface (CLI).

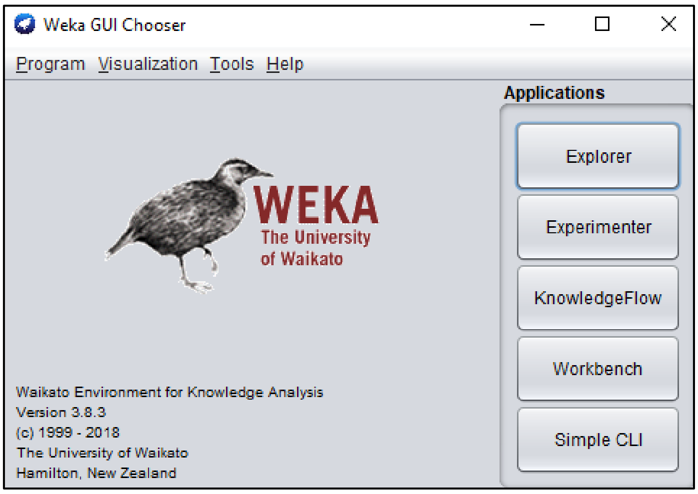


Figure 5.1. WEKA GUI Chooser and the Available Applications

The Explorer GUI, which gives access to all available facilities via menu selections and form filling, is the easiest way to use WEKA (Sobhana, 2014). However, due to memory limitations, the Explorer can be used for small to medium sized problems (Witten et al., 2016). The Experimenter GUI can be used to identify which algorithms and parameter settings give the best performance for the reviewed problem thanks to the capability of experimenting a wide variety of algorithms and settings simultaneously on the same dataset. The Knowledge Flow GUI, which is the best option for large datasets, is designed for developing a data stream by connecting components such as datasets, preprocessing tools, ML algorithms, evaluation methods, and visualization modules. The Workbench GUI is the combination of the Explorer, the Experimenter, and the Knowledge Flow in one application. The Simple CLI gives access to users for calling Java codes using the Java Library for WEKA (Witten et al., 2016).

The data is uploaded into WEKA in attribute-relation file format (ARFF), which is composed of pre-labeled diverse parts for attribute names, attribute types, values (categories), and the data itself (Patel et al., 2014). Datasets in other formats (i.e. MS Excel) should be converted to ARFF to be used in WEKA. An example ARFF file is given in Figure 5.2 that has various parts including dataset preprocessing part ('relation' part), attribute names and categories ('attribute' part), and data part.

In short, the WEKA workbench can be used for data preprocessing (i.e. filtering, standardization, normalization), attribute elimination (i.e. information gain, subset evaluation), data classification (i.e. Naïve Bayes, SVM, decision trees), regression (i.e. LR, SVM), clustering (i.e. k-means), association rules (i.e. Apriori algorithm), and data visualization (Sobhana, 2014).

In this thesis study, WEKA version 3.8.3, which is the latest stable version of the tool, is used for data classification tasks in dispute occurrence and potential compensation prediction problems along with resolution method selection problem.

```
@relation 'Dispute Prediction Dataset - Type-Status -
Delivery - Contract - CounterXP-
weka.filters.unsupervised.attribute.Remove-
R1,5,8-10,13-15,18,20,22-24,26'

@attribute ContractValue {1,2,3,4}
@attribute PlannedDuration {1,2,3,4,5}
@attribute Extension {1,2,3,4,5}
@attribute ConstructionType {1,2,3,4,5,6,7,8,9}
@attribute ConstructionStatus {1,2,3}
@attribute ContractorType {1,2,3}
@attribute ProjectDeliverySystem {1,2,3,4}
@attribute ContractType {1,2,3,4,5,6}
@attribute CounterSkills {1,2,3,4,5,6,7,8,9,10}
@attribute CounterResponse {1,2,3,4,5,6,7,8,9,10}
@attribute CounterXP {1,2,3,4,5,6,7,8,9,10}
@attribute CounterCoordination {1,2,3,4,5,6,7,8,9,10}
@attribute Communication {1,2,3,4,5,6,7,8,9,10}
@attribute Interrelations {1,2,3,4,5,6,7,8,9,10}
@attribute Dispute {0,1}

@data
2,1,1,4,1,1,3,2,8,7,10,8,8,7,1
4,1,1,6,1,1,3,6,5,5,4,3,3,3,1
4,1,1,5,2,1,2,2,1,1,3,3,1,1,1
2,2,2,4,1,1,3,2,4,7,7,6,9,6,1
1,4,3,1,1,1,2,4,7,7,7,7,7,8,1
2,2,1,8,1,1,3,6,9,9,9,9,9,9,0
3,2,1,8,1,1,3,5,6,5,7,5,9,8,1
3,2,1,8,1,1,3,6,10,9,8,8,9,10,1
3,2,1,8,1,1,3,6,5,9,7,7,5,4,1
2,2,1,4,1,1,3,6,6,8,9,8,6,6,1
4,1,1,1,2,1,2,6,7,5,8,7,10,10,0
1,4,2,7,1,1,3,5,5,5,5,5,8,8,1
4,1,1,4,1,1,3,6,5,5,5,3,7,7,1
4,1,1,4,1,1,3,6,2,2,3,2,2,3,1
4,1,2,4,1,1,3,6,8,8,9,8,8,8,0
2,3,1,4,1,1,1,6,5,8,5,5,6,7,1
4,1,1,8,1,1,2,6,2,1,1,1,3,3,1
3,3,3,7,1,1,2,6,7,8,5,2,7,7,1
1,3,3,5,1,3,3,3,4,7,4,3,5,7,1
```

Figure 5.2. An Example ARFF File Format used in WEKA

5.2. BINARY CLASSIFICATION PROBLEM OF DISPUTE OCCURRENCE PREDICTION

The binary classification problem of dispute occurrence prediction is solved by single and ensemble ML algorithms. The utilized single algorithms are (1) Naïve Bayes, (2) kNN, (3) J48, (4) MLP, (5) Polynomial kernel SVM, and (6) Gaussian RBF kernel SVM. Meanwhile, utilized ensemble techniques are (1) voting, (2) stacked generalization, and (3) AdaBoost algorithm. Configurations for each algorithm in WEKA and corresponding classification results are given in this section.

The dataset for dispute occurrence prediction involves 108 instances obtained from real construction projects. The classification problem here is whether the project will encounter disputes or not. In pursue of such a classification, this thesis study classifies the dataset using stratified 10-fold cross-validation with 10 repeats. Results for each run will be given. Then, the average accuracy for 10 runs are presented as the final classifier performance with a 95% CI constructed around average values. Thus, average accuracy values with upper and lower bounds will represent the final performance of each classifier under review.

5.2.1. Binary Classification for Dispute Occurrence Prediction Using Single ML Algorithms

The WEKA configuration details of each single ML algorithm and the obtained binary classification results are given in this section starting with the Naïve Bayes algorithm, which will be followed by the kNN, J48, MLP, polynomial kernel SVM, and Gaussian RBF kernel SVM, in the given order.

5.2.1.1. The Naïve Bayes Algorithm and its Configuration in WEKA

In WEKA version 3.8.3, the Naïve Bayes algorithm is contained in 'weka.classifiers.bayes.NaiveBayes' class. This class can handle binary and categorical attributes along with missing values (Frank et al., 2016). This classifier can also handle numeric attributes by assuming a distribution (Brownlee, 2018b)

The Naïve Bayes algorithm in WEKA has several options that can be adjusted. Options that gave the best classification performance are selected. The WEKA configuration for the Naïve Bayes is given in Figure 5.3. Critical settings here are the 'useKernelEstimator' setting and the 'useSupervisedDiscretization' setting. Both options can be enabled by setting it to 'True' or disabled by setting it to 'False'. For remaining settings, default values in WEKA are utilized.

The Naïve Bayes classifier can assume complex distributions such as kernel density functions rather than assuming a Gaussian distribution for the numeric data. Kernel estimators may result in a better match with the distribution of attributes in the dataset (Brownlee, 2018b). Consequently, the Naïve Bayes algorithm can achieve better classification performance using kernel estimators (Amin and Habib, 2015). However, this option does not change the classification performance since numeric attributes in reviewed datasets in this research were previously converted to categorical variables for computational reasons due to utilization of the Chi-Square tests. Therefore, in order to decrease computational complexity, this option is 'False' and disabled. The 'useSupervisedDiscretization' option can be 'True' when numeric attributes are used. This option will automatically convert numeric attributes to categorical ones in WEKA (Brownlee, 2018b). However, there are no numeric attributes in reviewed datasets and therefore, there is no need for enabling this option.

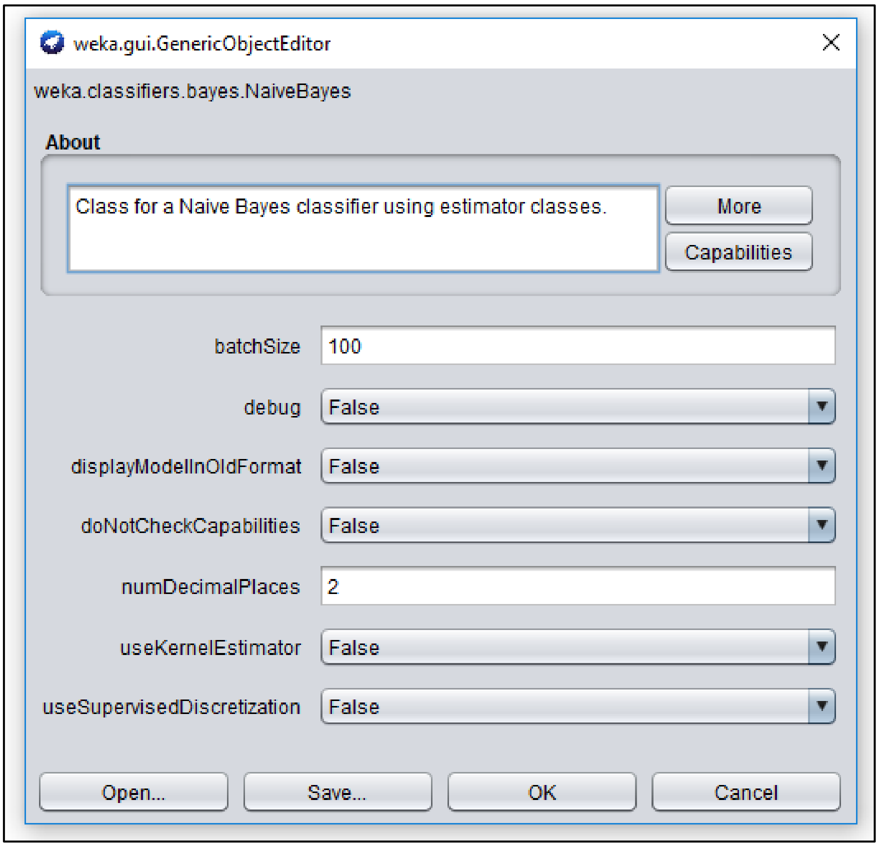


Figure 5.3. The Naïve Bayes Classifier Configuration in WEKA

Naïve Bayes classification results using the mentioned configuration are given in the next section.

5.2.1.2. Results from the Naïve Bayes Algorithm

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.1), Naïve Bayes classifiers have an average classification accuracy of '87.50%' with lower and upper bounds (86.60% - 88.40%) within 95% CI. In other words, the Naïve Bayes algorithm predicts the dispute occurrence in construction projects with an average success rate of '87.50%'.

The average for Kappa statistic value is '0.728' that can be interpreted as a substantial agreement between predicted and actual values in the dataset. Disputed projects are considered as positive instances and undisputed projects are considered as negative instances. The average precision value that indicates the positive predicting power of Naïve Bayes classifiers is '0.912'. The average sensitivity (recall) value is '0.893'. In other words, the success of the Naïve Bayes algorithm in identifying disputed (positive) projects is '89.3%'. Similarly, the average specificity value is '0.843' that indicates the Naïve Bayes algorithm achieved '84.3%' success in identifying undisputed (negative) projects. The average AUROC value is '0.953' that indicates the ability of the Naïve Bayes algorithm to avoid misclassifications. Considering that the ideal AUROC value is '1', the obtained AUROC value is almost perfect.

Table 5.1. 10-Times 10-Fold Cross-Validation Results for the Naïve Bayes

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes	Accuracy(%)	87.04	87.96	87.04	88.89	87.96	88.89	88.89	85.19	86.11	87.04	87.50
	Kappa	0.716	0.741	0.719	0.756	0.738	0.759	0.759	0.679	0.697	0.719	0.728
	Precision	0.900	0.925	0.912	0.914	0.913	0.926	0.926	0.897	0.899	0.912	0.912
	Recall	0.900	0.886	0.886	0.914	0.900	0.900	0.900	0.871	0.886	0.886	0.893
	Specificity	0.816	0.868	0.857	0.842	0.842	0.868	0.868	0.816	0.816	0.842	0.843
	AUROC	0.945	0.957	0.950	0.955	0.956	0.954	0.956	0.954	0.950	0.953	0.953

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	625	75
Undisputed	60	320

The corresponding confusion matrix can be seen under the results. Although the dataset involves 108 instances, the confusion matrix involves 1080 instances since there are 10 repeats with the same algorithm.

5.2.1.3. The KNN Algorithm and its Configuration in WEKA

In WEKA version 3.8.3, the kNN algorithm is contained in ‘weka.classifiers.lazy.IBk’ class. This class can handle binary, categorical, and numeric attributes. However, numeric attributes should be rescaled by normalization. Moreover, if the data has a Gaussian distribution, it should be standardized (Frank et al., 2016).

The kNN algorithm in WEKA has several settings that can be adjusted. Settings that gave the best classification performance are selected. The WEKA configuration for the kNN is given in Figure 5.4.

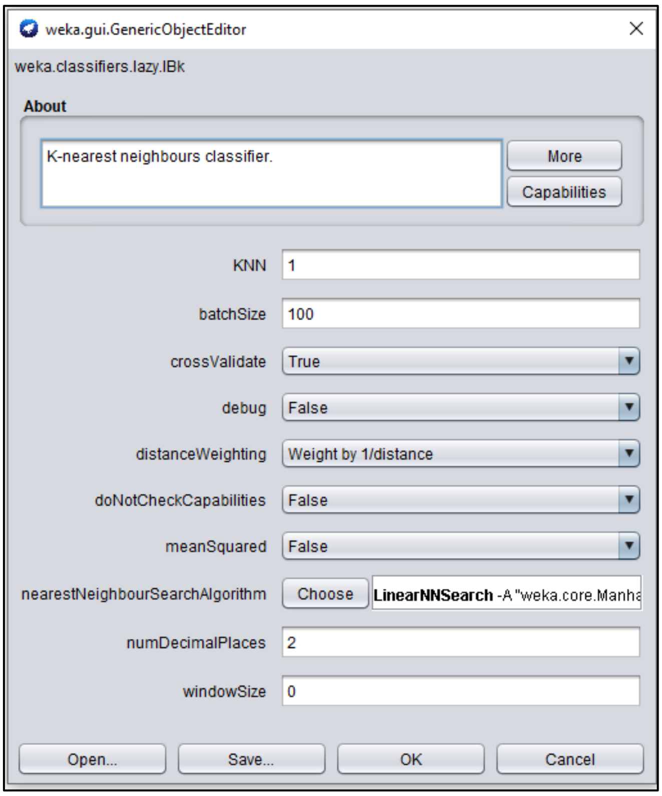


Figure 5.4. The KNN Classifier Configuration in WEKA

As mentioned earlier (Section 4.2.2), the value of k should normally be selected using cross-validation (Thakkar et al., 2011). Therefore, the 'crossValidate' setting, which is set to 'False' by default, should be adjusted as 'True'.

The algorithm has a 'distanceWeighting' setting that can adjust distance values between instances with a factor (1 / distance) or (1 – distance). There is also an option for no distance setting. The best performance is obtained from (1 / distance) weighting that assigns weights (votes) for k neighboring instances during classification of a new instance.

The 'meanSquared' setting determines whether to use the mean squared error rather than the mean absolute error when doing cross-validation for regression problems. Thus, it should be considered for regression problems only. For classification problems, this configuration is set as 'False'.

The 'nearestNeighborSearchAlgorithm' involves distance measurement functions. Among these functions 'LinearNNSearch' option involves distance measures such as Chebyshev, Euclidean, Manhattan, and Minkowski distance measures. As mentioned in Section 4.2.2, the Manhattan distance measure should be selected in this research. Indeed, among various trials, the Manhattan distance measure gave the best classification performance.

The most important parameter is the k value that can be adjusted by changing the value in the 'KNN' setting. In this thesis study, the k parameter is optimized using an external algorithm from the WEKA library. This algorithm is the 'cross-validation parameter selection' algorithm that is capable of determining the optimum value for a parameter using cross-validation. The cross-validation parameter selection algorithm is contained in 'weka.classifiers.meta.CVParameterSelection' class. The WEKA configuration for this algorithm is given in Figure 5.5. The two settings that should be organized for determination of the optimum k parameter are the 'CVParameters' setting and the 'classifier' setting. In the 'classifier' setting, the kNN algorithm should be selected as the classifier with the configuration in Figure 5.4. The 'CVParameters' setting is empty by default that is shown as ‘0.java.lang.String’. The user should enter the search range for the k value manually. The search range for the optimum k value for dispute occurrence prediction problem is between ‘1’ and ‘100’. Remembering that there are 108 instances in dispute occurrence prediction dataset, larger k values are not considered. By clicking on the ‘CVParameters’ option, a new tab opens for defining the parameter that will be searched by the algorithm. The user defines the parameter to be searched and in this case, it is ‘K’. Then, the range to be searched is defined for this case between ‘1’ and ‘100’. In order to try all integer values in this range, the number of steps is defined as ‘100’. Thus, the algorithm will divide the range into ‘100’ steps and try all values one by one to perform the classification. The k value that gives the best classification performance will be presented to the user as the optimum k value.

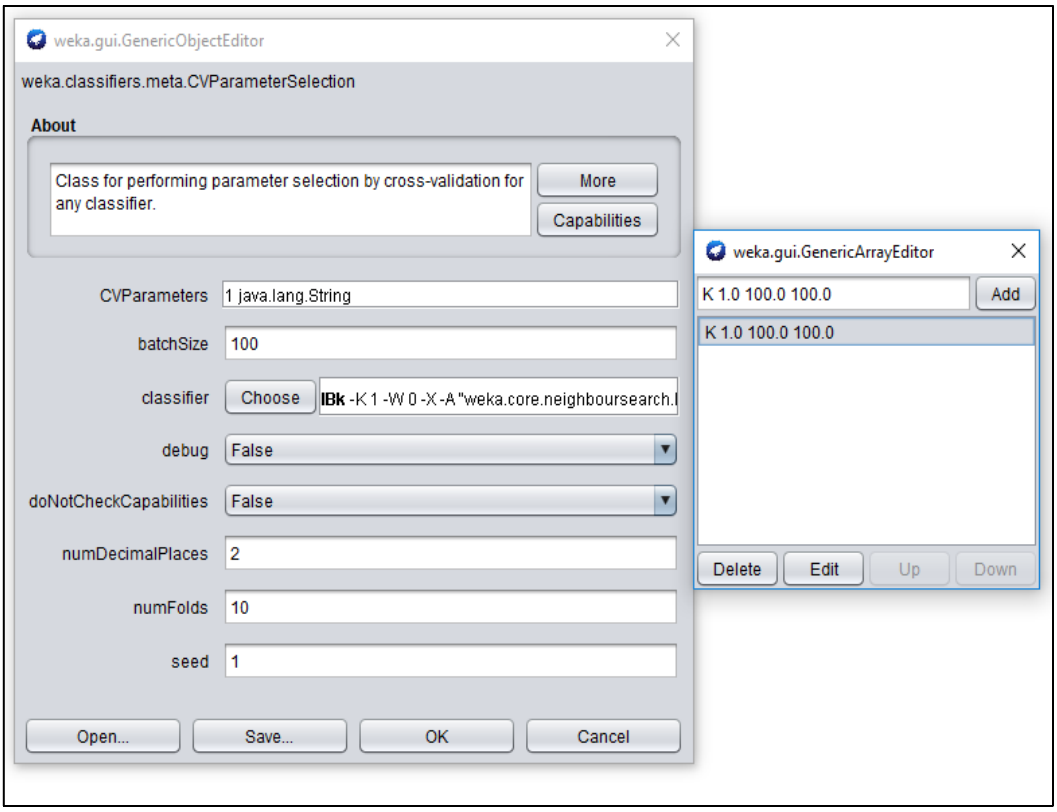


Figure 5.5. The CVParameterSelection Configuration for the KNN in WEKA

5.2.1.4. Results from the kNN Algorithm

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.2), the optimum k value is identified as ‘k = 3’. KNN classifiers have an average classification accuracy of ‘87.69%’ with lower and upper bounds (86.65% - 88.72%) within 95% CI. In other words, the kNN algorithm predicts the dispute occurrence in construction projects with an average success rate of ‘87.69%’.

The average for Kappa statistic value is ‘0.737’ that shows a substantial agreement. The average precision value that indicates the positive predicting power of kNN classifiers is ‘0.931’. The average sensitivity (recall) value is ‘0.874’ that means the success of the kNN algorithm in identifying disputed projects is ‘87.4%’. Similarly, the average specificity value is ‘0.881’ showing the kNN algorithm achieved ‘88.1%’ success in identifying undisputed projects. The average AUROC value is ‘0.928’ that is very close to the ideal value.

Table 5.2. 10-Times 10-Fold Cross-Validation Results for the KNN Algorithm

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN	Accuracy(%)	87.96	88.89	86.11	87.96	86.11	90.74	87.04	87.96	86.11	87.96	87.69
	Kappa	0.741	0.762	0.708	0.744	0.704	0.802	0.723	0.741	0.704	0.741	0.737
	Precision	0.925	0.939	0.937	0.938	0.923	0.955	0.924	0.925	0.923	0.925	0.931
	Recall	0.886	0.886	0.843	0.871	0.857	0.900	0.871	0.886	0.857	0.886	0.874
	Specificity	0.868	0.895	0.895	0.895	0.868	0.921	0.868	0.868	0.868	0.868	0.881
	AUROC	0.883	0.922	0.940	0.930	0.925	0.948	0.922	0.937	0.934	0.935	0.928

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	612	88
Undisputed	45	335

5.2.1.5. The J48 Algorithm and its Configuration in WEKA

In WEKA version 3.8.3, the J48 algorithm is contained in ‘weka.classifiers.trees.J48’ class. This class can generate pruned or unpruned decision trees and it can work with binary, categorical, and numeric attributes. Moreover, it is capable of handling missing values (Frank et al., 2016).

Settings in the J48 algorithm involve the type of pruning, the confidence threshold for pruning, and the minimum number of instances in leaf nodes. In default configuration of the J48 algorithm, the pruning is on and subtree raising is the preferred pruning technique. These settings are kept as default values such that the 'unpruned' option is set to 'False', the 'subtreeRaising' option is set to 'True', and the 'reducedErrorPruning' option is set to 'False'. When reduced error pruning is set to 'False', the 'numFolds' option, which determines the number of folds for reduced error pruning, becomes unnecessary. In addition, the 'useLaplace' option is set to 'True' that smoothed the counts at leaves based on Laplace (Lavesson and Davidsson, 2006). The configuration for the J48 algorithm in WEKA can be seen in Figure 5.6.

As it can be seen from Figure 5.6, the J48 decision tree has numerous parameters. However, only two of them affects the amount of pruning, which are the 'confidenceFactor' and the 'minNumObj' settings. The 'confidenceFactor' setting determines the confidence factor for pruning, or in other words, the size of the tree. (Molina et al:, 2012). The confidence factor parameter is associated with the effectiveness of the post-pruning. The lower confidence factor corresponds to a lesser confidence to the training data such that the error estimate for each node increases. Error estimates are increased by penalizing the nodes with few instances since confident assumptions related to classification errors on these nodes cannot be made. Consequently, the likelihood of pruning of such nodes increases in pursue of a more stable tree structure (Drazin and Montag, 2012). In short, the smaller confidence factor values result in more pruning and the default value for the confidence factor in WEKA is '0.25' (Witten et al., 2016). On the other hand, the 'minNumObj' setting determines the minimum number of instances in leaf nodes and the default value is '2' (Molina et al., 2012). Considering that the pruning process is a main determinant on the performance of the J48 algorithm, the optimum pruning should be performed (Lavesson and Davidsson, 2006). The confidence factor should be tested with different values when developing trees and the most appropriate value for the reviewed dataset should be determined (Drazin and Montag, 2012). This also applies for the minimum number of instances in leaf nodes.

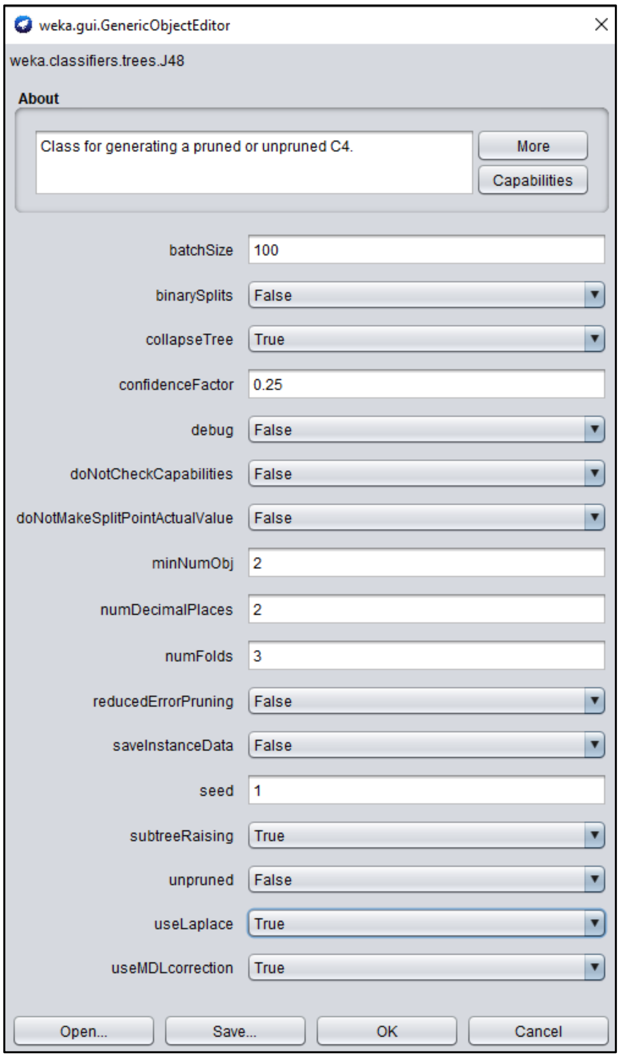


Figure 5.6. The J48 Classifier Configuration in WEKA

In order to perform an optimum pruning process, both parameters should be optimized simultaneously. It is not known which confidence factor and minimum number of instances in leaf nodes values generate the best results for the reviewed data classification problem. For this reason, an external algorithm, which is called the 'Grid Search Algorithm', is utilized in the WEKA tool as suggested in Hsu et al. (2003).

The grid search algorithm is capable of optimizing values of two different parameters at the same time by trying various value pairs of these parameters until the best cross-validation accuracy is achieved (Hsu et al., 2003). In grid search, parameters are varied with a fixed step size through a wide range of values and the cross-validation classification performance of every combination couple is assessed (Friedrichs and Igel, 2005). Although grid search is an exhaustive parameter search that requires significant computational time, this time requirement is not much more than requirements of other advanced methods. Moreover, grid search can be parallelized when search parameters are independent, while many advanced methods are hard to parallelize as they depend on iterative processes (Hsu et al., 2003). However, as stated earlier, grid search is only suitable for the search of two parameters in WEKA. When the best combination of parameters is determined, the whole training set is trained for one final time in order to develop the final classifier.

In WEKA, the grid search algorithm is contained under the class 'weka.classifiers.meta.GridSearch'. Configuration details of the grid search algorithm is given in Figure 5.7. In grid search, the J48 classifier should be selected under the 'classifier' setting. Then, the two parameters that will be optimized should be entered to 'XProperty' and 'YProperty' options using the names in WEKA tool such that 'confidenceFactor' and 'minNumObj' should be written to 'XProperty' and 'YProperty', respectively. The search range for both parameters should be determined. There are various suggestions in the literature for confidence factor and minimum number of instances in leaf nodes value search ranges. Lavesson and Davidsson (2006) suggested an extensive search in a small range between '0.02' and '0.5' with a step size of '0.02' for confidence factor and a range between '1' and '4' with a step size of '1' for minimum number of instances in leaf nodes. Meanwhile, Reif et al. (2011) suggested a less extensive search in a wide range between '0.05' and '5' to be searched in '10' steps for confidence factor and a range between '1' and '100' to be searched in '10' steps for minimum number of instances in leaf nodes.

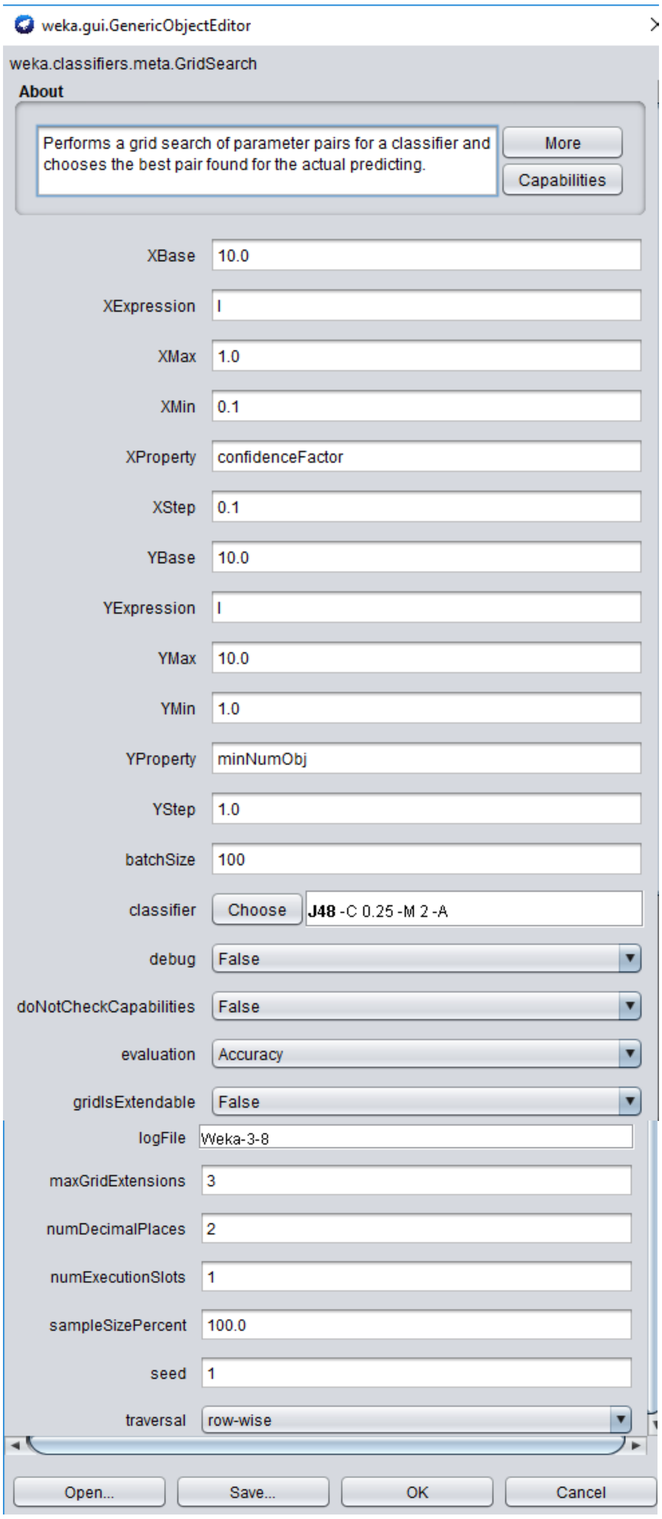


Figure 5.7. The Grid Search Configuration for the J48 Algorithm in WEKA

In this thesis study, a moderately wide range for confidence factor is selected between '0.1' and '1' with a step size of '0.1'. Considering the limited amount of instances in reviewed datasets in this research, the minimum number of instances in leaf nodes have a search range between '1' and '10' with a step size of '1'. The evaluation metric for parameter selection is the classification accuracy as stated in the 'evaluation' option as 'Accuracy'.

5.2.1.6. Results from the J48 Algorithm

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.3), J48 classifiers have an average classification accuracy of '88.98%' with lower and upper bounds (87.26% - 90.70%) within 95% CI. In other words, the J48 algorithm predicts the dispute occurrence in construction projects with an average success rate of '88.98%'.

The average for Kappa statistic value is '0.761' that shows a substantial agreement. The average precision value that indicates the positive predicting power of J48 classifiers is '0.927'. The average sensitivity (recall) value is '0.901' that means the success of the J48 algorithm in identifying disputed projects is '90.1%'. Similarly, the average specificity value is '0.868' showing the J48 algorithm achieved '86.8%' success in identifying undisputed projects. The average AUROC value is close to the ideal case as it is equal to '0.947'.

Table 5.3. 10-Times 10-Fold Cross-Validation Results for the J48 Algorithm

Classifier	**Performance Measure**	**Run Number**										**Avg.**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
J48	Accuracy(%)	87.96	85.19	87.96	90.74	92.59	90.74	86.11	91.67	87.96	88.89	88.98
	Kappa	0.731	0.687	0.741	0.799	0.841	0.799	0.704	0.816	0.738	0.756	0.761
	Precision	0.890	0.922	0.925	0.941	0.970	0.941	0.923	0.930	0.913	0.914	0.927
	Recall	0.929	0.843	0.886	0.914	0.914	0.914	0.857	0.943	0.900	0.914	0.901
	Specificity	0.789	0.868	0.868	0.895	0.947	0.895	0.868	0.868	0.842	0.842	0.868
	AUROC	0.936	0.952	0.951	0.952	0.949	0.959	0.927	0.967	0.931	0.948	0.947

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	631	69
Undisputed	50	330

5.2.1.7. The MLP and its Configuration in WEKA

In WEKA version 3.8.3, MLP networks are contained in 'weka.classifiers.functions.MultilayerPerceptron' class. This class can work with binary, categorical, and numeric attributes. Moreover, it is capable of handling missing values (Frank et al., 2016). The MLP configuration in WEKA tool involves numerous properties to be adjusted. Controlling parameters that affect the MLP performance significantly are discussed in this section. These controlling parameters are the number of epochs, momentum, learning rate, and the number of hidden layers. Other settings are used with default values. The configuration for the MLP network in WEKA can be seen in Figure 5.8.

During the training process, the training data is repeatedly presented to perceptrons in the MLP network in order to adjust connection weights until the error becomes lower than a predefined threshold value or a predetermined number of epochs is reached. Each cycle of calculations through all training instances is called an epoch (Chau, 2007). The number of epochs to train through is adjusted by the 'trainingTime' setting in WEKA. The default value for the number of epochs in WEKA is '500'. The classification is performed by increasing the number of epochs to '1000'; however, results showed that the computation time increased to more than double of the previous computation time without remarkable improvements in the performance. Therefore, the research adhered to the default number of epochs value of WEKA.

The determination of weights in the MLP tends to be a slow process with long durations required for computations. Considering the need to update weights repeatedly, MLP models are associated with slow convergence (Alpaydın, 2010). However, the momentum term can help in achieving better performances in terms of duration when updating the weights. This is done by associating a small proportion of the previous update value from the previous iteration with the current (new) weight change (Witten et al., 2016). Thus, momentum is the term that increases the speed of learning (Sözgen, 2009). The MLP in WEKA tool presents a 'momentum' setting for this purpose with the default value of ‘0.2’. The momentum can take values greater than ‘0’ and smaller than ‘1’ (Goodfellow et al., 2016).

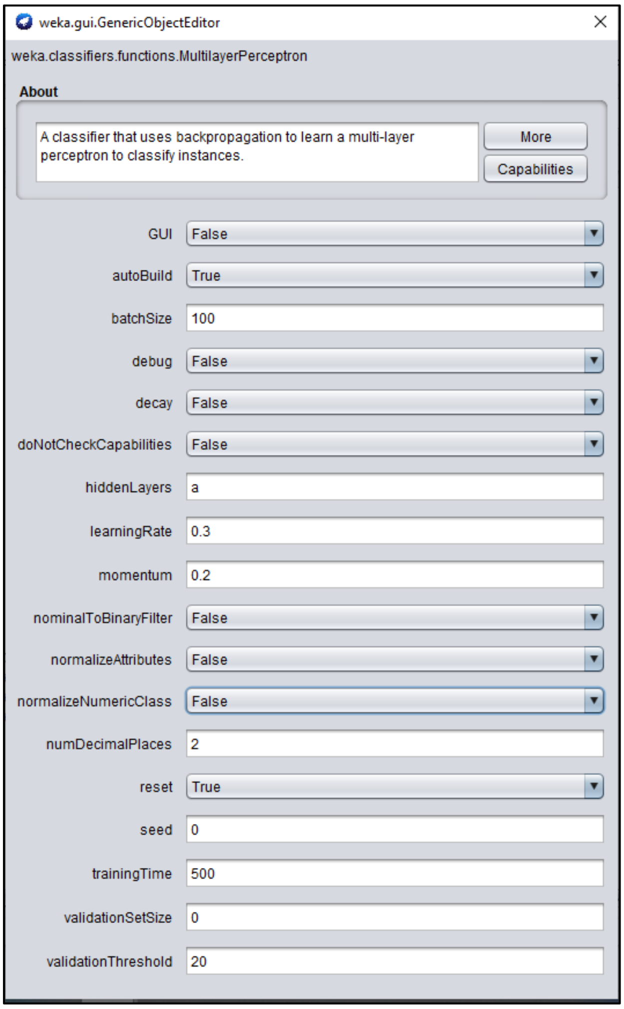


Figure 5.8. The MLP Classifier Configuration in WEKA

MLP networks are not limited to having one hidden layer, and more hidden layers with their own weights can be placed following the first hidden layer. However, as the number of hidden layers increases, it becomes more complex to analyze the MLP network and in practice, more than one hidden layers are rarely preferred (Alpaydın, 2010). In addition, as the number of layers increases, the number of perceptrons and the number of connections between these perceptrons increases. Consequently, the computational cost of the MLP network significantly increases. Therefore, in this thesis study, MLP networks with ‘0’, ‘1’, ‘2’, and ‘a’ number of hidden layers are experimented. In WEKA, the default value of hidden layers is ‘a’, which corresponds to the total number of inputs and outputs divided by ‘2’. For the dispute occurrence classification, the number of attributes are ‘14’ and the number of outputs is ‘1’ (only one output neuron for binary classification problems). According to this calculation, there should be ‘$(14 + 1) / 2 = 7.5 \approx 8$’ hidden layers for the dispute occurrence prediction MLP classifier. According to experiments with the mentioned number of hidden layers, the network giving the best classification performance is selected as the optimum number of hidden layers.

The amount of updates in the model at each epoch can be adjusted by determining the learning rate of the network (Brownlee, 2018b). The learning rate parameter determines the learning curve of the model with values between ‘0’ and ‘1’ (Sözgen, 2009). If the learning rate parameter is too large and the error function has several minima, the search may miss a minimum value; on the other hand, if this parameter is too small, the progress may be slow (Witten et al., 2016). For this reason, learning rate values are searched by an external cross-validation parameter selection algorithm (CVParameterSelection) to identify the model with the best cross-validation accuracy (the default is ‘0.3’). The configuration of the ‘CVParameterSelection’ for the MLP network can be seen in Figure 5.9. The two settings that should be organized for determination of the optimum learning rate parameter are the ‘CVParameters’ setting and the ‘classifier’ setting. In the ‘classifier’ setting, the MLP algorithm should be selected with the configuration in Figure 5.8. The user should enter the search range for the learning rate parameter manually to the ‘CVParameters’ setting. By clicking on the ‘CVParameters’ option, a new tab opens for defining the parameter that will be searched by the algorithm. The user defines the parameter to be searched and in this case, it is ‘L’. Then, the range to be searched is defined for this case between ‘0.1’ and ‘1’ within ‘10’ steps. The learning rate parameter that gives the best classification performance will be presented to the user as the optimum learning rate value.

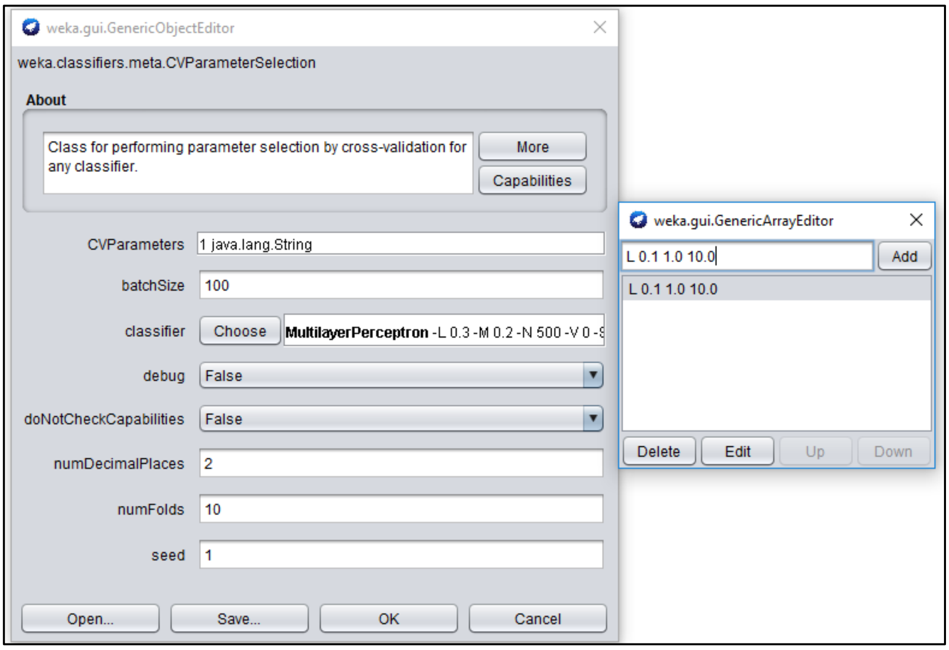


Figure 5.9. The CVParameterSelection Configuration for the MLP in WEKA

5.2.1.8. Results from the MLP

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.4), the best classification performance from MLP classifiers are obtained from the one with 'a' hidden layers (in this case 'a = 8'). MLP classifiers with 'a' hidden layers have an average classification accuracy of '83.52%' with lower and upper bounds (82.06% - 84.98%) within 95% CI. In other words, the MLP algorithm predicts the dispute occurrence in construction projects with an average success rate of '83.52%'.

The average for Kappa statistic value is '0.641' that shows a substantial agreement. The average precision value that indicates the positive predicting power of MLP classifiers is '0.879'. The average sensitivity (recall) value is '0.866' that means the success of the MLP algorithm in identifying disputed projects is '86.6%'. Similarly, the average specificity value is '0.779' showing the MLP achieved '77.9%' success in identifying undisputed projects. The average AUROC value is '0.894'.

Table 5.4. 10-Times 10-Fold Cross-Validation Results for the MLP

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
MLP	Accuracy(%)	86.11	80.56	84.26	84.26	82.41	84.26	84.26	79.63	84.26	85.19	83.52
	Kappa	0.690	0.571	0.657	0.653	0.621	0.661	0.665	0.553	0.657	0.679	0.641
	Precision	0.877	0.845	0.884	0.873	0.881	0.896	0.908	0.843	0.884	0.897	0.879
	Recall	0.914	0.857	0.871	0.886	0.843	0.857	0.843	0.843	0.871	0.871	0.866
	Specificity	0.763	0.711	0.789	0.763	0.789	0.816	0.842	0.711	0.789	0.816	0.779
	AUROC	0.910	0.879	0.895	0.901	0.878	0.915	0.906	0.853	0.908	0.894	0.894

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	606	94
Undisputed	84	296

5.2.1.9. The Polynomial Kernel SVM and its Configuration in WEKA

In WEKA version 3.8.3, the class ‘weka.classifiers.functions.SMO’ is used for the polynomial kernel SVM algorithm. The SMO algorithm, which is the short version of Sequential Minimal Optimization, refers to the optimization algorithm utilized in the SVM. This class can work with binary, categorical, and numeric attributes. Moreover, it is capable of handling missing values (Brownlee, 2018b).

The polynomial kernel SVM algorithm in WEKA has several settings that can be adjusted. Configuration that gave the best classification performance is selected. The WEKA configuration for the polynomial kernel SVM is given in Figure 5.10.

The first setting is the ‘buildCalibrationModels’ setting that can be ‘True’ or ‘False’. When it is set to ‘True’, the final SVM classifier generates probability estimates of class labels that is normally not available in the algorithm.

The most important settings for the polynomial kernel SVM are ‘c’ and ‘exponent’ settings. The ‘exponent’ setting is located inside the ‘kernel’ option after selecting ‘PolyKernel’ function. The default kernel function in SMO is the polynomial kernel. By clicking on the ‘kernel’ setting, a new window opens that enables selecting other kernel functions. In the SMO algorithm of WEKA, the kernel function can be either polynomial or RBF (Lavesson and Davidsson, 2006). In this case, the 'PolyKernel' option will be used.

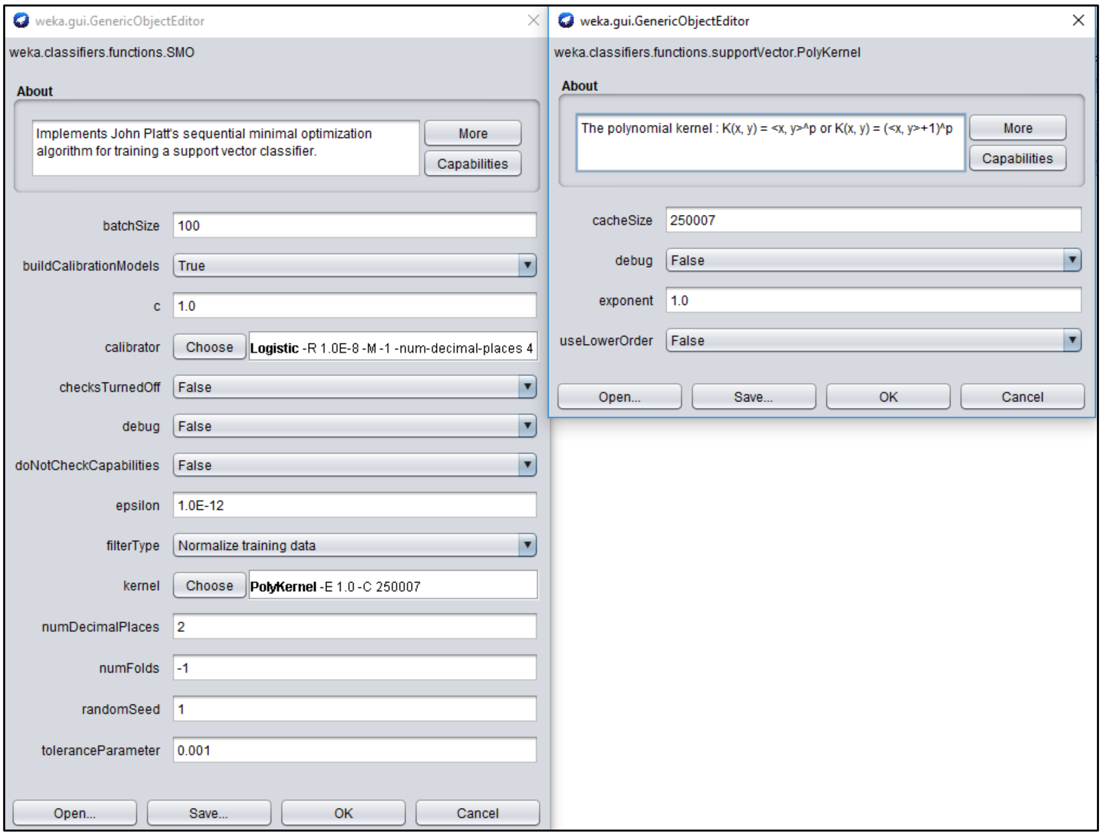


Figure 5.10. The Polynomial Kernel SVM Classifier Configuration in WEKA

The problem with the SVM algorithm is how to set the optimal penalty parameter C and the optimal kernel hyperparameters so that the prediction accuracy of the SVM classifier is maximized. As stated earlier, the penalty parameter C is selected beforehand and it determines the cost of violating the constraint (Dibike et al., 2001). In other words, penalty parameter C, which is also known as the soft margin constant, determines the trade-off between strictness of classification constraints and tolerated misclassifications (Lavesson and Davidsson, 2006). A penalty parameter C value of '0' means there will be no violations of the margin and more violation is allowed as this value increases (Brownlee, 2018a). In general, there is an optimum value for C hyperparameter (Drucker et al., 1999). Thus, the optimum value for the soft margin constant hyperparameter C should be determined. Similarly, the most appropriate kernel function along with optimum settings for kernel hyperparameters should also be identified beforehand (Cheng and Wu, 2009). In polynomial kernel SVM, the kernel hyperparameter to be optimized is the degree of the polynomial function, which is located in the ‘exponent’ setting. Therefore, there are two hyperparameters to be optimized simultaneously, which can be handled by the grid search algorithm, similar to the case in the J48 algorithm.

In practice, parameters for ML algorithms are usually determined by the grid search algorithm that searches a range of predefined values for two parameters at the same time with a fixed step size to present the best couple of values in terms of cross-validated classification accuracy (Friedrichs and Igel, 2005). Grid search is the simplest method for determining the penalty parameter C and kernel hyperparameters (Lin et al., 2008). Because of its simplicity and success in previous studies, the grid search algorithm is often preferred to other deterministic techniques (Mantovani et al., 2015). Moreover, Hsu et al. (2003) recommended using grid search algorithm to identify the optimal pair of C and kernel hyperparameters as the algorithm enables trials on various C and kernel hyperparameter pairs to reveal the pair that gives the best cross-validation accuracy.

In WEKA, the grid search algorithm is contained under the class ‘weka.classifiers.meta.GridSearch’. Configuration details of the grid search algorithm is given in Figure 5.11. In grid search, the SMO classifier with the configuration mentioned in Figure 5.10 should be selected under the ‘classifier’ setting. Then, the two hyperparameters that will be optimized should be entered to ‘XProperty’ and ‘YProperty’ options using the names in WEKA tool such that ‘C’ and ‘kernel.exponent’ should be written to ‘XProperty’ and ‘YProperty’, respectively.

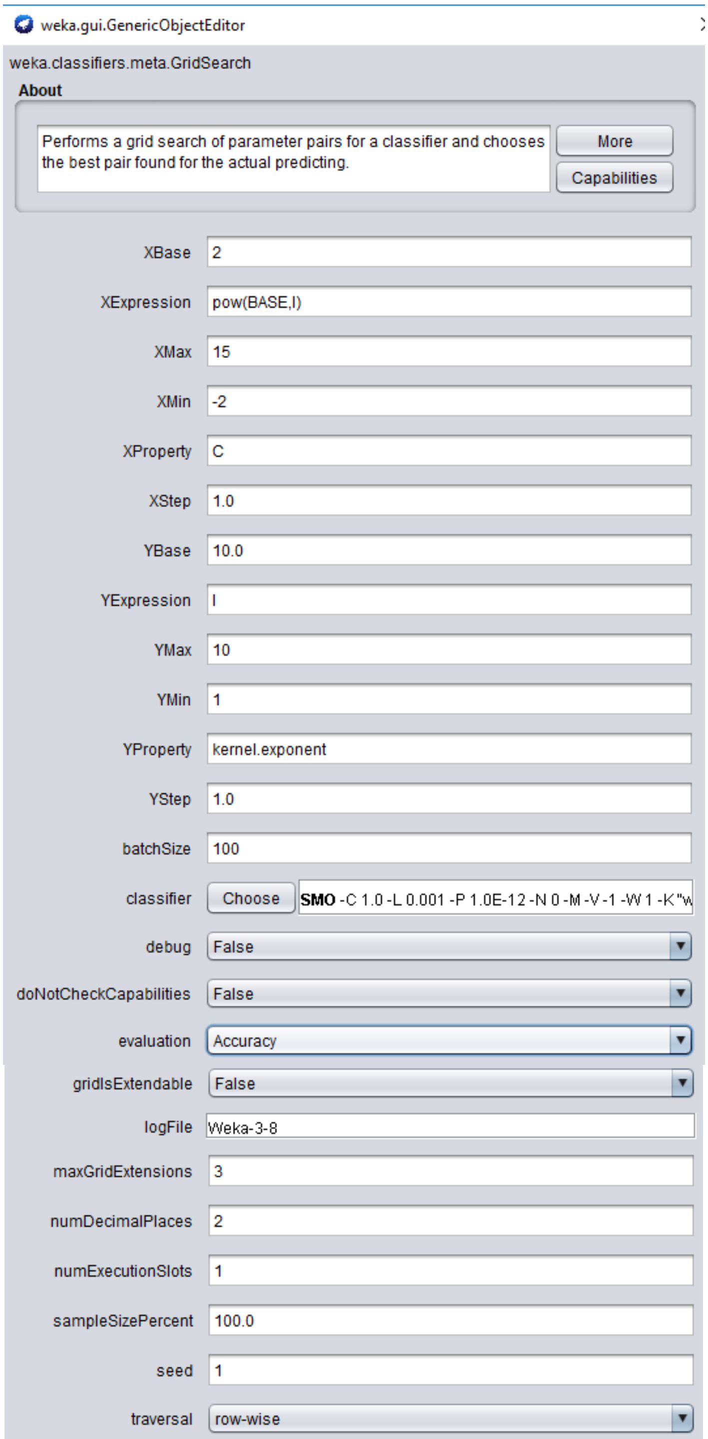


Figure 5.11. The Grid Search Configuration for the Polynomial Kernel SVM Algorithm in WEKA

The search range for both hyperparameters should be determined. There are various search range suggestions in the literature for penalty parameter C and polynomial kernel function degree. An excessively large range of values may waste computational power and on the other hand, a too small range can make it impossible to find the optimal solution (Lin et al., 2008). In an experimental study, it is revealed that in order to identify good parameters, trials with exponentially growing sequences, such that values of $\{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$ for C hyperparameter, can be beneficial (Hsu et al., 2003). In another study, the range for C hyperparameter values is determined as $\{2^{-2}, \ldots, 2^{12}\}$ (Reif et al., 2011). In another experimental study on SVM tuning, the range for C hyperparameter is given as $\{2^{-2}, \ldots, 2^{15}\}$ (Mantovani et al., 2015). In the light of these, by combining suggested ranges, the search range for penalty parameter C is selected as an exponentially growing sequence of $\{2^{-2}, \ldots, 2^{15}\}$ in this thesis study. Consequently, the mentioned search range for the C hyperparameter is defined in WEKA by setting the 'XMax' option to '15', the 'XMin' option to '-2', the 'XBase' option to '2', the 'XExpression' option to 'pow(BASE,I)', and the 'XStep' option to '1'.

There are no strict suggestions for the search range of polynomial kernel function degree. The simplest polynomial kernel function is the linear form that separates the data with a straight line and as the degree of the function increases, an increasingly meandering separation will be obtained (Brownlee, 2018a). Increasing the degree too much may result in an overfitted model. Thus, the polynomial kernel function degree search range is selected as $\{1, 2, \ldots, 10\}$. The mentioned search range for the degree of the polynomial kernel function is defined in WEKA by setting the 'YMax' option to '10', the 'YMin' option to '1', the 'YBase' option to '10', the 'YExpression to 'I', and the 'YStep' option to '1'.

Finally, the evaluation metric for hyperparameter selection is the classification accuracy as stated in the 'evaluation' option as 'Accuracy'. Default values of WEKA are kept for remaining settings.

5.2.1.10. Results from the Polynomial Kernel SVM

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.5), polynomial kernel SVM classifiers have an average classification accuracy of ‘89.91%’ with lower and upper bounds (88.85% - 90.96%) within 95% CI. In other words, the polynomial kernel SVM algorithm predicts the dispute occurrence in construction projects with an average success rate of ‘89.91%’.

The average for Kappa statistic value is ‘0.777’ that shows a substantial agreement. The average precision value that indicates the positive predicting power of polynomial kernel SVM classifiers is ‘0.917’. The average sensitivity (recall) value is ‘0.929’ that means the success of the polynomial kernel SVM algorithm in identifying disputed projects is ‘92.9%’. Similarly, the average specificity value is ‘0.845’ showing the polynomial kernel SVM achieved ‘84.5%’ success in identifying undisputed projects. The average AUROC value is ‘0.887’.

Table 5.5. 10-Times 10-Fold Cross-Validation Results for Poly. Kernel SVM

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM	Accuracy(%)	89.81	91.67	89.81	87.96	89.81	89.81	90.74	90.74	91.67	87.04	89.91
	Kappa	0.778	0.816	0.775	0.738	0.775	0.773	0.795	0.795	0.818	0.712	0.777
	Precision	0.928	0.930	0.915	0.913	0.915	0.904	0.917	0.917	0.942	0.889	0.917
	Recall	0.914	0.943	0.929	0.900	0.929	0.943	0.943	0.943	0.929	0.914	0.929
	Specificity	0.868	0.868	0.842	0.842	0.842	0.816	0.842	0.842	0.895	0.789	0.845
	AUROC	0.891	0.906	0.885	0.871	0.885	0.879	0.892	0.892	0.912	0.852	0.887

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	650	50
Undisputed	59	321

5.2.1.11. The RBF Kernel SVM and its Configuration in WEKA

In WEKA version 3.8.3, the class ‘weka.classifiers.functions.LibSVM’ is used for the Gaussian RBF kernel SVM algorithm. The LibSVM library for SVM applications is currently one of the most widely used tools that is capable of solving binary and multiclass classification problems with the goal of helping users to use SVM applications easily. Moreover, it can generate probability estimates of predictions (Chang and Lin, 2011). It is an enhancement to the SMO algorithm available in WEKA especially in terms of computational costs. The LibSVM package should be externally loaded into WEKA via package manager. This class can work with binary, categorical, and numeric attributes. Moreover, it is capable of handling missing values (Frank et al., 2016). The Gaussian RBF kernel SVM algorithm in WEKA has several settings that can be adjusted. The configuration that gave the best classification performance is selected. The WEKA configuration for the Gaussian RBF kernel SVM is given in Figure 5.12.

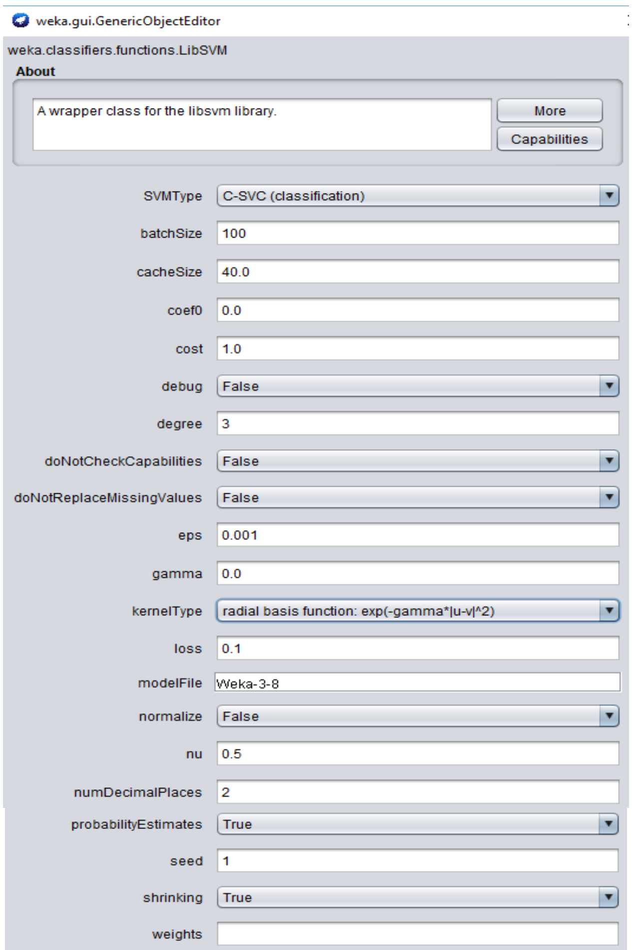


Figure 5.12. The RBF Kernel SVM Classifier Configuration in WEKA

In the LibSVM panel, the first setting is the ‘SVMType’ setting that should be set to ‘C-SVC’ for data classification problems. Secondly, the ‘kernelType’ should be set to RBF. After determining the kernel function as RBF kernel, using SVM means solving the problem of determining the optimal penalty parameter C (cost of constraint violation) and optimal kernel hyperparameters. These hyperparameter settings determine the classification performance of the SVM algorithm (Cheng and Wu, 2009). Hyperparameters to be optimized in the case of RBF kernel SVM are the penalty parameter C (here, cost) and the RBF kernel hyperparameter ‘gamma’ (Hsu et al., 2003). Similar to the case in the polynomial kernel SVM, the grid search algorithm is an appropriate tool to identify the best C and gamma for the RBF kernel SVM (Cheng and Wu, 2009). In their guideline study for support vector classification, Hsu et al. (2003) recommended using cross-validated grid search for determining C and gamma values by experimenting various pairs simultaneously to obtain the best cross-validation accuracy. When the best pair is identified, the whole training set is trained for the last time with identified values to present the final classifier.

Configuration details of the grid search algorithm is given in Figure 5.13. In grid search, the LibSVM classifier with the configuration mentioned in Figure 5.12 should be selected under the ‘classifier’ setting. Then, the two hyperparameters that will be optimized should be entered to ‘XProperty’ and ‘YProperty’ options using the names in WEKA tool such that ‘cost’ and ‘gamma’ should be written to ‘XProperty’ and ‘YProperty’, respectively.

The search range for both hyperparameters should be determined. There are various search range suggestions in the literature for penalty parameter C (cost) and RBF kernel function sigma (or gamma) value. The Gaussian RBF kernel hyperparameter sigma (please refer to Eq. [50]) is the spread parameter that affects the generalization performance of SVM classifiers significantly. Experimental studies proved that both large and small sigma values could cause poor generalization performances. When sigma goes to zero, all training instances are considered as support vectors. This will separate the problem perfectly for the training data however, new instances cannot be separated as the SVM classifier is overfitted to the training data. On the other hand, when sigma goes to infinity, the SVM cannot recognize new data points. Experiments proved that the classification accuracy is low at first and gets higher with the increasing sigma values however, the accuracy drops again as the sigma continues to increase. This shows that there is an optimum sigma (consequently, gamma) value for the classification task (Wang et al., 2003). The simulation of Wang et al. (2003) showed that sigma values between '0.1' and '20' is an appropriate search range. Hsu et al. (2003) recommended using a sequentially growing range for the gamma hyperparameter such that $\{2^{-15}, 2^{-13}, \ldots, 2^{3}\}$. Similar to this, Reif et al. (2011) used values of $\{2^{-10}, 2^{-9}, \ldots, 2^{4}\}$ for gamma hyperparameter. The common point in all these suggestions is that large gamma values are not preferred for trails. In the light of these, this thesis study searched the range of $\{2^{-15}, 2^{-14}, \ldots, 2^{4}\}$ for the gamma hyperparameter. Consequently, the mentioned search range for the RBF kernel gamma hyperparameter is defined in WEKA by setting the 'YMax' option to '4', the 'YMin' option to '-15', the 'YBase' option to '2', the 'YExpression' option to 'pow(BASE,I)', and the 'YStep' option to '1'.

In RBF kernel SVM, the C hyperparameter controls the maximal distance of a support vector for significant contribution to the decision function for a given width hyperparameter gamma. Lower values for C means a larger portion of instances are considered as support vectors. Considering that the algorithm only uses support vectors for computations, the increasing number of support vectors will correspond to more computations for evaluating the decision function and an RBF structure with more nodes (Belousov et al., 2002a). As explained in Section 5.2.1.9, the search range for penalty parameter C is selected as an exponentially growing sequence of $\{2^{-2}, \ldots, 2^{15}\}$ in this thesis study. However, empirical studies revealed that when the RBF kernel width hyperparameter gamma is adjusted accordingly, lower C values generated good generalization performance and C values between '50' and '100' are recommended for parsimonious solutions (Belousov et al., 2002a). Considering this new recommendation, an additional range between '1' and '100' for C values are also searched. Consequently, the mentioned search range for the C hyperparameter is defined in WEKA by setting the 'XMax' option to '15', the 'XMin' option to '-2', the 'XBase' option to '2', the 'XExpression' option to 'pow(BASE,I)', and the 'XStep' option to '1'. For the additional search range, the 'XMax' option is set to '100', the 'XMin' option is set to '1', the 'XBase' option is set to '10', the 'XExpression' option is set to 'I', and the 'XStep' option is set to '1'.

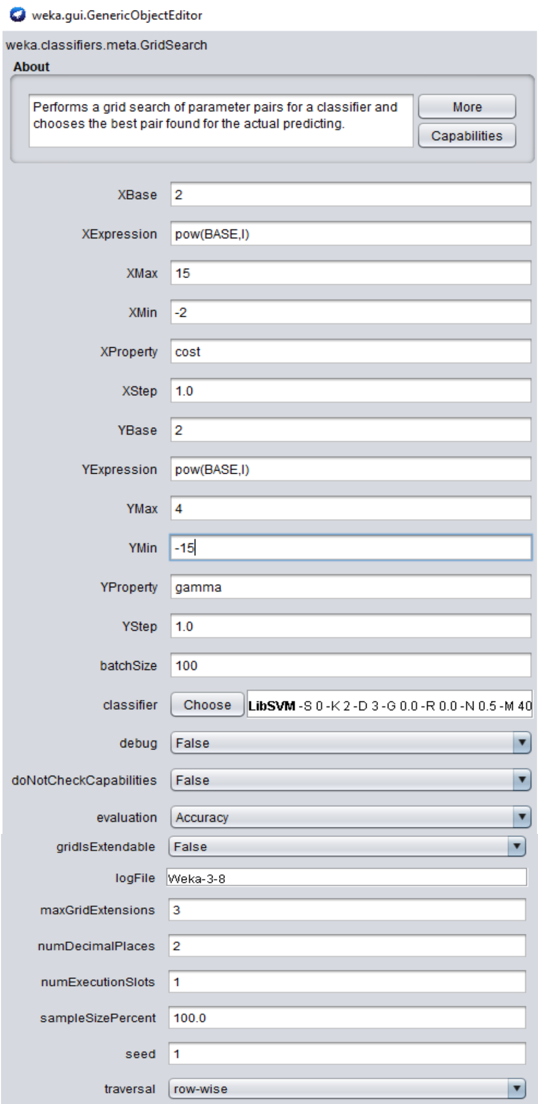


Figure 5.13. The Grid Search Configuration for the Gaussian RBF Kernel SVM Algorithm in WEKA

Finally, the evaluation metric for hyperparameter selection is the classification accuracy as stated in the 'evaluation' option as 'Accuracy'. Default values of WEKA are kept for remaining settings.

5.2.1.12. Results from the Gaussian RBF Kernel SVM

According to results obtained from 10-fold cross-validation with 10 repeats (Table 5.6), Gaussian RBF kernel SVM classifiers have an average classification accuracy of '90.46%' with lower and upper bounds (89.17% - 91.75%) within 95% CI. In other words, the Gaussian RBF kernel SVM algorithm predicts the dispute occurrence in construction projects with an average success rate of '90.46%'.

The average for Kappa statistic value is '0.790' that shows a substantial agreement. The average precision value that indicates the positive predicting power of Gaussian RBF kernel SVM classifiers is '0.925'. The average sensitivity (recall) value is '0.929' that means the success of the Gaussian RBF kernel SVM algorithm in identifying disputed projects is '92.9%'. Similarly, the average specificity value is '0.861' showing the Gaussian RBF kernel SVM achieved '86.1%' success in identifying undisputed projects. The average AUROC value is '0.894'.

Table 5.6. 10-Times 10-Fold Cross-Validation Results for RBF Kernel SVM

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM	Accuracy(%)	90.74	94.44	89.81	91.67	89.81	91.67	89.81	89.81	87.96	88.89	90.46
	Kappa	0.795	0.878	0.773	0.818	0.778	0.818	0.778	0.773	0.738	0.756	0.790
	Precision	0.917	0.957	0.904	0.942	0.928	0.942	0.928	0.904	0.913	0.914	0.925
	Recall	0.943	0.957	0.943	0.929	0.914	0.929	0.914	0.943	0.900	0.914	0.929
	Specificity	0.842	0.921	0.816	0.895	0.868	0.895	0.868	0.816	0.842	0.842	0.861
	AUROC	0.891	0.939	0.879	0.912	0.891	0.912	0.891	0.879	0.871	0.878	0.894

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	650	50
Undisputed	53	327

5.2.2. Comparison of Results from Single Classifiers

Table 5.7 shows the 10-times 10-fold cross-validation results of single classifiers with their best parameter settings. This table is used for comparing performances of single classifiers with each other.

Table 5.7. 10-Times 10-Fold Cross-Validation Performance of All Single Classifiers

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Avg. Precision	Avg. Recall (TPR)	Avg. Specificity	Avg. AUROC	Rank
Naïve Bayes	87.50	[86.60 – 88.40]	0.728	0.912	0.893	0.843	**0.953**	5
KNN	87.69	[86.65 – 88.72]	0.737	**0.931**	0.874	**0.881**	0.928	4
J48	88.98	[87.26 – 90.70]	0.761	0.927	0.901	0.868	0.947	**3**
MLP	83.52	[82.06 – 84.98]	0.641	0.879	0.866	0.779	0.894	6
Poly. Kernel SVM	89.91	[88.85 – 90.96]	0.777	0.917	0.929	0.845	0.887	**2**
RBF Kernel SVM	**90.46**	**[89.17 – 91.75]**	**0.790**	0.925	**0.929**	0.861	0.894	**1**

The best average classification accuracy is obtained from the Gaussian RBF kernel SVM algorithm that achieved '90.46%' average classification accuracy. It is followed by the polynomial kernel SVM algorithm that achieved '89.91%' average classification accuracy. The third place belongs to J48 classifiers with '88.98%' average classification accuracy.

The best average Kappa statistic value is obtained from the Gaussian RBF kernel SVM algorithm as '0.790'.

The best average precision value comes from the kNN algorithm as '0.931'. This is slightly better than the RBF kernel SVM that has the second best average precision value as '0.925'.

Best algorithms in identification of disputed projects are the two versions of the SVM algorithm. They both have an average recall (sensitivity) value of '0.929'. In other words, they achieved '92.9%' average success in identifying positive instances.

The best classifier in identification of undisputed projects is the kNN classifier with an average specificity value of '0.881'. In other words, kNN classifiers achieved '88.1%' average success in identifying negative instances.

The best average AUROC value is obtained from the Naïve Bayes algorithm as '0.953'. This is an almost ideal AUROC value. All algorithms produced impressing AUROC values with the lowest value as '0.887'.

In the light of these comparisons, it is observed that the Gaussian RBF kernel SVM is superior to others in terms of average classification accuracy, average Kappa, and average true positive rate (recall) measures. In addition, it is the second best algorithm with a slightly worse performance behind the kNN algorithm in terms of precision measure.

Since all evaluation metrics have close values compared to each other, it is decided to rank algorithms according to the primary evaluation criterion, which is the average classification accuracy. Thus, the top three algorithms for the dispute occurrence prediction are (1) Gaussian RBF kernel SVM, (2) polynomial kernel SVM, and (3) J48 decision trees. The average classification accuracy of all single classifiers within 95% CI can be seen in Figure 5.14.

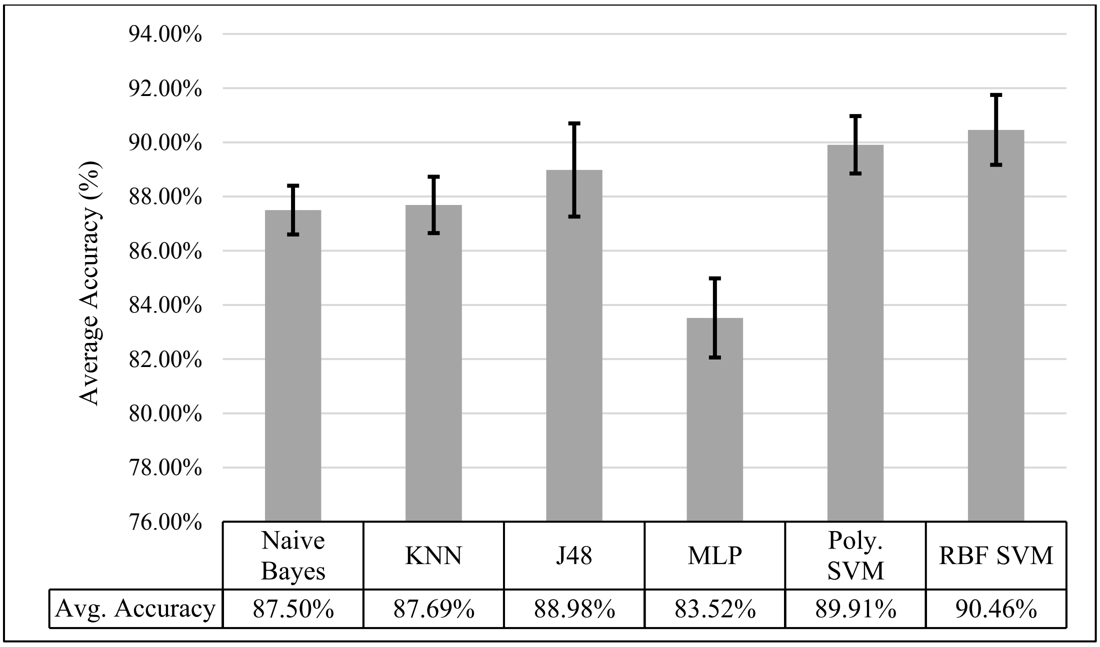


Figure 5.14. Average Classification Accuracies of Single Classifiers within 95% CI

5.2.3. Binary Classification for Dispute Occurrence Prediction Using Ensemble ML Algorithms

Configuration details of each ensemble ML technique and obtained binary classification results are given in this section starting with the voting technique, which will be followed by the stacked generalization technique and the AdaBoost algorithm, in the given order.

5.2.3.1. The Voting Technique and its Configuration in WEKA

In WEKA version 3.8.3, the voting technique is contained in 'weka.classifiers.meta.Vote' class. The WEKA configuration for the voting technique is given in Figure 5.15.

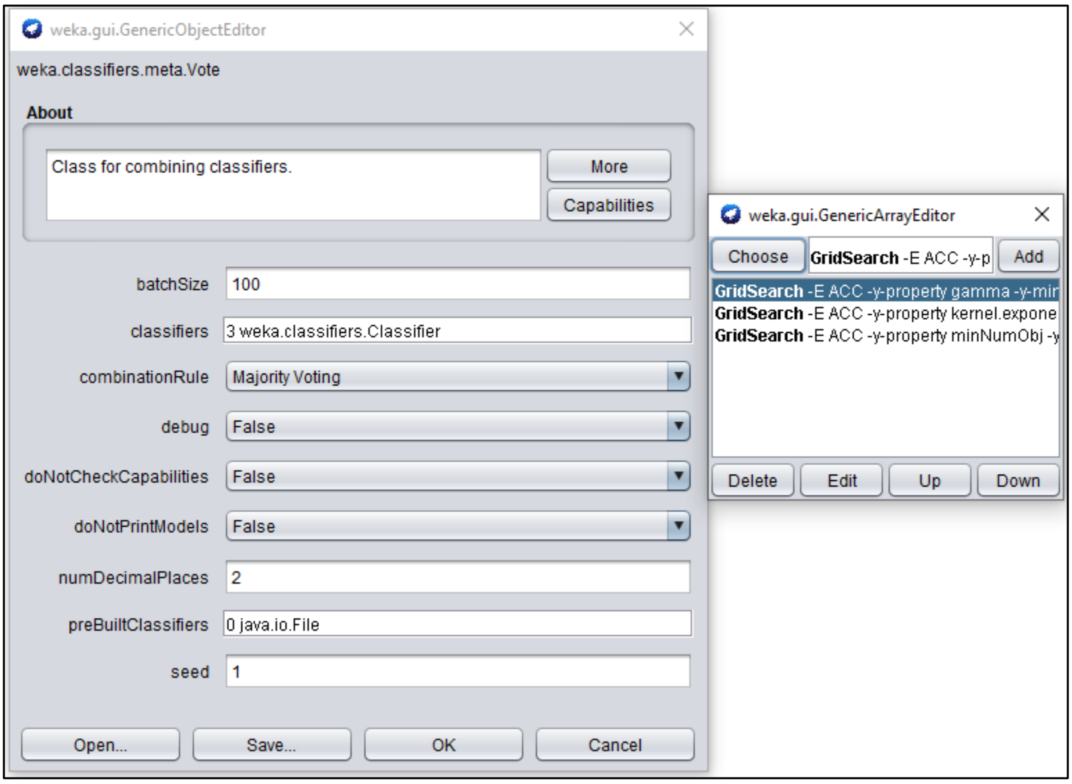


Figure 5.15. The Voting Technique Configuration in WEKA

In this research, results of the top three base (single) classifiers in terms of classification accuracy out of six experimented single algorithms are considered during voting. For dispute occurrence prediction, these algorithms are (1) Gaussian RBF kernel SVM, (2) polynomial kernel SVM, and (3) J48 decision trees. These three algorithms are defined in the voting technique using the 'classifiers' option with their corresponding configurations.

Among various voting strategies in the literature, the majority voting and the average of probabilities techniques are experimented in this research. The voting strategy can be selected by adjusting the 'combinationRule' setting. For this dataset, the majority voting generated better results than the average of probabilities technique. In the majority voting, a project is classified as 'disputed' if two out of three algorithms classify the project as a disputed project and similarly, a project is classified as 'undisputed' if two out of three algorithms classify the project as an undisputed project (Witten et al., 2016). For remaining settings, default values of WEKA are used.

5.2.3.2. Results from the Voting Technique

10-fold cross-validation results with 10 repeats obtained from the majority voting technique are given in Table 5.8. Ensemble classifiers obtained from majority voting have an average classification accuracy of '91.11%' with lower and upper bounds (89.93% - 92.29%) within 95% CI. In other words, ensemble classifiers predict the dispute occurrence in construction projects with an average success rate of '91.11%'.

The average for Kappa statistic value is '0.806' that shows a perfect agreement. The average precision value that indicates the positive predicting power of the voting technique is '0.937'. The average sensitivity (recall) value is '0.926' that means the success of ensemble classifiers in identifying disputed projects is '92.6%'. Similarly, the average specificity value is '0.884' showing ensemble classifiers achieved '88.4%' success in identifying undisputed projects. The average AUROC value is '0.905'.

Table 5.8. 10-Times 10-Fold Cross-Validation Results for the Majority Voting

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
Majority Voting	Accuracy(%)	90.74	94.44	90.74	90.74	91.67	91.67	89.81	91.67	91.67	87.96	91.11
	Kappa	0.795	0.880	0.795	0.802	0.821	0.818	0.778	0.816	0.821	0.735	0.806
	Precision	0.917	0.971	0.917	0.955	0.955	0.942	0.928	0.930	0.955	0.901	0.937
	Recall	0.943	0.943	0.943	0.900	0.914	0.929	0.914	0.943	0.914	0.914	0.926
	Specificity	0.842	0.947	0.842	0.921	0.921	0.895	0.868	0.868	0.921	0.816	0.884
	AUROC	0.892	0.945	0.892	0.911	0.918	0.912	0.891	0.906	0.918	0.865	0.905

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	648	52
Undisputed	44	336

10-fold cross-validation results with 10 repeats obtained from the average of probabilities voting technique are given in Table 5.9. Ensemble classifiers obtained from the average of probabilities voting have an average classification accuracy of ‘90.83%’ with lower and upper bounds (89.98% - 91.69%) within 95% CI. In other words, ensemble classifiers predict the dispute occurrence in construction projects with an average success rate of ‘90.83%’.

Table 5.9. 10-Times 10-Fold Cross-Validation Results for the Average of Probabilities Voting

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
Average of Prob. Voting	Accuracy(%)	88.89	90.74	89.81	90.74	92.59	89.81	90.74	90.74	92.59	91.67	90.83
	Kappa	0.759	0.802	0.783	0.799	0.841	0.778	0.799	0.792	0.840	0.821	0.801
	Precision	0.926	0.955	0.954	0.941	0.970	0.928	0.941	0.905	0.956	0.955	0.943
	Recall	0.900	0.900	0.886	0.914	0.914	0.914	0.914	0.957	0.929	0.914	0.914
	Specificity	0.868	0.921	0.921	0.895	0.947	0.868	0.895	0.816	0.921	0.921	0.897
	AUROC	0.965	0.976	0.967	0.978	0.980	0.973	0.973	0.973	0.974	0.969	0.973

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	640	60
Undisputed	39	341

The average for Kappa statistic value is ‘0.801’ that shows a perfect agreement. The average precision value that indicates the positive predicting power of ensemble classifiers is ‘0.943’. The average sensitivity (recall) value is ‘0.914’ that means the success of ensemble classifiers in identifying disputed projects is ‘91.4%’. Similarly, the average specificity value is ‘0.897’ showing ensemble classifiers achieved ‘89.7%’ success in identifying undisputed projects. The average AUROC value is ‘0.973’.

5.2.3.3. The Stacked Generalization and its Configuration in WEKA

In WEKA version 3.8.3, the Stacked Generalization is contained in ‘weka.classifiers.meta.Stacking’ class. In stacked generalization, the primary algorithm, which is the base-learner, is defined in the ‘classifiers’ setting by selecting the relevant algorithm. The secondary algorithm, which is the meta-learner, is defined in the ‘metaClassifier’ setting by selecting the relevant algorithm. Figure 5.16 shows an example configuration of the ensemble algorithm obtained by combining the kNN algorithm as base-learner and the Naïve Bayes algorithm as meta-learner.

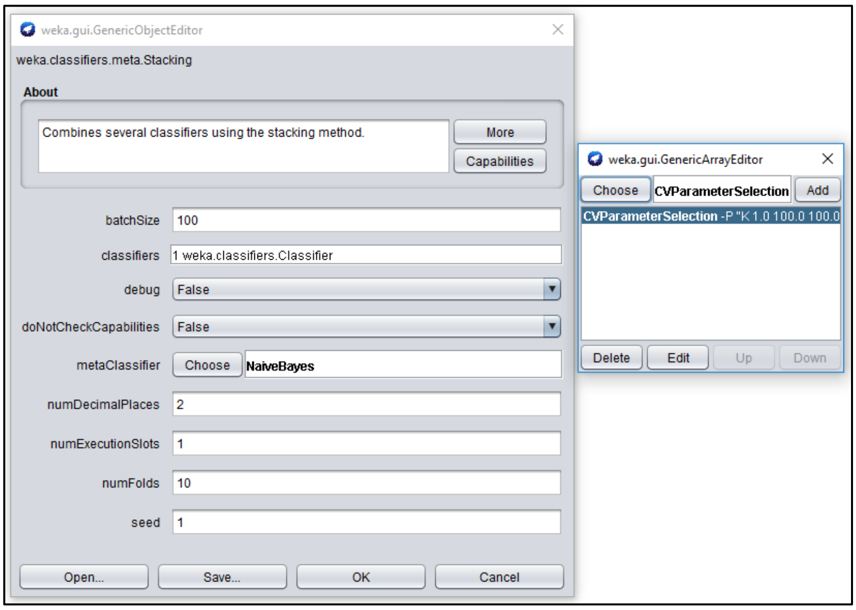


Figure 5.16. The Stacked Generalization Configuration in WEKA

In this thesis study, the top three single classifiers are combined with remaining experimented ML algorithms during stacking. For dispute occurrence prediction, the top three algorithms are (1) Gaussian RBF kernel SVM, (2) polynomial kernel SVM, and (3) J48 decision trees. These top three algorithms will be experimented as base-learners one by one and they will be combined with remaining five ML algorithms as meta-learners in turns. This is because; in stacking, same algorithms should not be stacked together. Therefore, the top three classifiers are combined with the remaining five algorithms so that ‘15’ ensemble classifiers are obtained.

5.2.3.4. Results from the Stacking Technique

When the classification accuracy of single classifiers contained in the ensemble model is as high as possible and classifiers are selected as diverse as possible, the ensemble model can outperform performances of single classifiers it contains (Alpaydın, 2010). However, experiments showed that not all stacked ensemble classifiers achieved better classification accuracies than single ones. Therefore, results of all 15 stacked classifiers are not given. Instead, results of ensemble models that outperformed single algorithms they contain are given. For this purpose, the classification accuracy of the ensemble model is compared with accuracies of both single classifiers they contain.

When the base-learner is the polynomial kernel SVM algorithm, none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models ‘Polynomial kernel SVM + Naïve Bayes’, ‘Polynomial kernel SVM + KNN’, ‘Polynomial kernel SVM + J48’, ‘Polynomial kernel SVM + MLP’, and ‘Polynomial kernel SVM + Gaussian RBF kernel SVM’ are not considered.

Similarly, when the base-learner is the J48 algorithm, none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models containing the J48 algorithm as base-learner are not considered.

When the base-learner is the Gaussian RBF kernel SVM, all ensemble models achieved better classification accuracies than single algorithms they contain. However, all combinations made with the Gaussian RBF kernel SVM algorithm gave the exact same classification results except AUROC values. This is because of the base-learner algorithm. As stated in Section 4.3.2, the base-learner does the most of the work and the meta-learner is like an arbiter (Witten et al., 2016). Here, the performance of the Gaussian RBF kernel SVM as the base-learner dominates the stacked classifier performance. In all five stacked classifiers obtained by combining RBF kernel SVM with remaining algorithms, the average classification accuracy, the average Kappa, the average precision, the average recall, and the average specificity values are the same. Thus, they also generate exact confusion matrices. Consequently, results of these five stacked classifiers will not be given separately. Instead, the stacked classifier that generated the best AUROC value is given. Among the five stacked classifiers, the best AUROC value is obtained from the ensemble classifier that combined 'Gaussian RBF kernel SVM + Polynomial kernel SVM'. Classification results of this stacked ensemble classifier can be seen from Table 5.10.

Table 5.10. 10-Times 10-Fold Cross-Validation Results for the 'RBF Kernel SVM + Poly. Kernel SVM' Stacked Ensemble Classifier

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Stacking RBF SVM + Poly. SVM	Accuracy(%)	91.67	93.52	93.52	91.67	90.74	90.74	89.81	90.74	91.67	87.04	91.11
	Kappa	0.816	0.859	0.857	0.821	0.797	0.799	0.778	0.795	0.818	0.709	0.805
	Precision	0.930	0.957	0.944	0.955	0.929	0.941	0.928	0.917	0.942	0.878	0.932
	Recall	0.943	0.943	0.957	0.914	0.929	0.914	0.914	0.943	0.929	0.929	0.931
	Specificity	0.868	0.921	0.895	0.921	0.868	0.895	0.868	0.842	0.895	0.763	0.874
	AUROC	0.906	0.932	0.926	0.918	0.898	0.905	0.891	0.892	0.912	0.846	0.903

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	652	48
Undisputed	48	332

According to 10-fold cross-validation results with 10 repeats obtained from stacked ensemble classifiers that combined 'RBF kernel SVM + Polynomial kernel SVM', the average classification accuracy is ‘91.11%’ with lower and upper bounds (89.78% - 92.44%) within 95% CI. In other words, ensemble classifiers predict the dispute occurrence in construction projects with an average success rate of ‘91.11%’.

The average for Kappa statistic value is ‘0.805’ that shows a perfect agreement. The average precision value that indicates the positive predicting power of ensemble classifiers is ‘0.932’. The average sensitivity (recall) value is ‘0.931’ that means the success of ensemble classifiers in identifying disputed projects is ‘93.1%’. Similarly, the average specificity value is ‘0.874’ showing ensemble classifiers achieved ‘87.4%’ success in identifying undisputed projects. The average AUROC value is ‘0.903’.

5.2.3.5. The AdaBoost Algorithm and its Configuration in WEKA

In WEKA version 3.8.3, the AdaBoost algorithm is contained in ‘weka.classifiers.meta.AdaBoostM1’ class. The AdaBoost algorithm is used to develop a strong classifier out of several weak classifiers of a specific learning algorithm (Freund and Schapire, 1996). Figure 5.17 shows the configuration for the AdaBoost algorithm.

The weak learner for the AdaBoost algorithm is selected by using the ‘classifiers’ setting. All six single ML algorithms are selected as the weak classifier one by one. Another important setting is the ‘useResampling’ setting that determines whether to use resampling technique instead of the reweighting mechanism in the AdaBoost algorithm (Witten et al., 2016). It is observed that resampling did not improve the classification accuracy in experiments. Default values of WEKA are used for remaining settings.

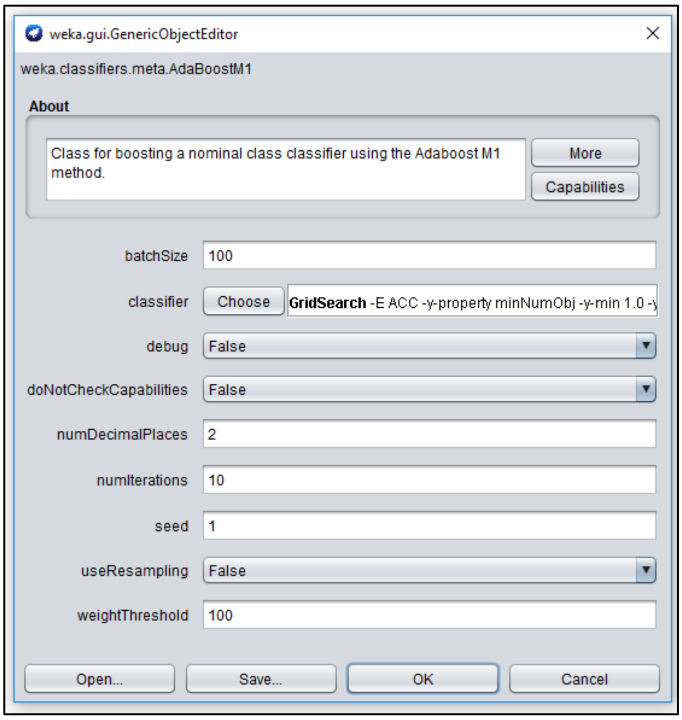


Figure 5.17. The AdaBoost Algorithm Configuration in WEKA

5.2.3.6. Results from the AdaBoost Algorithm

In this thesis study, all six single ML algorithms are boosted by the AdaBoost algorithm. However, as explained in Section 4.3.3, the boosting process might perform poorly if single classifiers are too complex for the amount available training data (Witten et al., 2016). In this thesis study, among all AdaBoost experiments, the performance of the classifier is improved in the Naïve Bayes and the MLP algorithms only. Indeed, these two algorithms were the weakest ones in all six single ML algorithms. The AdaBoost results of remaining algorithms are not taken into account, as they do not enhance the performance.

10-fold cross-validation results with 10 repeats obtained from the AdaBoost algorithm that combined Naïve Bayes classifiers to form an ensemble classifier are given in Table 5.11. Boosted ensemble Naïve Bayes classifiers have an average classification accuracy of '88.06%' with lower and upper bounds (87.20% - 88.91%) within 95%

CI. In other words, boosted ensemble Naïve Bayes classifiers predict the dispute occurrence in construction projects with an average success rate of '88.06%'.

Table 5.11. 10-Times 10-Fold Cross-Validation Results for the AdaBoost Algorithm with Ensemble Naïve Bayes Classifiers

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
AdaBoost Naïve Bayes	Accuracy(%)	87.04	88.89	88.89	87.04	88.89	88.89	86.11	89.81	87.96	87.04	88.06
	Kappa	0.719	0.759	0.759	0.716	0.753	0.753	0.697	0.775	0.735	0.716	0.738
	Precision	0.912	0.926	0.926	0.900	0.903	0.903	0.899	0.915	0.901	0.900	0.908
	Recall	0.886	0.900	0.900	0.900	0.929	0.929	0.886	0.929	0.914	0.900	0.907
	Specificity	0.842	0.868	0.868	0.816	0.816	0.816	0.816	0.842	0.816	0.816	0.832
	AUROC	0.953	0.964	0.959	0.957	0.961	0.964	0.947	0.952	0.933	0.954	0.954

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	635	65
Undisputed	64	316

The average for Kappa statistic value is '0.738' that shows a substantial agreement. The average precision value that indicates the positive predicting power of ensemble classifiers is '0.908'. The average sensitivity (recall) value is '0.907' that means the success of ensemble classifiers in identifying disputed projects is '90.7%'. Similarly, the average specificity value is '0.832' showing ensemble classifiers achieved '83.2%' success in identifying undisputed projects. The average AUROC value is '0.954', which is an almost ideal value.

10-fold cross-validation results with 10 repeats obtained from the AdaBoost algorithm that combined MLP classifiers to form an ensemble classifier are given in Table 5.12. Boosted ensemble MLP classifiers have an average classification accuracy of '83.70%' with lower and upper bounds (81.68% - 85.73%) within 95% CI. In other words, boosted ensemble MLP classifiers predict the dispute occurrence in construction projects with an average success rate of '83.70%'.

The average for Kappa statistic value is '0.651' that shows a substantial agreement. The average precision value that indicates the positive predicting power of ensemble classifiers is ‘0.895’. The average sensitivity (recall) value is ‘0.849’ that means the success of ensemble classifiers in identifying disputed projects is ‘84.9%’. Similarly, the average specificity value is ‘0,816’ showing ensemble classifiers achieved ‘81.6%’ success in identifying undisputed projects. The average AUROC value is ‘0.908’.

Table 5.12. 10-Times 10-Fold Cross-Validation Results for the AdaBoost Algorithm with Ensemble MLP Classifiers

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
AdaBoost MLP	Accuracy(%)	84.26	82.41	77.78	86.11	83.33	82.41	88.89	84.26	84.26	83.33	83.70
	Kappa	0.661	0.630	0.530	0.704	0.643	0.626	0.759	0.649	0.661	0.643	0.651
	Precision	0.896	0.905	0.859	0.923	0.894	0.892	0.926	0.863	0.896	0.894	0.895
	Recall	0.857	0.814	0.786	0.857	0.843	0.829	0.900	0.900	0.857	0.843	0.849
	Specificity	0.816	0.842	0.763	0.868	0.816	0.816	0.868	0.737	0.816	0.816	0.816
	AUROC	0.918	0.871	0.898	0.930	0.895	0.901	0.933	0.905	0.916	0.912	0.908

Confusion Matrix

	Predicted	
Actual	Disputed	Undisputed
Disputed	594	106
Undisputed	70	310

5.2.4. Comparison of Results from Ensemble Classifiers

Table 5.13 shows the 10-times 10-fold cross-validation results of ensemble classifiers that performed better than their single counterparts did. This table is used for comparing performances of ensemble classifiers with each other.

Ensemble classifiers obtained from the majority voting technique and the stacked generalization method (combining the Gaussian RBF kernel SVM as base-learner and the polynomial kernel SVM as meta-learner) generated the best average classification accuracies as ‘91.11%’.

The improvement in average classification accuracy by using the majority voting technique is ‘+0.65%’ to the best single classifier (RBF kernel SVM), ‘+1.20%’ to the second best single classifier (polynomial kernel SVM), and ‘+2.13%’ to the third best single classifier (J48). The improvement in average classification accuracy by using the stacked generalization method is '+0.65%' to base-learner (RBF kernel SVM) and '+1.20%' to meta-learner (polynomial kernel SVM).

Although both ensemble classifiers gave the same average classification accuracy values, ensemble classifiers obtained from the majority voting technique can said to be better than ensemble classifiers obtained from the stacked generalization method considering the remaining evaluation metrics. Indeed, the majority voting technique is superior in terms of average Kappa, average precision, average specificity, and average AUROC values compared to the stacking method. Stacking is stronger than the majority voting in only identification of disputed projects (recall value).

Ensemble classifiers obtained from the AdaBoost algorithm improved their single counterparts. Boosted Naïve Bayes classifiers improved the average classification accuracy of the single Naïve Bayes algorithm by '0.56%'. Similarly, boosted MLP classifiers improved the average accuracy of the single MLP algorithm by '0.18%'. However, both ensemble models obtained from the AdaBoost algorithm performed worse than the majority voting technique and the stacking method.

Table 5.13. 10-Times 10-Fold Cross-Validation Performance of Ensemble Classifiers

Algorithm	**Avg. Accuracy (%)**	**%95 CI Accuracy (%)**	**Avg. Kappa**	**Avg. Prec.**	**Avg. Recall (TPR)**	**Avg. Spec.**	**Avg. AUROC**	**Improve. Base Learner Accuracy**	**Improve. Meta Learner Accuracy**
Majority Voting	91.11	[89.93-92.29]	0.806	0.937	0.926	0.884	0.905	+0.65% to best base learner	NA
Stacking RBF SVM + Poly. SVM	91.11	[89.78-92.44]	0.805	0.932	0.931	0.874	0.903	+0.65%	+1.20%
AdaBoost Naïve Bayes	88.06	[87.20-88.91]	0.738	0.908	0.907	0.832	0.954	+0.56%	NA
AdaBoost MLP	83.70	[81.68-85.73]	0.651	0.895	0.849	0.816	0.908	+0.18%	NA

5.2.5. Comparison of All Classifiers for Dispute Occurrence Prediction

Table 5.14 shows the 10-times 10-fold cross-validation results of all classifiers (single and ensemble) for dispute occurrence prediction together for comparison.

Table 5.14. Comparison of All Dispute Occurrence Prediction Classifiers

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Avg. Precision	Avg. Recall (TPR)	Avg. Specificity	Avg. AUROC	Rank
Majority Voting	**91.11**	**[89.93 – 92.29]**	**0.806**	**0.937**	0.926	**0.884**	0.905	1
Stacking RBF SVM + Poly.SVM	**91.11**	[89.78 – 92.44]	0.805	0.932	**0.931**	0.874	0.903	2
RBF Kernel SVM	90.46	[89.17 – 91.75]	0.790	0.925	0.929	0.861	0.894	3
Poly. Kernel SVM	89.91	[88.85 – 90.96]	0.777	0.917	0.929	0.845	0.887	4
J48	88.98	[87.26 – 90.70]	0.761	0.927	0.901	0.868	0.947	5
KNN	87.69	[86.65 – 88.72]	0.737	0.931	0.874	0.881	0.928	6
AdaBoost Naïve Bayes	88.06	[87.20 – 88.91]	0.738	0.908	0.907	0.832	**0.954**	7
Naïve Bayes	87.50	[86.60 – 88.40]	0.728	0.912	0.893	0.843	0.953	8
AdaBoost MLP	83.70	[81.68 – 85.73]	0.651	0.895	0.849	0.816	0.908	9
MLP	83.52	[82.06 – 84.98]	0.641	0.879	0.866	0.779	0.894	10

As it can be seen from Table 5.14, ensemble classifiers outperformed single classifiers in terms of average prediction accuracy. The first two best performing algorithms are the ensemble ones. The best single algorithm (Gaussian RBF kernel SVM) has the third rank in overall comparison. The average classification accuracy of all classifiers for dispute occurrence prediction within 95% CI can be seen in Figure 5.18.

The ensemble classifier obtained from the majority voting technique gave the best performance in every measure other than the average recall and the average AUROC. The best classifier in terms of recall is another ensemble classifier, which is obtained from the stacking method combining Gaussian RBF kernel SVM as base-learner and polynomial kernel SVM as meta-learner. The AdaBoost algorithm on Naïve Bayes classifiers generates the best AUROC value.

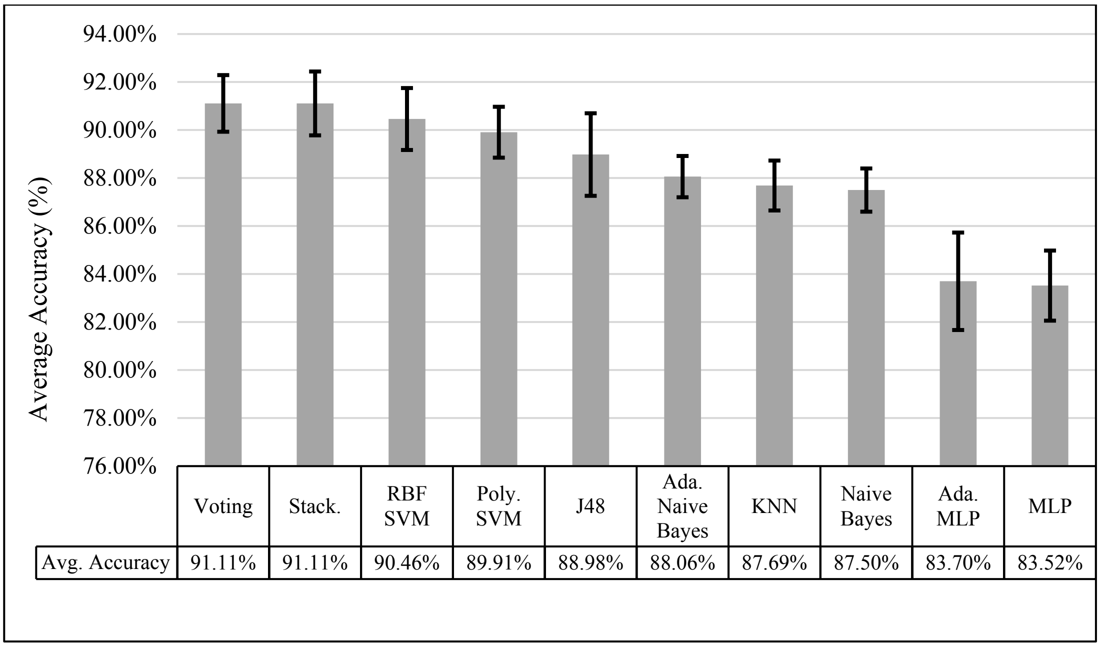


Figure 5.18. Avg. Classification Accuracies of All Classifiers within 95% CI

In the light of foregoing observations, the final model for dispute occurrence prediction is the ensemble classifier obtained from the majority voting technique combining classification decisions of Gaussian RBF kernel SVM, polynomial kernel SVM, and J48 decision trees.

5.3. MULTICLASS CLASSIFICATION PROBLEMS

The binary classification problem of dispute occurrence prediction is handled by six different single ML algorithms so far. These algorithms are well equipped to solve binary data classification problems. However, potential compensation prediction and the problem of resolution method selection should be treated differently than the dispute occurrence prediction problem. This is because potential compensation prediction and resolution method selection problems have more than two classes. As mentioned earlier, when there are more than two class labels that can be assigned to an instance, the problem is a multiclass classification problem. Multiclass classification problems are solved either naturally by extending the binary classification capabilities of algorithms or artificially by decomposing the problem into several binary classification tasks (i.e. one-versus-one (OVO), one-versus-all (OVA), random correction code (RCC), exhaustive correction code(ECC)) (Li et al., 2004; Aly, 2005). Experiments by using both approaches are performed in this thesis study and results are compared.

Among experimented single ML algorithms, the Naïve Bayes algorithm is naturally extensible for multiclass problems and literature has shown that it can achieve significantly good classification performance (Thakkar et al., 2011).

In the kNN algorithm, the distance between the new instance and every other training instances are calculated using a distance measure (i.e. Manhattan distance), the k number of smallest distances are identified, and the most represented class in these k instances are assigned to the new instance as the class label (Aly, 2005). Therefore, the kNN algorithm can naturally handle multiclass classification problems.

Similarly, in the C4.5 decision tress (the J48 algorithm), there are leaf nodes that represent class labels. If there are K classes, there will be K leaf nodes (Aly, 2005). When a new instance should be classified, the decision tree structure is followed from the root node to the leaf node according to tests on attribute values of the new instance. The final leaf node that is reached is the class label of the new instance. Thus, the J48 algorithm is also capable of naturally handling multiclass classification problems (Thakkar et al., 2011).

MLP networks can also handle multiclass classification problems naturally by increasing the number of output neurons to K in case of K class problems instead of using '1' output neuron like the case in binary classification (Thakkar et al., 2011).

Unlike other single ML algorithms mentioned so far, the SVM algorithm is designed for binary classification problems only (Cortes and Vapnik, 1995). Therefore, the SVM algorithm or the multiclass classification problem should be adjusted to reach a solution. However, the problem in the literature is how to effectively extend the SVM algorithm for multiclass classification problems and it should be determined experimentally (Hsu and Lin, 2002). One method is to use all classes in a single optimization formulation (Mayoraz and Alpaydin, 1999). However, it is computationally more expensive to solve a multiclass problem than to solve a binary problem of the same size (Hsu and Lin, 2002). Another method is based on using external methods such as OVO, OVA, etc. and decomposing the multiclass problem into several binary problems (An et al., 2007). In this thesis study, it is preferred to convert the multiclass problem into several binary classification problems for the SVM algorithm.

The WEKA workbench supports four multiclass problem decomposition techniques. These are (1) one-vs-one (OVO), (2) one-vs-all (OVA), (3) random correction code (RCC), and (4) exhaustive correction code (ECC). All these techniques will be explained briefly.

The 'one-vs-one' (OVO) method, which is also called the pairwise classification, is a multiclass solution technique that is based on training classifiers for each pair of classes (Li et al., 2004). Thus, in case of a dataset composed of k classes, this method will generate a total of [k (k - 1) / 2] independent classifiers with each classifier trained on data from two classes (Hsu and Lin, 2002). For potential compensation prediction, there are '4' class labels representing compensation types and consequently, there will be '6' binary classifiers. Similarly, for resolution method selection problem, there are '6' class labels and consequently, there will be '15' binary classifiers. Then, the prediction of each classifier for the class label of a new instance will have a vote and the class that gets the majority of votes is assigned to the new instance (Raziff et al., 2017). Although this technique requires constructing several binary classifiers, it still has low computational cost. This is because each pairwise learning problem only involves instances of the two classes under consideration (Witten et al., 2016).

For a problem with K classes (K > 2), the 'one-vs-all' (OVA) method defines K binary problems with each problem separating one class from all other classes combined (Alpaydın, 2010). This method is also called 'one-vs-rest' and it produces several binary datasets by discriminating each class against the union of remaining classes

(Witten et al., 2016). Thus, if there are K classes, this method constructs K classifiers and unlike the OVO method that pairs each class with another class one by one, every class is paired with remaining classes at once in the OVA technique (Raziff et al., 2017). In other words, instances belonging to class i are considered as positive instances and remaining instances that do not belong to class i are considered as negative instances (Hsu and Lin, 2002). During classification with the OVA technique, each classifier will produce a confidence figure of their predictions and the prediction of the classifier that has the highest confidence for a new instance is assigned as the final class label (Witten et al., 2016). This method may perform poorly if the number of classes is excessively large (Raziff et al., 2017). On the other hand, it has been proven empirically that the OVA technique is a competitive method when parameters of the base classifier are tuned appropriately (Witten et al., 2016).

Practically, both OVO and OVA techniques are special cases of a method called the error-correcting output codes (Alpaydın, 2010). The method is proposed by Dietterich and Bakiri (1994) where the multiclass problem is decomposed into a set of binary problems and classes are represented by bits of code words in pursue of enhancing the classification performance. In more simple terms, the error-correcting output codes method changes the representation of classes by using an ensemble of binary classifiers to decide individual bits of the code word for the output class where each bit can take a value of '0' or '1'. In error-correcting output codes method, an initial matrix of code words are constructed depending on the number of classes and then, ensemble classifiers are trained depending on the code word so that it follows binary classification rules. For a test dataset, each code word classifier is evaluated and the classifier that is closest to the test dataset is presented as the final classifier (Raziff et al., 2017).

The WEKA workbench has two extensions of the original error-correcting output codes. These are the random correction code (RCC) and the exhaustive correction code (ECC) techniques. The ECC technique often generates accurate classifiers for multiclass problems. However, as the number of classes increases, the number of classifiers that should be constructed increases exponentially (Li et al., 2004). In other words, when the number of classes is large, this technique becomes infeasible as too many classifiers have to be generated and in such cases, RCC technique can be preferred (Witten et al., 2016). The difference between ECC and RCC techniques is that in RCC, the code word matrix can be randomized at the initial construction (Raziff et al., 2017).

Multiclass classification problems of potential compensation prediction and resolution method selection in this thesis study are solved without converting the problem into binary tasks whenever it is possible. Meanwhile, all techniques that can be used to decompose the problem into binary problems are also experimented whenever it is possible. Thus, both the natural approach and approaches using OVO, OVA, RCC, and ECC techniques are experimented and results are given in following sections for each of the six single ML algorithms.

5.3.1. Multiclass Classification Using Single ML Algorithms

WEKA configuration details of each single ML algorithm and obtained multiclass classification results are given in this section starting with the Naïve Bayes algorithm, which will be followed by the kNN, J48, MLP, polynomial kernel SVM, and Gaussian RBF kernel SVM, in the given order.

It should be noted that unlike the case in binary classification where average values for evaluation metrics are given, in multiclass classification average values are used only for classification accuracy and Kappa statistic values. Other performance metrics are given in weighted average values so that class populations are reflected to the performance of classifiers.

5.3.1.1. The Naïve Bayes Algorithm and its Configuration in WEKA

The Naïve Bayes algorithm can naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. The configuration of the Naïve Bayes algorithm in WEKA for binary classification can be used exactly the same way to obtain multiclass solutions. In addition to this, both multiclass classification problems can be solved by decomposing them into several binary problems. In order to do this, the class 'weka.classifiers.meta.MultiClassClassifier' should be selected. Configuration details for the Naïve Bayes algorithm using decomposition techniques in WEKA can be seen in Figure 5.19.

To select the Naïve Bayes algorithm, the 'classifier' setting should be set to Naïve Bayes with the configuration used in binary classification task (Figure 5.3). The 'method' setting allows the user to select the decomposition technique among OVO, OVA, RCC, and ECC. Upon selecting the OVO technique, the 'use PairwiseCoupling' setting can set to 'True' for better performance. This setting is only applicable for the OVO technique. Default values in WEKA are used for remaining settings.

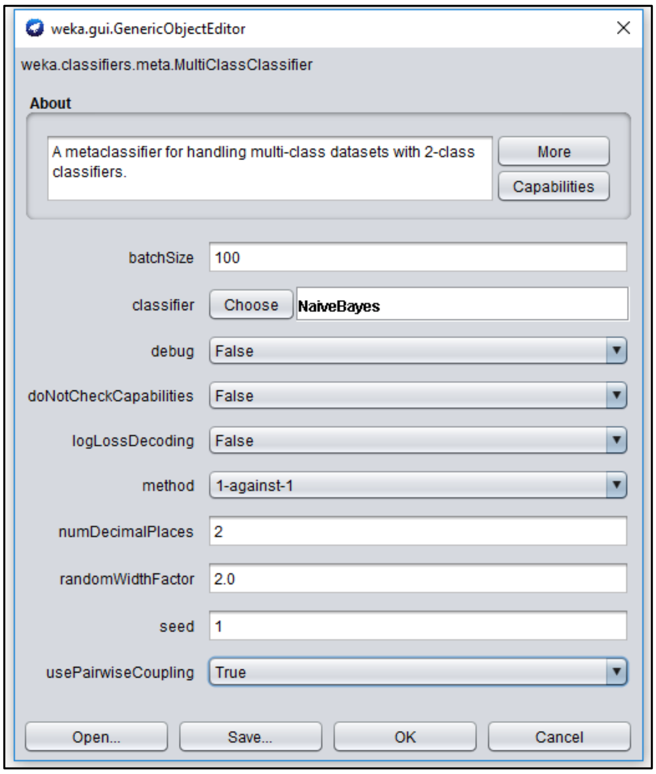


Figure 5.19. The Multiclass Naïve Bayes Classifier Configuration in WEKA

5.3.1.2. Results from the Naïve Bayes Algorithm for Potential Compensation Prediction

The solution without using decomposition techniques for Naïve Bayes is exactly equal to the solution obtained from using OVO technique with pairwise coupling. Thus, the classification obtained by 10-fold cross-validation with 10 repeats is given as results from the Naïve Bayes algorithm using OVO technique in Table 5.15.

According to these results, Naïve Bayes classifiers using OVO technique with pairwise coupling have an average classification accuracy of '79.27%' with lower and upper bounds (78.26% - 80.28%) within 95% CI. In other words, the Naïve Bayes OVO algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '79.27%'.

Table 5.15. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes OVO Pairwise Coupling	Accuracy(%)	80.49	80.49	76.83	79.27	79.27	79.27	80.49	76.83	80.49	79.27	79.27
	Kappa	0.692	0.692	0.640	0.674	0.674	0.674	0.695	0.640	0.692	0.675	0.675
	Precision	0.788	0.788	0.757	0.776	0.776	0.776	0.794	0.757	0.788	0.777	0.778
	Recall	0.805	0.805	0.768	0.793	0.793	0.793	0.805	0.768	0.805	0.793	0.793
	Specificity	0.903	0.903	0.899	0.902	0.902	0.902	0.908	0.899	0.903	0.901	0.902
	AUROC	0.924	0.923	0.915	0.921	0.921	0.916	0.920	0.922	0.918	0.920	0.920

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	50	40	0	30
Cost Comp. Only	33	337	0	10
Time Comp. Only	0	0	11	39
Cost & Time Comp.	0	0	18	252

The average for Kappa statistic value is '0.675' that shows a substantial agreement. The weighted average precision value of Naïve Bayes OVO classifiers is '0.778'. The weighted average sensitivity (recall) value is '0.793' that means the success of Naïve Bayes OVO classifiers in identifying true positive instances is '79.3%'. Similarly, the weighted average specificity value is '0.902' showing the algorithm achieved '90.2%'

success in identifying true negative instances. The weighted average AUROC value is ‘0.920’, which is the highest AUROC value among all experimented Naïve Bayes classifiers for potential compensation prediction.

In the second experiment with the Naïve Bayes algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.16), Naïve Bayes classifiers using OVA technique have an average classification accuracy of ‘80.61%’ with lower and upper bounds (80.11% - 81.10%) within 95% CI. In other words, the Naïve Bayes OVA algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘80.61%’.

Table 5.16. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes OVA	Accuracy(%)	80.49	81.71	80.49	80.49	79.27	80.49	81.71	80.49	80.49	80.49	80.61
	Kappa	0.688	0.709	0.688	0.692	0.670	0.688	0.709	0.688	0.685	0.688	0.691
	Precision	0.765	0.803	0.765	0.788	0.770	0.765	0.803	0.765	0.752	0.765	0.774
	Recall	0.805	0.817	0.805	0.805	0.793	0.805	0.817	0.805	0.805	0.805	0.806
	Specificity	0.899	0.905	0.899	0.903	0.892	0.899	0.905	0.899	0.885	0.899	0.899
	AUROC	0.919	0.918	0.911	0.915	0.918	0.911	0.920	0.916	0.917	0.916	0.916

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	49	41	0	30
Cost Comp. Only	22	348	0	10
Time Comp. Only	0	1	4	45
Cost & Time Comp.	0	1	9	260

The average for Kappa statistic value is ‘0.691’ that shows a substantial agreement. The weighted average precision value of Naïve Bayes OVA classifiers is ‘0.774’. The weighted average sensitivity (recall) value is ‘0.806’ that means the success of Naïve Bayes OVA classifiers in identifying true positive instances is ‘80.6%’. Similarly, the weighted average specificity value is ‘0.899’ showing the algorithm achieved ‘89.9%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.916’, which is a remarkably high AUROC value showing the success of the algorithm.

In the third experiment with the Naïve Bayes algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.17), Naïve Bayes classifiers using RCC technique have an average classification accuracy of ‘78.54%’ with lower and upper bounds (77.16% - 79.91%) within 95% CI. In other words, the Naïve Bayes RCC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘78.54%’.

Table 5.17. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes RCC	Accuracy(%)	79.27	78.05	80.49	81.71	79.27	74.39	78.05	78.05	78.05	78.05	78.54
	Kappa	0.667	0.654	0.689	0.705	0.663	0.592	0.641	0.653	0.649	0.654	0.657
	Precision	0.748	0.753	0.771	0.781	0.736	0.704	0.717	0.757	0.763	0.770	0.750
	Recall	0.793	0.780	0.805	0.817	0.793	0.744	0.780	0.780	0.780	0.780	0.785
	Specificity	0.888	0.891	0.904	0.897	0.874	0.870	0.864	0.892	0.884	0.893	0.886
	AUROC	0.869	0.857	0.856	0.905	0.896	0.879	0.877	0.891	0.888	0.883	0.880

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	41	48	1	30
Cost Comp. Only	20	350	0	10
Time Comp. Only	1	6	4	39
Cost & Time Comp.	1	4	16	249

The average for Kappa statistic value is ‘0.657’ that shows a substantial agreement. The weighted average precision value of Naïve Bayes RCC classifiers is ‘0.750’. The weighted average sensitivity (recall) value is ‘0.785’ that means the success of classifiers in identifying true positive instances is ‘78.5%’. Similarly, the weighted average specificity value is ‘0.886’ showing the algorithm achieved ‘88.6%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.880’. In the light of these, it can be said that the RCC technique performed slightly worse than other decomposition techniques in all performance measures.

In the final experiment with the Naïve Bayes algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.18), Naïve Bayes classifiers using ECC technique have an average classification accuracy of '80.00%' with lower and upper bounds (79.26% - 80.74%) within 95% CI. In other words, the Naïve Bayes ECC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '80.00%'.

Table 5.18. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes ECC	Accuracy(%)	80.49	80.49	79.27	79.27	78.05	80.49	79.27	80.49	81.71	80.49	80.00
	Kappa	0.688	0.687	0.667	0.667	0.646	0.687	0.663	0.684	0.703	0.688	0.678
	Precision	0.765	0.762	0.748	0.747	0.720	0.762	0.736	0.769	0.771	0.765	0.755
	Recall	0.805	0.805	0.793	0.793	0.780	0.805	0.793	0.805	0.817	0.805	0.800
	Specificity	0.899	0.894	0.888	0.892	0.872	0.894	0.874	0.891	0.887	0.899	0.889
	AUROC	0.920	0.918	0.914	0.920	0.920	0.912	0.920	0.918	0.918	0.916	0.918

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	46	44	0	30
Cost Comp. Only	20	350	0	10
Time Comp. Only	0	6	0	44
Cost & Time Comp.	0	3	7	260

The average for Kappa statistic value is '0.678' that shows a substantial agreement. The weighted average precision value of Naïve Bayes ECC classifiers is '0.755'. The weighted average sensitivity (recall) value is '0.800' that means the success of classifiers in identifying true positive instances is '80.0%'. Similarly, the weighted average specificity value is '0.889' showing the algorithm achieved '88.9%' success in identifying true negative instances. The weighted average AUROC value is '0.918', which is a remarkably high AUROC value showing the success of the algorithm.

Considering performance of experiments using the Naïve Bayes algorithm are very close to each other, the primary evaluation criterion (average prediction accuracy) can be used for determining the best Naïve Bayes classifier for potential compensation prediction. Thus, it can be said that the best performing Naïve Bayes classifier is obtained from the OVA technique that achieved '80.61%' average prediction accuracy.

5.3.1.3. Results from the Naïve Bayes Algorithm for Resolution Method Selection

In the first experiment with the Naïve Bayes algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.19), Naïve Bayes classifiers using OVO technique with pairwise coupling have an average classification accuracy of '80.37%' with lower and upper bounds (78.95% - 81.79%) within 95% CI. In other words, the Naïve Bayes OVO algorithm predicts the resolution method to be used in construction disputes with an average success rate of '80.37%'.

Table 5.19. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using OVO Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes OVO Pairwise Coupling	Accuracy(%)	83.33	79.63	77.78	79.63	81.48	79.63	81.48	77.78	83.33	79.63	80.37
	Kappa	0.785	0.738	0.710	0.735	0.760	0.735	0.760	0.710	0.782	0.739	0.745
	Precision	0.848	0.808	0.800	0.812	0.837	0.813	0.837	0.794	0.855	0.807	0.821
	Recall	0.833	0.796	0.778	0.796	0.815	0.796	0.815	0.778	0.833	0.796	0.804
	Specificity	0.937	0.929	0.912	0.923	0.932	0.917	0.932	0.913	0.931	0.935	0.926
	AUROC	0.959	0.949	0.949	0.953	0.955	0.953	0.954	0.952	0.953	0.956	0.953

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	70	13	7	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	42	0	8	0
Mediation	0	0	0	31	0	19
SEA	1	0	0	0	61	38
Negotiation	0	0	0	0	20	170

The average for Kappa statistic value is '0.745' that shows a substantial agreement. The weighted average precision value of Naïve Bayes OVO classifiers is '0.821'. The weighted average sensitivity (recall) value is '0.804' that means the success of Naïve Bayes OVO classifiers in identifying true positive instances is '80.4%'. Similarly, the weighted average specificity value is ‘0.926’ showing the algorithm achieved ‘92.6%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.953’, which is an almost perfect value.

In the second experiment with the Naïve Bayes algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.20), Naïve Bayes classifiers using OVA technique have an average classification accuracy of ‘83.15%’ with lower and upper bounds (81.44% - 84.85%) within 95% CI. In other words, the Naïve Bayes OVA algorithm predicts the resolution method to be used in construction disputes with an average success rate of ‘83.15%’.

Table 5.20. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using OVA Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes OVA	Accuracy(%)	83.33	79.63	81.48	85.19	81.48	85.19	87.04	81.48	81.48	85.19	83.15
	Kappa	0.785	0.738	0.759	0.807	0.760	0.807	0.830	0.758	0.758	0.809	0.781
	Precision	0.842	0.803	0.829	0.870	0.823	0.884	0.900	0.839	0.839	0.858	0.849
	Recall	0.833	0.796	0.815	0.852	0.815	0.852	0.870	0.815	0.815	0.852	0.832
	Specificity	0.937	0.923	0.919	0.933	0.925	0.935	0.937	0.921	0.921	0.946	0.930
	AUROC	0.959	0.953	0.952	0.960	0.958	0.958	0.959	0.956	0.959	0.958	0.957

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	75	9	6	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	49	0	1	0
Mediation	0	0	0	34	0	16
SEA	1	0	0	0	55	44
Negotiation	0	0	0	0	14	176

The average for Kappa statistic value is ‘0.781’ that shows a substantial agreement. The weighted average precision value of Naïve Bayes OVA classifiers is ‘0.849’. The weighted average sensitivity (recall) value is ‘0.832’ that means the success of Naïve Bayes OVA classifiers in identifying true positive instances is ‘83.2%’. Similarly, the weighted average specificity value is ‘0.930’ showing the algorithm achieved ‘93.0%’

success in identifying true negative instances. The weighted average AUROC value is ‘0.957’, which is an almost perfect value.

In the third experiment with the Naïve Bayes algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.21), Naïve Bayes classifiers using RCC technique have an average classification accuracy of ‘78.33%’ with lower and upper bounds (75.75% - 80.91%) within 95% CI. In other words, the Naïve Bayes RCC algorithm predicts the resolution method to be used in construction disputes with an average success rate of ‘78.33%’.

Table 5.21. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using RCC Technique for Resolution Method Selection

Classifier	**Performance Measure**	**Run Number**										**Avg.**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Naïve Bayes RCC	Accuracy(%)	79.63	83.33	75.93	77.78	75.93	77.78	72.22	75.93	83.33	81.48	78.33
	Kappa	0.739	0.785	0.692	0.712	0.690	0.712	0.645	0.692	0.784	0.760	0.721
	Precision	0.802	0.846	0.760	0.797	0.765	0.784	0.737	0.755	0.843	0.845	0.793
	Recall	0.796	0.833	0.759	0.778	0.759	0.778	0.722	0.759	0.833	0.815	0.783
	Specificity	0.932	0.945	0.924	0.922	0.919	0.919	0.917	0.927	0.934	0.929	0.927
	AUROC	0.941	0.926	0.926	0.931	0.932	0.924	0.910	0.926	0.952	0.931	0.930

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	76	10	3	0	0	1
Arbitration	4	52	0	0	4	0
DRB	1	0	49	0	0	0
Mediation	0	1	0	31	0	18
SEA	4	6	3	1	51	35
Negotiation	1	3	2	4	16	164

The average for Kappa statistic value is ‘0.721’ that shows a substantial agreement. The weighted average precision value of Naïve Bayes RCC classifiers is ‘0.793’. The weighted average sensitivity (recall) value is ‘0.783’ that means the success of classifiers in identifying true positive instances is ‘78.3%’. Similarly, the weighted average specificity value is ‘0.927’ showing the algorithm achieved ‘92.7%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.930’.

In the light of these, it can be said that the RCC technique performed slightly worse than other decomposition techniques in all performance measures.

In the final experiment with the Naïve Bayes algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.22), Naïve Bayes classifiers using ECC technique have an average classification accuracy of '85.93%' with lower and upper bounds (84.50% - 87.35%) within 95% CI. In other words, the Naïve Bayes ECC algorithm predicts the resolution method to be used in construction disputes with an average success rate of '85.93%'.

Table 5.22. 10-Times 10-Fold Cross-Validation Results of the Naïve Bayes Algorithm Using ECC Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Naïve Bayes ECC	Accuracy(%)	87.04	83.33	83.33	85.19	87.04	87.04	87.04	83.33	87.04	88.89	85.93
	Kappa	0.832	0.784	0.784	0.807	0.831	0.831	0.831	0.782	0.831	0.856	0.817
	Precision	0.874	0.839	0.839	0.870	0.881	0.900	0.900	0.855	0.900	0.896	0.875
	Recall	0.870	0.833	0.833	0.852	0.870	0.870	0.870	0.833	0.870	0.889	0.859
	Specificity	0.941	0.929	0.929	0.933	0.935	0.937	0.937	0.923	0.937	0.952	0.935
	AUROC	0.962	0.959	0.955	0.961	0.963	0.960	0.963	0.957	0.963	0.963	0.961

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	83	7	0	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	50	0	0	0
Mediation	0	0	0	37	0	13
SEA	1	0	0	0	54	45
Negotiation	0	0	0	0	10	180

The average for Kappa statistic value is '0.817' that shows a perfect agreement. The weighted average precision value of Naïve Bayes ECC classifiers is '0.875'. The weighted average sensitivity (recall) value is '0.859' that means the success of classifiers in identifying true positive instances is '85.9%'. Similarly, the weighted average specificity value is '0.935' showing the algorithm achieved '93.5%' success in identifying true negative instances. The weighted average AUROC value is '0.961', which is the highest AUROC value among all experimented Naïve Bayes classifiers.

Considering results of experiments using the Naïve Bayes algorithm, it can be said that the best performing Naïve Bayes classifier is obtained from the ECC technique that achieved ‘85.93%’ average prediction accuracy for resolution method selection. In addition, this classifier is superior to other experimented Naïve Bayes classifiers in all remaining performance measures.

5.3.1.4. The KNN Algorithm and its Configuration in WEKA

The kNN algorithm can naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. The configuration of the kNN algorithm in WEKA for binary classification can be used exactly the same way to obtain multiclass solutions. In addition to this, both multiclass classification problems can be solved by decomposing them into several binary problems using the class ‘weka.classifiers.meta.MultiClassClassifier’. Configuration details for the kNN algorithm using decomposition techniques in WEKA can be seen in Figure 5.20.

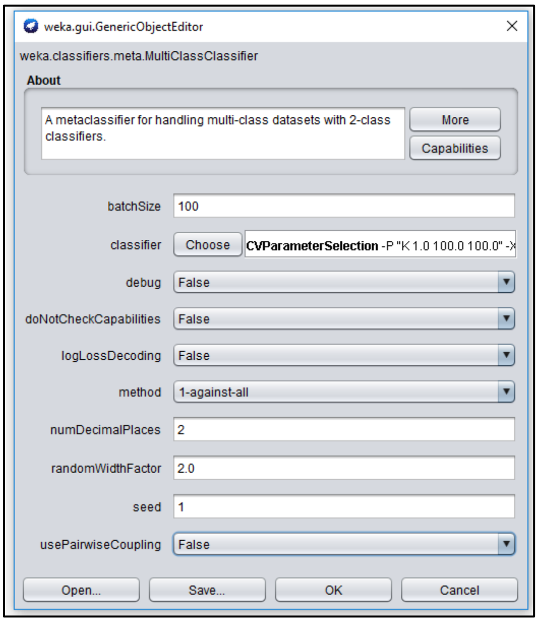


Figure 5.20. The Multiclass KNN Classifier Configuration in WEKA

To select the kNN algorithm, the ‘classifier’ setting should be set to kNN with the configuration used in binary classification task (‘CVParameterSelection’) (Figure 5.4

and Figure 5.5). The only difference from binary case is that the k parameter search range is between '1' and '82' for potential compensation prediction and '1' and '54' for resolution method selection. The 'method' setting allows the user to select the decomposition technique. Other than the natural solution obtained from the kNN algorithm, results from using decomposition techniques of OVO, OVA, RCC, and ECC are compared to each other. Default values in WEKA are used for remaining settings.

5.3.1.5. Results from the KNN Algorithm for Potential Compensation Prediction

In the first experiment, 10-fold cross-validation results with 10 repeats obtained from using the kNN algorithm without any decomposition technique (natural solution) are given in Table 5.23.

Table 5.23. 10-Times 10-Fold Cross-Validation Results of KNN Algorithm without Using Decomposition Techniques for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN No Decomp.	Accuracy(%)	80.49	76.83	76.83	79.27	76.83	80.49	80.49	74.39	80.49	80.49	78.66
	Kappa	0.689	0.636	0.635	0.671	0.636	0.688	0.688	0.597	0.688	0.689	0.661
	Precision	0.753	0.723	0.725	0.742	0.723	0.752	0.750	0.701	0.750	0.753	0.737
	Recall	0.805	0.768	0.768	0.793	0.768	0.805	0.805	0.744	0.805	0.805	0.787
	Specificity	0.898	0.891	0.888	0.896	0.891	0.898	0.893	0.880	0.893	0.898	0.893
	AUROC	0.911	0.915	0.908	0.913	0.915	0.903	0.916	0.905	0.921	0.916	0.912

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	49	42	0	29
Cost Comp. Only	37	333	0	10
Time Comp. Only	0	3	0	47
Cost & Time Comp.	5	0	2	263

The optimum k parameter is determined as k=3 by the cross-validation parameter selection algorithm. Similar to previous experiments with the kNN, the best results are obtained from using (1 / weight) distance weighting method with the Manhattan distance measure. According to these results, kNN classifiers have an average multiclass classification accuracy of '78.66%' with lower and upper bounds (77.05%

- 80.26%) within 95% CI. In other words, the kNN algorithm without using decomposition techniques predicts the potential compensation type that can be acquired in a dispute with an average success rate of '78.66%'.

The average for Kappa statistic value is '0.661' that shows a substantial agreement. The weighted average precision value is '0.737'. The weighted average sensitivity (recall) value is '0.787' that means the success of classifiers in identifying true positive instances is '78.7%'. Similarly, the weighted average specificity value is '0.893' showing the algorithm achieved '89.3%' success in identifying true negative instances. The weighted average AUROC value is '0.912', which is the highest AUROC value among all experimented kNN classifiers for potential compensation prediction.

In the second experiment with the kNN algorithm, the OVO technique is utilized. This time, using pairwise coupling generated worse results than using no pairwise coupling. Thus, the OVO technique is used with no pairwise coupling. The kNN OVO algorithm generates six distinct classifiers and according to 10-fold cross-validation results with 10 repeats (Table 5.24), kNN classifiers using OVO technique with no pairwise coupling have an average classification accuracy of '77.68%' with lower and upper bounds (76.67% - 78.69%) within 95% CI. In other words, the kNN OVO algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '77.68%'.

The average for Kappa statistic value is '0.644' that shows a substantial agreement. The weighted average precision value is '0.728'. The weighted average sensitivity (recall) value is '0.777' that means the success of classifiers in identifying true positive instances is '77.7%'. Similarly, the weighted average specificity value is '0.886' showing the algorithm achieved '88.6%' success in identifying true negative instances. The weighted average AUROC value is '0.888'.

Table 5.24. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
KNN	Accuracy(%)	79.27	78.05	76.83	78.05	78.05	78.05	78.05	74.39	76.83	79.27	77.68
OVO	Kappa	0.671	0.648	0.638	0.650	0.645	0.644	0.649	0.597	0.631	0.671	0.644
	Precision	0.751	0.729	0.737	0.722	0.714	0.714	0.732	0.707	0.719	0.751	0.728
No	Recall	0.793	0.780	0.768	0.780	0.780	0.780	0.780	0.744	0.768	0.793	0.777
Pairwise	Specificity	0.897	0.882	0.898	0.885	0.877	0.877	0.886	0.881	0.884	0.897	0.886
Coupling	AUROC	0.899	0.888	0.891	0.893	0.883	0.874	0.887	0.878	0.883	0.900	0.888

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	41	49	0	30
Cost Comp. Only	33	337	1	9
Time Comp. Only	0	1	0	49
Cost & Time Comp.	3	0	8	259

In the third experiment with the kNN algorithm, the OVA technique is utilized. The kNN OVA algorithm generates four distinct classifiers (since there are four output classes) and according to 10-fold cross-validation results with 10 repeats (Table 5.25), kNN classifiers using OVA technique have an average classification accuracy of '76.34%' with lower and upper bounds (74.59% - 78.10%) within 95% CI. In other words, the kNN OVA algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '76.34%'.

The average for Kappa statistic value is '0.629' that shows a substantial agreement. The weighted average precision value is '0.732'. The weighted average sensitivity (recall) value is '0.763' that means the success of classifiers in identifying true positive instances is '76.3%'. Similarly, the weighted average specificity value is '0,893' showing the algorithm achieved '89.3%' success in identifying true negative instances. The weighted average AUROC value is '0.904'.

Table 5.25. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
KNN OVA	Accuracy(%)	79.27	78.05	73.17	78.05	73.17	78.05	75.61	74.39	74.39	79.27	76.34
	Kappa	0.671	0.654	0.584	0.653	0.584	0.655	0.619	0.599	0.598	0.672	0.629
	Precision	0.751	0.739	0.714	0.748	0.712	0.746	0.732	0.715	0.717	0.748	0.732
	Recall	0.793	0.780	0.732	0.780	0.732	0.780	0.756	0.744	0.744	0.793	0.763
	Specificity	0.897	0.895	0.889	0.896	0.885	0.900	0.897	0.887	0.878	0.901	0.893
	AUROC	0.907	0.915	0.911	0.900	0.902	0.902	0.905	0.908	0.904	0.890	0.904

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	49	41	0	30
Cost Comp. Only	38	332	4	6
Time Comp. Only	0	1	1	48
Cost & Time Comp.	6	0	16	248

In the fourth experiment with the kNN algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.26), kNN classifiers using RCC technique have an average classification accuracy of '76.59%' with lower and upper bounds (74.95% - 78.22%) within 95% CI.

Table 5.26. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
KNN RCC	Accuracy(%)	79.27	74.39	74.39	79.27	75.61	75.61	76.83	73.17	79.27	78.05	76.59
	Kappa	0.671	0.591	0.601	0.670	0.614	0.618	0.634	0.582	0.673	0.653	0.631
	Precision	0.751	0.695	0.716	0.741	0.704	0.726	0.732	0.706	0.773	0.738	0.728
	Recall	0.793	0.744	0.744	0.793	0.756	0.756	0.768	0.732	0.793	0.780	0.766
	Specificity	0.897	0.863	0.890	0.896	0.881	0.892	0.889	0.885	0.898	0.895	0.889
	AUROC	0.859	0.846	0.856	0.874	0.866	0.862	0.843	0.860	0.884	0.875	0.863

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	47	43	0	30
Cost Comp. Only	36	333	4	7
Time Comp. Only	0	0	1	49
Cost & Time Comp.	9	4	10	247

The average for Kappa statistic value is ‘0.631’ that shows a substantial agreement. The weighted average precision value is ‘0.728’. The weighted average sensitivity (recall) value is ‘0.766’ that means the success of classifiers in identifying true positive instances is ‘76.6%’. Similarly, the weighted average specificity value is ‘0.889’ showing the algorithm achieved ‘88.9%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.863’.

In the final experiment with the kNN algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.27), kNN classifiers using ECC technique have an average classification accuracy of ‘76.59%’ with lower and upper bounds (75.59% - 77.58%) within 95% CI.

The average for Kappa statistic value is ‘0.631’ that shows a substantial agreement. The weighted average precision value is ‘0.730’. The weighted average sensitivity (recall) value is ‘0.766’ that means the success of classifiers in identifying true positive instances is ‘76.6%’. Similarly, the weighted average specificity value is ‘0.892’ showing the algorithm achieved ‘89.2%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.896’.

Table 5.27. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN ECC	Accuracy(%)	78.05	76.83	76.83	78.05	74.39	75.61	76.83	74.39	78.05	76.83	76.59
	Kappa	0.652	0.631	0.636	0.653	0.597	0.618	0.637	0.604	0.654	0.633	0.631
	Precision	0.735	0.717	0.737	0.738	0.707	0.719	0.743	0.725	0.745	0.729	0.730
	Recall	0.780	0.768	0.768	0.780	0.744	0.756	0.768	0.744	0.780	0.768	0.766
	Specificity	0.890	0.880	0.894	0.895	0.881	0.891	0.899	0.895	0.900	0.893	0.892
	AUROC	0.904	0.906	0.905	0.888	0.889	0.896	0.896	0.890	0.893	0.892	0.896

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	48	42	0	30
Cost Comp. Only	37	333	3	7
Time Comp. Only	0	2	0	48
Cost & Time Comp.	9	0	14	247

Considering results of experiments using the kNN algorithm, it can be said that the best performing kNN classifier is obtained from the natural solution (no decomposition technique used) that achieved '78.66%' average prediction accuracy for potential compensation prediction. In addition, this classifier is superior to other experimented kNN classifiers in all remaining performance measures. However, as given in Section 5.3.1.2, the Naïve Bayes OVA technique outperformed the kNN algorithm for potential compensation prediction.

5.3.1.6. Results from the kNN Algorithm for Resolution Method Selection

In the first experiment, 10-fold cross-validation results with 10 repeats obtained from using the kNN algorithm without any decomposition technique (natural solution) are given in Table 5.28. The optimum k parameter is determined as k=3 by the cross-validation parameter selection algorithm. Similar to previous experiments with the kNN, the best results are obtained from using (1 / weight) distance weighting method with the Manhattan distance measure. According to these results, kNN classifiers have an average multiclass classification accuracy of '73.52%' with lower and upper bounds (70.79% - 76.24%) within 95% CI. In other words, the kNN algorithm without using decomposition techniques predicts the resolution method to be used in construction disputes with an average success rate of '73.52%'.

The average for Kappa statistic value is '0.661' that shows a substantial agreement. The weighted average precision value is '0.760'. The weighted average sensitivity (recall) value is '0.735' that means the success of classifiers in identifying true positive instances is '73.5%'. Similarly, the weighted average specificity value is '0.913' showing the algorithm achieved '91.3%' success in identifying true negative instances. The weighted average AUROC value is '0.894'.

Table 5.28. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm without Using Decomposition Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN **No Decomp.**	Accuracy(%)	81.48	70.37	70.37	72.22	77.78	74.07	74.07	68.52	72.22	74.07	73.52
	Kappa	0.764	0.624	0.620	0.646	0.714	0.669	0.668	0.598	0.643	0.671	0.661
	Precision	0.822	0.744	0.734	0.741	0.802	0.772	0.759	0.716	0.738	0.769	0.760
	Recall	0.815	0.704	0.704	0.722	0.778	0.741	0.741	0.685	0.722	0.741	0.735
	Specificity	0.937	0.912	0.896	0.914	0.914	0.910	0.917	0.900	0.907	0.922	0.913
	AUROC	0.903	0.880	0.879	0.894	0.897	0.888	0.916	0.892	0.900	0.888	0.894

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	72	10	3	0	0	5
Arbitration	6	54	0	0	0	0
DRB	0	0	42	0	8	0
Mediation	0	1	0	32	0	17
SEA	0	0	7	0	61	32
Negotiation	0	0	0	0	54	136

In the second experiment with the kNN algorithm, the OVO technique is utilized. However, the resolution method selection model learns from a dataset composed of 54 instances and the kNN algorithm using OVO decomposition technique cannot output classification results due to lack of enough instances. Therefore, the kNN OVO algorithm is not considered in comparisons.

In the third experiment with the kNN algorithm, the OVA technique is utilized. The kNN OVA algorithm generates six distinct classifiers (since there are six output classes) and according to 10-fold cross-validation results with 10 repeats (Table 5.29), kNN classifiers using OVA technique have an average classification accuracy of '73.89' with lower and upper bounds (71.60% - 76.18%) within 95% CI. In other words, the kNN OVA algorithm predicts the resolution method to be used in construction disputes with an average success rate of '73.89%'.

The average for Kappa statistic value is '0.665' that shows a substantial agreement. The weighted average precision value is '0.758' showing the positive predicting power of the algorithm. The weighted average sensitivity (recall) value is '0.739' that means the success of classifiers in identifying true positive instances is ‘73.9%’. Similarly, the weighted average specificity value is ‘0.911’ showing the algorithm achieved ‘91.1%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.888’.

Table 5.29. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using OVA Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
KNN OVA	Accuracy(%)	81.48	74.07	72.22	74.07	74.07	72.22	74.07	70.37	70.37	75.93	73.89
	Kappa	0.764	0.667	0.644	0.664	0.666	0.643	0.668	0.621	0.621	0.691	0.665
	Precision	0.822	0.767	0.736	0.756	0.751	0.745	0.759	0.745	0.725	0.772	0.758
	Recall	0.815	0.741	0.722	0.741	0.741	0.722	0.741	0.704	0.704	0.759	0.739
	Specificity	0.937	0.910	0.907	0.903	0.911	0.898	0.917	0.900	0.902	0.921	0.911
	AUROC	0.888	0.909	0.864	0.902	0.906	0.869	0.882	0.876	0.887	0.901	0.888

Confusion Matrix

Actual	Predicted: Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	75	9	1	0	0	5
Arbitration	9	51	0	0	0	0
DRB	0	0	42	0	8	0
Mediation	0	1	0	33	0	16
SEA	0	0	5	0	58	37
Negotiation	0	0	0	0	50	140

In the fourth experiment with the kNN algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.30), kNN classifiers using RCC technique have an average classification accuracy of ‘73.15%’ with lower and upper bounds (70.87% - 75.42%) within 95% CI. In other words, the kNN RCC algorithm predicts the resolution method to be used in construction disputes with an average success rate of ‘73.15%’.

The average for Kappa statistic value is ‘0.657’ that shows a substantial agreement. The weighted average precision value is ‘0.749’ showing the positive predictive power of the algorithm. The weighted average sensitivity (recall) value is ‘0.732’ that means the success of classifiers in identifying true positive instances is ‘73.2%’. Similarly, the weighted average specificity value is ‘0.913’ showing the algorithm achieved

‘91.3%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.874’.

Table 5.30. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using RCC Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN RCC	Accuracy(%)	77.78	75.93	74.07	72.22	72.22	72.22	70.37	66.67	75.93	74.07	73.15
	Kappa	0.718	0.693	0.668	0.646	0.643	0.645	0.622	0.574	0.691	0.667	0.657
	Precision	0.785	0.773	0.767	0.748	0.738	0.738	0.723	0.689	0.769	0.760	0.749
	Recall	0.778	0.759	0.741	0.722	0.722	0.722	0.704	0.667	0.759	0.741	0.732
	Specificity	0.931	0.923	0.910	0.916	0.907	0.902	0.908	0.896	0.923	0.913	0.913
	AUROC	0.869	0.870	0.860	0.883	0.888	0.861	0.871	0.863	0.890	0.888	0.874

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	69	10	5	0	0	6
Arbitration	4	56	0	0	0	0
DRB	0	0	42	0	8	0
Mediation	0	1	1	32	0	16
SEA	1	3	6	0	56	34
Negotiation	0	1	0	3	46	140

In the final experiment with the kNN algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.31), kNN classifiers using ECC technique have an average classification accuracy of ‘74.63%’ with lower and upper bounds (72.46% - 76.80%) within 95% CI. In other words, the kNN ECC algorithm predicts the predicts the resolution method to be used in construction disputes with an average success rate of ‘74.63%’.

The average for Kappa statistic value is ‘0.674’ that shows a substantial agreement. The weighted average precision value is ‘0.769’ showing the positive predictive power of the algorithm. The weighted average sensitivity (recall) value is ‘0.746’ that means the success of classifiers in identifying true positive instances is ‘74.6%’. Similarly, the weighted average specificity value is ‘0.910’ showing the algorithm achieved ‘91.0%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.908’, which is the highest AUROC value among experimented kNN algorithms for resolution method selection.

Table 5.31. 10-Times 10-Fold Cross-Validation Results of the KNN Algorithm Using ECC Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
KNN ECC	Accuracy(%)	79.63	74.07	74.07	72.22	75.93	75.93	74.07	68.52	74.07	77.78	74.63
	Kappa	0.739	0.667	0.668	0.645	0.690	0.690	0.664	0.593	0.667	0.715	0.674
	Precision	0.804	0.767	0.774	0.755	0.776	0.783	0.754	0.719	0.759	0.797	0.769
	Recall	0.796	0.741	0.741	0.722	0.759	0.759	0.741	0.685	0.741	0.778	0.746
	Specificity	0.927	0.910	0.910	0.906	0.915	0.906	0.906	0.885	0.910	0.925	0.910
	AUROC	0.898	0.921	0.901	0.915	0.909	0.899	0.917	0.880	0.915	0.920	0.908

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	72	7	6	0	0	5
Arbitration	2	58	0	0	0	0
DRB	0	0	42	0	8	0
Mediation	0	1	0	33	0	16
SEA	0	0	4	0	55	41
Negotiation	0	0	0	0	47	143

Considering results of experiments using the kNN algorithm, it can be said that the best performing kNN classifier is obtained from the ECC technique that achieved ‘74.63%’ average prediction accuracy for resolution method selection. In addition, this classifier is superior to other experimented kNN classifiers in all remaining performance measures except the specificity measure. In terms of specificity, the best classifiers are obtained from the kNN algorithm without using decomposition techniques and the kNN RCC algorithm. However, as given in Section 5.3.1.3, the Naïve Bayes ECC technique (85.93% average prediction accuracy) significantly outperformed the kNN ECC algorithm for resolution method selection.

5.3.1.7. The J48 Algorithm and its Configuration in WEKA

The J48 algorithm can naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. The configuration of the J48 algorithm in WEKA for binary classification can be used exactly the same way to obtain multiclass solutions. In addition to this, both problems can be solved by decomposing them into several binary problems using the class 'weka.classifiers.meta.MultiClassClassifier'. Configuration details for the J48 algorithm using decomposition techniques in WEKA can be seen in Figure 5.21.

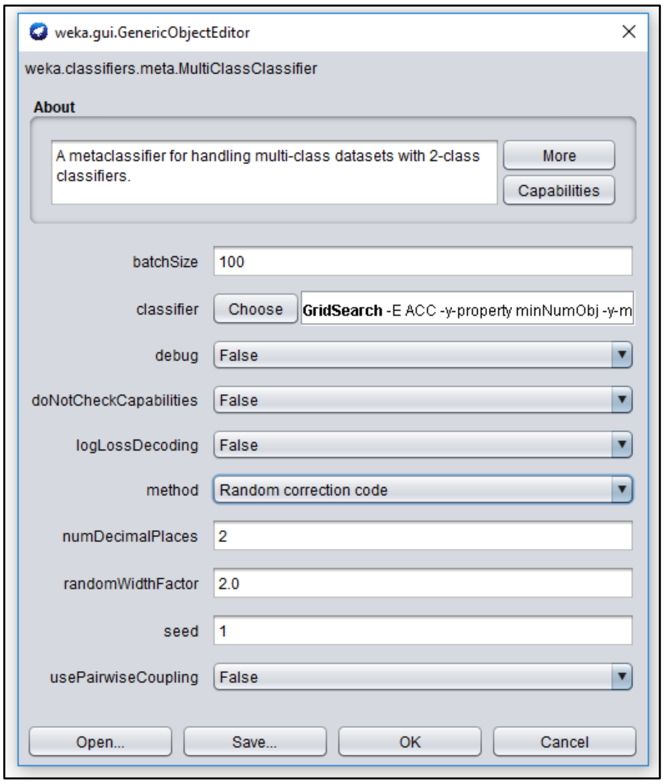


Figure 5.21. The Multiclass J48 Classifier Configuration in WEKA

To select the J48 algorithm, the 'classifier' setting should be set to J48 with the configuration used in binary classification task ('GridSearch') (Figure 5.6 and Figure 5.7). The 'method' setting allows the user to select the decomposition technique. The J48 algorithm is tested using all available decomposition techniques, which are OVO, OVA, RCC, and ECC. In addition, a solution is obtained when no decomposition techniques used (natural solution). Default values in WEKA are used for remaining settings.

5.3.1.8. Results from the J48 Algorithm for Potential Compensation Prediction

In the first experiment, 10-fold cross-validation results with 10 repeats obtained from using the J48 algorithm without any decomposition technique (natural solution) are given in Table 5.32. The grid search algorithm was used to optimize two parameters; confidence factor and minimum number of instances at leaf nodes. The optimum confidence factor is determined as '1' while, the optimum value for minimum number of instances at leaf nodes is '4'. According to these results, J48 classifiers without using any decomposition technique have an average multiclass classification accuracy of '76.59%' with lower and upper bounds (75.99% - 77.91%) within 95% CI. In other words, the J48 algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '76.95%' that makes it the best J48 classifier in terms of classification accuracy for potential compensation prediction.

Table 5.32. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm without Using Decomposition Techniques for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 No Decomp.	Accuracy(%)	78.05	78.05	78.05	78.05	74.39	76.83	76.83	78.05	75.61	75.61	76.95
	Kappa	0.631	0.631	0.636	0.631	0.578	0.613	0.615	0.631	0.595	0.599	0.616
	Precision	0.620	0.620	0.699	0.620	0.613	0.618	0.625	0.620	0.615	0.670	0.632
	Recall	0.780	0.780	0.780	0.780	0.744	0.768	0.768	0.780	0.756	0.756	0.769
	Specificity	0.851	0.851	0.860	0.851	0.846	0.849	0.856	0.851	0.850	0.857	0.852
	AUROC	0.817	0.807	0.823	0.799	0.788	0.820	0.808	0.813	0.812	0.826	0.811

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	2	88	0	30
Cost Comp. Only	5	365	1	9
Time Comp. Only	0	0	0	50
Cost & Time Comp.	1	0	5	264

The highest average for Kappa statistic value among J48 classifiers belongs to this algorithm with a Kappa value of '0.616' that shows a substantial agreement. The weighted average precision value is '0.632'. The highest weighted average sensitivity (recall) value is also obtained from this algorithm as '0.769' that means the success of classifiers in identifying true positive instances is '76.9%'. The weighted average specificity value is '0.852' showing the algorithm achieved '85.2%' success in identifying true negative instances. The weighted average AUROC value is '0.811'.

In the second experiment with the J48 algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.33), J48 classifiers using OVO technique have an average classification accuracy of '75.37%' with lower and upper bounds (73.32% - 77.41%) within 95% CI. In other words, the J48 OVO algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '75.37%'.

The average for Kappa value is '0.605' that shows a substantial agreement. The highest weighted average precision value among experimented J48 algorithms is obtained from this algorithm as '0.686'. The weighted average sensitivity (recall) value is '0.754' that means the success of classifiers in identifying true positive instances is '75.4%'. The highest weighted average specificity value is also obtained from this algorithm as '0.868' showing the algorithm achieved '86.8%' success in identifying true negative instances. The weighted average AUROC value is '0.836'.

Table 5.33. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 OVO Pairwise Coupling	Accuracy(%)	73.17	78.05	71.95	76.83	76.83	69.51	76.83	78.05	75.61	76.83	75.37
	Kappa	0.566	0.641	0.568	0.628	0.628	0.514	0.618	0.650	0.611	0.623	0.605
	Precision	0.647	0.706	0.701	0.702	0.702	0.627	0.673	0.722	0.692	0.690	0.686
	Recall	0.732	0.780	0.720	0.768	0.768	0.695	0.768	0.780	0.756	0.768	0.754
	Specificity	0.853	0.868	0.884	0.875	0.875	0.845	0.858	0.885	0.872	0.866	0.868
	AUROC	0.840	0.846	0.848	0.828	0.823	0.834	0.813	0.855	0.803	0.870	0.836

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	25	65	0	30
Cost Comp. Only	41	329	0	10
Time Comp. Only	0	0	0	50
Cost & Time Comp.	4	0	2	264

In the third experiment with the J48 algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.34), J48 classifiers using OVA technique have an average classification accuracy of '75.12%' with lower and upper bounds (73.94% - 76.30%) within 95 CI. In other words, the J48 OVA algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '75.12%'.

Table 5.34. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 OVA	Accuracy(%)	76.83	75.61	75.61	74.39	74.39	71.95	76.83	76.83	73.17	75.61	75.12
	Kappa	0.613	0.596	0.604	0.581	0.580	0.553	0.615	0.618	0.560	0.595	0.592
	Precision	0.617	0.615	0.677	0.655	0.620	0.659	0.625	0.673	0.611	0.622	0.637
	Recall	0.768	0.756	0.756	0.744	0.744	0.720	0.768	0.768	0.732	0.756	0.751
	Specificity	0.851	0.847	0.864	0.850	0.852	0.859	0.856	0.858	0.846	0.851	0.853
	AUROC	0.864	0.833	0.837	0.831	0.819	0.826	0.887	0.841	0.767	0.856	0.836

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	6	85	0	29
Cost Comp. Only	16	353	3	8
Time Comp. Only	1	0	0	49
Cost & Time Comp.	7	0	6	257

The average for Kappa statistic value is '0.592' that shows a moderate agreement. The weighted average precision value is '0.637'. The weighted average sensitivity (recall) value is '0.751' that means the success of classifiers in identifying true positive instances is '75.1%'. Similarly, the weighted average specificity value is '0.853' showing the algorithm achieved '85.3%' success in identifying true negative instances. The weighted average AUROC value is '0.836'.

In the fourth experiment with the J48 algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.35), J48 classifiers using RCC technique have an average classification accuracy of '74.51%'

with lower and upper bounds (72.61% - 76.42%) within 95% CI. In other words, the J48 RCC classifier have an average success rate of '74.51%'.

The average for Kappa statistic value is '0.584' that shows a moderate agreement. The weighted average precision value is '0.665'. The weighted average sensitivity (recall) value is '0.745' that means the success of classifiers in identifying true positive instances is '74.5%'. Similarly, the weighted average specificity value is '0.852' showing the algorithm achieved '85.2%' success in identifying true negative instances. The weighted average AUROC value is '0.845'.

Table 5.35. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 RCC	Accuracy(%)	78.05	74.39	73.17	76.83	75.61	69.51	75.61	75.61	75.61	70.73	74.51
	Kappa	0.631	0.579	0.568	0.615	0.597	0.519	0.600	0.599	0.600	0.532	0.584
	Precision	0.620	0.647	0.724	0.675	0.622	0.649	0.694	0.688	0.694	0.641	0.665
	Recall	0.780	0.744	0.732	0.768	0.756	0.695	0.756	0.756	0.756	0.707	0.745
	Specificity	0.851	0.841	0.847	0.845	0.855	0.856	0.858	0.853	0.858	0.854	0.852
	AUROC	0.841	0.818	0.814	0.881	0.847	0.837	0.863	0.840	0.850	0.863	0.845

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	10	80	0	30
Cost Comp. Only	18	350	2	10
Time Comp. Only	1	1	2	46
Cost & Time Comp.	6	5	10	249

In the final experiment with the J48 algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.36), J48 classifiers using ECC technique have an average classification accuracy of '75.73%' with lower and upper bounds (74.69% - 76.78%) within 95% CI. In other words, J48 ECC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '75.73%'. This is the second best J48 classifier in terms of classification accuracy with a slightly worse performance than the best performing J48 classifier (natural solution).

The average for Kappa statistic value is ‘0.601’ that shows a substantial agreement. The weighted average precision value is ‘0.642’. The weighted average sensitivity (recall) value is ‘0.757’ that means the success of classifiers in identifying true positive instances is ‘75.7%’. Similarly, the weighted average specificity value is ‘0.855’ showing the algorithm achieved ‘85.5%’ success in identifying true negative instances. Among experimented J48 classifiers, the highest weighted average AUROC value is obtained from this algorithm as ‘0.889’.

Table 5.36. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 ECC	Accuracy(%)	75.61	76.83	75.61	76.83	74.39	74.39	76.83	78.05	73.17	75.61	75.73
	Kappa	0.596	0.623	0.604	0.613	0.580	0.583	0.615	0.631	0.566	0.597	0.601
	Precision	0.615	0.704	0.687	0.618	0.620	0.656	0.625	0.620	0.654	0.622	0.642
	Recall	0.756	0.768	0.756	0.768	0.744	0.744	0.768	0.780	0.732	0.756	0.757
	Specificity	0.847	0.867	0.865	0.849	0.852	0.855	0.856	0.851	0.854	0.851	0.855
	AUROC	0.893	0.894	0.884	0.888	0.889	0.875	0.924	0.869	0.873	0.904	0.889

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	6	84	0	30
Cost Comp. Only	17	353	2	8
Time Comp. Only	1	0	0	49
Cost & Time Comp.	1	0	7	262

In the light of classification results from J48 classifiers for potential compensation prediction, performances of various J48 algorithms are very similar. There is no single J48 classifier that outperforms others in majority of the performance measures. Therefore, the primary evaluation criterion, which is the average classification accuracy, is used to determine the best J48 classifier. According to this, it can be said that the best performing J48 classifier is obtained from the one with no decomposition technique (natural solution). It achieved ‘76.95%’ average prediction accuracy for potential compensation prediction. However, this classifier is outperformed by the Naïve Bayes OVA and the multiclass kNN algorithms.

5.3.1.9. Results from the J48 Algorithm for Resolution Method Selection

In the first experiment, 10-fold cross-validation results with 10 repeats obtained from using the multiclass J48 algorithm (natural solution) are given in Table 5.37. According to these results, multiclass J48 classifiers have an average multiclass classification accuracy of '80.37%' with lower and upper bounds (77.78% - 82.96%) within 95% CI. In other words, the J48 algorithm predicts the resolution method to be used in construction disputes with an average success rate of '80.37%'.

Table 5.37. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm without Using Decomposition Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 No Decomp.	Accuracy(%)	81.48	83.33	85.19	72.22	83.33	81.48	79.63	79.63	79.63	77.78	80.37
	Kappa	0.767	0.791	0.814	0.652	0.791	0.767	0.744	0.744	0.744	0.721	0.753
	Precision	0.830	0.850	0.866	0.751	0.850	0.830	0.810	0.811	0.812	0.791	0.820
	Recall	0.815	0.833	0.852	0.722	0.833	0.815	0.796	0.796	0.796	0.778	0.804
	Specificity	0.960	0.963	0.965	0.940	0.963	0.960	0.951	0.957	0.956	0.954	0.957
	AUROC	0.937	0.945	0.944	0.886	0.955	0.946	0.947	0.947	0.924	0.941	0.937

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	58	14	18	0	0	0
Arbitration	11	49	0	0	0	0
DRB	10	0	40	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	89	11
Negotiation	0	0	0	0	42	148

The average for Kappa statistic value is '0.753' that shows a substantial agreement. The weighted average precision value is '0.820'. The weighted average sensitivity (recall) value is '0.804' that means the success of classifiers in identifying true positive instances is '80.4%'. The weighted average specificity value is '0.957' showing the algorithm achieved '95.7%' success in identifying true negative instances. The weighted average AUROC value is '0.937'.

In the second experiment with the J48 algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats

(Table 5.38), J48 classifiers using OVO technique have an average classification accuracy of '82.41%' with lower and upper bounds (80.51% - 84.31%) within 95% CI. In other words, the J48 OVO algorithm predicts the resolution method to be used in construction disputes with an average success rate of '82.41%'.

Table 5.38. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using OVO Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 OVO Pairwise Coupling	Accuracy(%)	79.63	87.04	81.48	79.63	81.48	79.63	85.19	85.19	83.33	81.48	82.41
	Kappa	0.737	0.834	0.765	0.741	0.765	0.741	0.813	0.813	0.786	0.765	0.776
	Precision	0.805	0.868	0.829	0.806	0.821	0.814	0.869	0.860	0.831	0.822	0.833
	Recall	0.796	0.870	0.815	0.796	0.815	0.796	0.852	0.852	0.833	0.815	0.824
	Specificity	0.926	0.947	0.950	0.940	0.945	0.946	0.963	0.957	0.935	0.943	0.945
	AUROC	0.941	0.932	0.954	0.933	0.948	0.945	0.968	0.962	0.955	0.952	0.949

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	81	9	0	0	0	0
Arbitration	13	47	0	0	0	0
DRB	6	0	44	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	70	30
Negotiation	0	0	0	0	37	153

The average for Kappa value is '0.776' that shows a substantial agreement. The weighted average precision value is '0.833'. The weighted average sensitivity (recall) value is '0.824' that means the success of classifiers in identifying true positive instances is '82.4%'. The weighted average specificity value is '0.945' showing the algorithm achieved '94.5%' success in identifying true negative instances. The weighted average AUROC value is '0.949'.

In the third experiment with the J48 algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.39), J48 classifiers using OVA technique have an average classification accuracy of '80.56%' with lower and upper bounds (78.87% - 82.24%) within 95 CI. In other words, the J48

OVA algorithm predicts the resolution method to be used in construction disputes with an average success rate of '80.56%'.

The average for Kappa statistic value is '0.754' that shows a substantial agreement. The weighted average precision value is '0.834'. The weighted average sensitivity (recall) value is '0.806' that means the success of classifiers in identifying true positive instances is '80.6%'. Similarly, the weighted average specificity value is '0.952' showing the algorithm achieved '95.2%' success in identifying true negative instances. The weighted average AUROC value is '0.929'.

Table 5.39. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using OVA Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 OVA	Accuracy(%)	79.63	79.63	83.33	81.48	81.48	85.19	79.63	77.78	77.78	79.63	80.56
	Kappa	0.743	0.741	0.789	0.768	0.765	0.813	0.742	0.719	0.718	0.743	0.754
	Precision	0.820	0.844	0.860	0.852	0.829	0.873	0.821	0.813	0.798	0.826	0.834
	Recall	0.796	0.796	0.833	0.815	0.815	0.852	0.796	0.778	0.778	0.796	0.806
	Specificity	0.954	0.946	0.954	0.958	0.946	0.965	0.948	0.950	0.944	0.956	0.952
	AUROC	0.941	0.926	0.913	0.916	0.939	0.941	0.928	0.928	0.928	0.932	0.929

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	71	19	0	0	0	0
Arbitration	1	57	2	0	0	0
DRB	6	0	23	0	26	1
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	85	15
Negotiation	0	0	7	0	34	149

In the fourth experiment with the J48 algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.40), J48 classifiers using RCC technique have an average classification accuracy of '85.00%' with lower and upper bounds (83.41% - 86.59%) within 95% CI. In other words, the J48 random correction code classifier predicts the resolution method to be used in construction disputes with an average success rate of '85.00%'.

The average for Kappa statistic value is '0.811' that shows a perfect agreement. The weighted average precision value is '0.869'. The weighted average sensitivity (recall) value is '0.850' that means the success of classifiers in identifying true positive instances is '85.0%'. Similarly, the weighted average specificity value is '0.963' showing the algorithm achieved '96.3' success in identifying true negative instances. The weighted average AUROC value is '0.940'.

Table 5.40. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using RCC Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 RCC	Accuracy(%)	85.19	87.04	87.04	85.19	85.19	85.19	85.19	87.04	83.33	79.63	85.00
	Kappa	0.813	0.836	0.836	0.814	0.813	0.813	0.814	0.836	0.791	0.744	0.811
	Precision	0.872	0.886	0.886	0.866	0.870	0.876	0.872	0.896	0.851	0.819	0.869
	Recall	0.852	0.870	0.870	0.852	0.852	0.852	0.852	0.870	0.833	0.796	0.850
	Specificity	0.966	0.961	0.961	0.965	0.959	0.967	0.965	0.967	0.963	0.957	0.963
	AUROC	0.940	0.943	0.943	0.938	0.936	0.960	0.914	0.949	0.946	0.927	0.940

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	71	9	8	0	0	2
Arbitration	1	59	0	0	0	0
DRB	3	0	42	1	4	0
Mediation	0	0	1	48	0	1
SEA	0	2	0	0	88	10
Negotiation	0	0	0	0	39	151

In the final experiment with the J48 algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.41), J48 classifiers using ECC technique have an average classification accuracy of '86.48%' with lower and upper bounds (85.08% - 87.88%) within 95% CI. In other words, J48 ECC algorithm predicts the resolution method to be used in construction disputes with an average success rate of '86.48%' that makes it the best J48 classifier in terms of classification accuracy for resolution method selection.

Among all experimented J48 classifiers, the highest average for Kappa statistic value is obtained from this algorithm as '0.830' that shows a perfect agreement. The highest weighted average precision value is also obtained from this algorithm as ‘0.879’. Similarly, the highest weighted average sensitivity (recall) value (‘0.879’) and the highest weighted average specificity value (‘0.865’) is obtained from J48 ECC algorithm. In other words, the algorithm achieved ‘87.9%’ success in identifying true positive instances and ‘86.5%’ success in identifying true negative instances. The highest weighted average AUROC value is also obtained from the J48 ECC algorithm as ‘0.964’, which is almost an ideal AUROC value.

Table 5.41. 10-Times 10-Fold Cross-Validation Results of the J48 Algorithm Using ECC Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
J48 ECC	Accuracy(%)	85.19	85.19	90.74	88.89	85.19	85.19	85.19	87.04	85.19	87.04	86.48
	Kappa	0.814	0.814	0.883	0.860	0.814	0.814	0.814	0.837	0.814	0.837	0.830
	Precision	0.866	0.866	0.921	0.905	0.866	0.866	0.866	0.884	0.869	0.884	0.879
	Recall	0.852	0.852	0.907	0.889	0.852	0.852	0.852	0.870	0.852	0.870	0.865
	Specificity	0.965	0.965	0.973	0.971	0.965	0.965	0.965	0.968	0.966	0.968	0.967
	AUROC	0.956	0.956	0.969	0.968	0.965	0.967	0.966	0.955	0.972	0.961	0.964

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	75	6	9	0	0	0
Arbitration	0	60	0	0	0	0
DRB	8	0	42	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	90	10
Negotiation	0	0	0	0	40	150

In the light of results from J48 classifiers for resolution method selection, the J48 ECC algorithm outperformed others in every performance measure. Thus, it is the best performing J48 classifier achieving ‘86.48%’ average prediction accuracy. In addition, the J48 ECC algorithm outperformed the Naïve Bayes ECC and kNN ECC algorithms in terms of prediction accuracy for resolution method selection. Finally, the J48 ECC algorithm is perfect in predicting arbitration and mediation cases and relatively powerful in predicting litigation, DRB, and SEA cases.

5.3.1.10. The MLP and its Configuration in WEKA

The MLP algorithm can naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. The configuration of the MLP algorithm in WEKA for binary classification can be used exactly the same way to obtain multiclass solutions. In addition to this, both problems can be solved by decomposing them into several binary problems using the class 'weka.classifiers.meta.MultiClassClassifier'. Similar to the binary classification with MLP algorithm, best results are obtained from using 'a' hidden layers. For potential compensation prediction, 'a' corresponds to '7' ((9 attributes + 4 classes) / 2 = 6.5 ≈ 7). For resolution method selection, 'a' corresponds to '7' ((7 attributes + 6 classes) / 2 = 6.5 ≈ 7). Configuration details for the MLP algorithm using decomposition techniques in WEKA can be seen in Figure 5.22.

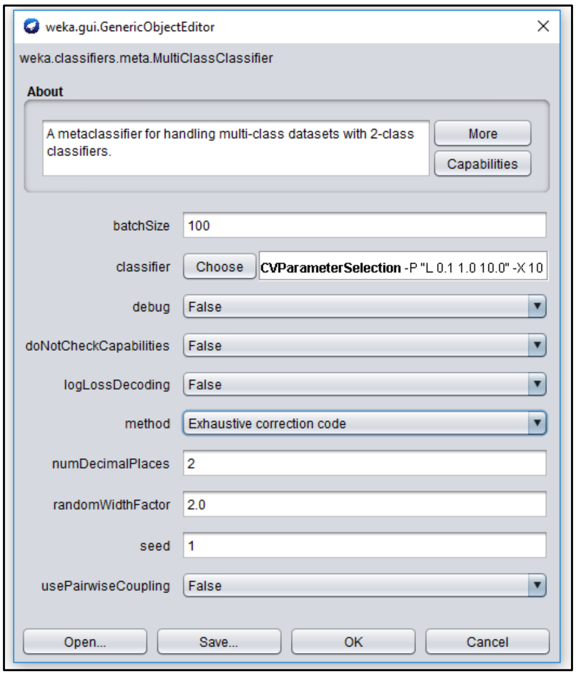


Figure 5.22. The Multiclass MLP Classifier Configuration in WEKA

To select the MLP algorithm, the 'classifier' setting should be set to MLP with the configuration used in binary classification task ('CVParameterSelection' with 'a' hidden layers) (Figure 5.8 and Figure 5.9). The 'method' setting allows the user to select the decomposition technique. Default values in WEKA are used for remaining settings.

5.3.1.11. Results from the MLP for Potential Compensation Prediction

10-fold cross-validation results with 10 repeats obtained from using the MLP algorithm without any decomposition technique are given in Table 5.42. According to these results, MLP classifiers with no decomposition techniques have an average multiclass classification accuracy of '66.95%' with lower and upper bounds (64.50% - 69.40%) within 95% CI. In other words, the MLP multiclass algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '66.95%'.

The average for Kappa statistic value is '0.491' that shows a moderate agreement. The weighted average precision value is '0.657'. The weighted average sensitivity (recall) value is '0.670' that means the success of classifiers in identifying true positive instances is '67.0%'. Similarly, the weighted average specificity value is '0.858' showing the algorithm achieved '85.8%' success in identifying true negative instances. The weighted average AUROC value is '0.824'.

Table 5.42. 10-Times 10-Fold Cross-Validation Results of MLP Algorithm without Using Decomposition Techniques for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP No Decomp.	Accuracy(%)	74.39	65.85	67.07	64.63	64.63	68.29	68.29	64.63	69.51	62.20	66.95
	Kappa	0.607	0.484	0.494	0.454	0.442	0.509	0.506	0.454	0.531	0.429	0.491
	Precision	0.740	0.671	0.676	0.620	0.592	0.657	0.659	0.632	0.684	0.637	0.657
	Recall	0.744	0.659	0.671	0.646	0.646	0.683	0.683	0.646	0.695	0.622	0.670
	Specificity	0.891	0.875	0.858	0.845	0.830	0.861	0.852	0.840	0.873	0.853	0.858
	AUROC	0.860	0.815	0.853	0.844	0.816	0.829	0.797	0.810	0.799	0.815	0.824

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	16	67	10	27
Cost Comp. Only	72	297	6	5
Time Comp. Only	3	1	10	36
Cost & Time Comp.	5	9	30	226

In the second experiment with the MLP algorithm, the OVO technique is used with pairwise coupling. According to 10-fold cross-validation results with 10 repeats (Table 5.43), MLP OVO classifiers with pairwise coupling have an average multiclass classification accuracy of ‘68.54%’ with lower and upper bounds (67.01% - 70.06%) within 95% CI. In other words, the MLP OVO algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘68.54%’.

The average for Kappa statistic value is ‘0.512’ that shows a moderate agreement. The weighted average precision value is ‘0.658’. The weighted average sensitivity (recall) value is ‘0.686’ that means the success of classifiers in identifying true positive instances is ‘68.6%’. The weighted average specificity value is ‘0.859’ showing the algorithm achieved ‘85.9%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.837’.

Table 5.43. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP OVO Pairwise Coupling	Accuracy(%)	71.95	69.51	67.07	65.85	69.51	67.07	68.29	71.95	67.07	67.07	68.54
	Kappa	0.561	0.523	0.491	0.481	0.517	0.497	0.503	0.567	0.487	0.494	0.512
	Precision	0.687	0.656	0.648	0.660	0.632	0.668	0.639	0.703	0.630	0.653	0.658
	Recall	0.720	0.695	0.671	0.659	0.695	0.671	0.683	0.720	0.671	0.671	0.686
	Specificity	0.869	0.858	0.857	0.860	0.848	0.860	0.855	0.878	0.851	0.858	0.859
	AUROC	0.829	0.855	0.871	0.843	0.846	0.821	0.817	0.853	0.826	0.811	0.837

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	13	77	7	23
Cost Comp. Only	76	295	5	4
Time Comp. Only	0	0	9	41
Cost & Time Comp.	0	0	25	245

In the third experiment with the MLP algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.44), MLP classifiers using OVA technique have an average classification accuracy of ‘68.63%’

with lower and upper bounds (67.03% - 70.23%) within 95% CI. In other words, the MLP OVA algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '68.63%' that makes it the best performing MLP algorithm for potential compensation prediction.

Among all experimented MLP algorithms, the highest average Kappa statistic value ('0.512') and the highest weighted average precision value ('0.666') is obtained from the MLP OVA algorithm. In addition, the highest weighted average sensitivity (recall) value ('0.686') and the highest weighted average specificity value ('0.864') are also obtained from this algorithm compared to other MLP classifiers. In other words, this algorithm achieved '68.6%' success in identifying true positive instances and '86.4%' success in identifying true negative instances. The weighted average AUROC value is '0.842'.

Table 5.44. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP OVA	Accuracy(%)	71.95	68.29	73.17	68.29	67.07	67.97	67.07	68.29	68.29	65.85	68.63
	Kappa	0.564	0.511	0.576	0.504	0.489	0.489	0.502	0.511	0.503	0.479	0.512
	Precision	0.725	0.666	0.699	0.637	0.643	0.646	0.681	0.670	0.638	0.655	0.666
	Recall	0.720	0.683	0.732	0.683	0.671	0.671	0.671	0.683	0.683	0.659	0.686
	Specificity	0.879	0.868	0.868	0.853	0.861	0.851	0.881	0.863	0.857	0.861	0.864
	AUROC	0.850	0.853	0.862	0.835	0.833	0.820	0.833	0.855	0.833	0.850	0.842

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	20	65	8	27
Cost Comp. Only	65	303	9	3
Time Comp. Only	3	2	2	43
Cost & Time Comp.	3	3	27	237

In the fourth experiment with the MLP algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.45), MLP classifiers using RCC technique have an average classification accuracy of '68.02%' with lower and upper bounds (66.16% - 69.88%) within 95% CI. In other words, the

MLP RCC classifier algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '68.02%'.

The average for Kappa statistic value is '0.503' that shows a moderate agreement. The weighted average precision value is '0.648'. The weighted average sensitivity (recall) value is '0.681' that means the success of classifiers in identifying true positive instances is '68.1%'. The weighted average specificity value is '0.860' showing the algorithm achieved '86.0%' success in identifying true negative instances. The weighted average AUROC value is '0.832'.

Table 5.45. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP RCC	Accuracy(%)	69.51	67.07	67.07	71.95	64.34	63.41	69.51	69.51	68.29	69.51	68.02
	Kappa	0.521	0.492	0.502	0.565	0.446	0.433	0.529	0.518	0.497	0.523	0.503
	Precision	0.637	0.655	0.671	0.697	0.610	0.606	0.677	0.641	0.622	0.662	0.648
	Recall	0.695	0.671	0.671	0.720	0.646	0.634	0.695	0.695	0.683	0.695	0.681
	Specificity	0.868	0.862	0.879	0.874	0.834	0.844	0.874	0.857	0.842	0.863	0.860
	AUROC	0.819	0.794	0.851	0.866	0.821	0.843	0.814	0.842	0.825	0.844	0.832

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	12	69	12	27
Cost Comp. Only	61	304	10	5
Time Comp. Only	4	2	2	42
Cost & Time Comp.	5	3	22	240

In the final experiment with the MLP algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.46), MLP classifiers using ECC technique have an average classification accuracy of '68.48%' with lower and upper bounds (66.81% - 70.15%) within 95% CI. In other words, the MLP ECC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '68.48%'.

The average for Kappa statistic value is '0.505' that shows a moderate agreement. The weighted average precision value is '0.637'. The weighted average sensitivity (recall) value is ‘0.684’ that means the success of classifiers in identifying true positive instances is ‘68.4%’. The weighted average specificity value is ‘0.853’ showing the algorithm achieved ‘85.3%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.866’, which is the highest AUROC value among experimented MLP algorithms.

Table 5.46. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP ECC	Accuracy(%)	69.51	73.17	69.51	68.29	65.85	68.29	69.51	67.07	68.93	64.63	68.48
	Kappa	0.514	0.577	0.514	0.504	0.465	0.501	0.526	0.493	0.504	0.449	0.505
	Precision	0.619	0.686	0.613	0.637	0.617	0.636	0.666	0.655	0.634	0.603	0.637
	Recall	0.695	0.732	0.695	0.683	0.659	0.683	0.695	0.671	0.683	0.646	0.684
	Specificity	0.848	0.870	0.848	0.853	0.838	0.852	0.867	0.862	0.851	0.841	0.853
	AUROC	0.873	0.892	0.887	0.862	0.855	0.861	0.860	0.857	0.857	0.857	0.866

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	10	77	6	27
Cost Comp. Only	71	301	5	3
Time Comp. Only	2	2	1	45
Cost & Time Comp.	0	1	20	249

Considering results from MLP experiments, it can be seen that the MLP OVA algorithm outperformed others in every performance measure except the AUROC value for potential compensation prediction. Moreover, the MLP OVA algorithm has the second best AUROC value with a slightly worse performance behind the MLP ECC algorithm. Thus, it can be said that the best performing MLP classifier is obtained from the OVA technique that achieved ‘68.63%’ average prediction accuracy for potential compensation prediction. However, considering previous experiments, the performance of the MLP algorithm in multiclass classification of potential compensation types is relatively low compared to Naïve Bayes, kNN, and J48 algorithms.

5.3.1.12. Results from the MLP for Resolution Method Selection

10-fold cross-validation results with 10 repeats obtained from using the MLP algorithm without any decomposition technique are given in Table 5.47. According to these results, the MLP multiclass algorithm has an average multiclass classification accuracy of ‘83.15%’ with lower and upper bounds (81.04% - 85.26%) within 95% CI. In other words, the MLP multiclass algorithm predicts the resolution method to be used in construction projects with an average success rate of ‘83.15%’.

The average for Kappa statistic value is ‘0.788’ that shows a substantial agreement. The weighted average precision value is ‘0.851’. The weighted average sensitivity (recall) value is ‘0.831’ that means the success of classifiers in identifying true positive instances is ‘83.1%’. Similarly, the weighted average specificity value is ‘0.951’ showing the algorithm achieved ‘95.1%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.958’.

Table 5.47. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm without Using Decomposition Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP No Decomp.	Accuracy(%)	87.04	83.33	77.78	81.48	85.19	83.33	79.63	83.33	87.04	83.33	83.15
	Kappa	0.836	0.791	0.721	0.766	0.813	0.791	0.744	0.789	0.836	0.790	0.788
	Precision	0.881	0.857	0.800	0.839	0.869	0.858	0.827	0.846	0.881	0.848	0.851
	Recall	0.870	0.833	0.778	0.815	0.852	0.833	0.796	0.833	0.870	0.833	0.831
	Specificity	0.961	0.954	0.934	0.944	0.958	0.952	0.940	0.948	0.961	0.953	0.951
	AUROC	0.965	0.959	0.942	0.963	0.964	0.957	0.957	0.960	0.964	0.950	0.958

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	75	15	0	0	0	0
Arbitration	2	58	0	0	0	0
DRB	0	0	50	0	0	0
Mediation	0	0	0	48	0	2
SEA	0	0	0	0	78	22
Negotiation	0	0	0	0	50	140

In the second experiment with the MLP algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats

(Table 5.48), the MLP algorithm using OVO technique with pairwise coupling has an average classification accuracy of ‘81.30%’ with lower and upper bounds (79.71% - 82.88%) within 95% CI. In other words, the MLP OVO algorithm predicts the resolution method to be used in construction disputes with an average success rate of ‘81.30%’.

The average for Kappa statistic value is ‘0.766’ that shows a substantial agreement. The weighted average precision value is ‘0.844’. The weighted average sensitivity (recall) value is ‘0.813’ that means the success of classifiers in identifying true positive instances is ‘81.3%’. The weighted average specificity value is ‘0.947’ showing the algorithm achieved ‘94.7%’ success in identifying true negative instances. Among experimented MLP algorithms for resolution method selection, the highest weighted average AUROC value is obtained from this algorithm as ‘0.963’, which is almost an ideal AUROC value.

Table 5.48. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using OVO Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP OVO Pairwise Coupling	Accuracy(%)	81.48	79.63	79.63	79.63	85.19	79.63	81.48	79.63	85.19	81.48	81.30
	Kappa	0.769	0.747	0.745	0.743	0.814	0.746	0.768	0.746	0.814	0.767	0.766
	Precision	0.847	0.837	0.821	0.826	0.869	0.836	0.846	0.836	0.869	0.851	0.844
	Recall	0.815	0.796	0.796	0.796	0.852	0.796	0.815	0.796	0.852	0.815	0.813
	Specificity	0.948	0.944	0.938	0.938	0.957	0.946	0.950	0.946	0.957	0.948	0.947
	AUROC	0.958	0.967	0.949	0.967	0.968	0.962	0.964	0.960	0.970	0.960	0.963

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	77	13	0	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	48	0	2	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	78	22
Negotiation	0	0	0	0	64	126

In the third experiment with the MLP algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.49), the MLP algorithm using OVA technique has an average classification accuracy of '82.04%' with lower and upper bounds (80.06% - 84.02%) within 95% CI. In other words, the MLP OVA algorithm predicts the resolution method to be used in construction projects with an average success rate of '82.04%'.

The average for Kappa statistic value is '0.774' that shows a substantial agreement. The weighted average precision value is '0.842'. The weighted average sensitivity (recall) value is '0.820' that means the success of classifiers in identifying true positive instances is '82.0%'. The weighted average specificity value is '0.951' showing the algorithm achieved '95.1%' success in identifying true negative instances. The weighted average AUROC value is '0.961'.

Table 5.49. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using OVA Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP OVA	Accuracy(%)	85.19	81.48	75.93	81.48	83.33	83.33	79.63	81.48	83.33	85.19	82.04
	Kappa	0.813	0.768	0.699	0.768	0.791	0.790	0.744	0.766	0.790	0.813	0.774
	Precision	0.866	0.846	0.794	0.837	0.853	0.854	0.817	0.831	0.854	0.870	0.842
	Recall	0.852	0.815	0.759	0.815	0.833	0.833	0.796	0.815	0.833	0.852	0.820
	Specificity	0.959	0.950	0.934	0.955	0.955	0.957	0.942	0.946	0.957	0.959	0.951
	AUROC	0.970	0.962	0.952	0.964	0.965	0.960	0.951	0.968	0.963	0.959	0.961

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	65	17	8	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	50	0	0	0
Mediation	0	0	0	47	0	3
SEA	0	0	0	0	80	20
Negotiation	0	0	0	2	47	141

In the fourth experiment with the MLP algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.50), the MLP algorithm using RCC technique has an average classification accuracy of '80.74%' with lower and upper bounds (77.87% - 83.62%). In other words, the MLP RCC algorithm predicts the resolution method to be used in construction projects with an average success rate of '80.74%'.

The average for Kappa statistic value is '0.757' that shows a substantial agreement. The weighted average precision value is '0.826'. The weighted average sensitivity (recall) value is '0.807' that means the success of classifiers in identifying true positive instances is '80.7%'. The weighted average specificity value is '0.948' showing the algorithm achieved '94.8%' success in identifying true negative instances. The weighted average AUROC value is '0.917'.

Table 5.50. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using RCC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP RCC	Accuracy(%)	85.19	83.33	74.07	85.19	79.63	81.48	74.07	79.63	81.48	83.33	80.74
	Kappa	0.814	0.791	0.672	0.812	0.744	0.766	0.669	0.743	0.770	0.789	0.757
	Precision	0.866	0.843	0.773	0.862	0.821	0.823	0.748	0.821	0.851	0.851	0.826
	Recall	0.852	0.833	0.741	0.852	0.796	0.815	0.741	0.796	0.815	0.833	0.807
	Specificity	0.961	0.963	0.922	0.959	0.940	0.953	0.925	0.945	0.960	0.955	0.948
	AUROC	0.924	0.911	0.899	0.933	0.909	0.922	0.889	0.928	0.949	0.908	0.917

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	68	15	4	0	2	1
Arbitration	1	53	0	0	5	1
DRB	1	1	46	0	1	1
Mediation	1	0	0	47	0	2
SEA	3	1	0	0	77	19
Negotiation	1	0	7	0	37	145

In the final experiment with the MLP algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.51), the MLP algorithm using ECC technique has an average classification accuracy of '83.33%' with lower and upper bounds (80.54% - 86.13%) within 95% CI. In other words, the MLP ECC algorithm predicts the resolution method to be used in construction projects with an average success rate of '83.33%' that makes it the most successful one among the MLP classifiers for resolution method selection.

Among all MLP classifiers, the highest average for Kappa statistic value (‘0.790’) and the highest weighted average precision value (‘0.857’) is obtained from the MLP ECC algorithm. Similarly, the highest weighted average sensitivity (recall) value (‘0.833’) and the highest specificity value (‘0.953’) are also obtained from this algorithm. In other words, the MLP ECC algorithm achieved ‘83.3%’ accuracy in identifying true positive instances and ‘95.3%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.948’.

Table 5.51. 10-Times 10-Fold Cross-Validation Results of the MLP Algorithm Using ECC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
MLP ECC	Accuracy(%)	83.33	83.33	75.93	85.19	83.33	85.19	77.78	87.04	83.33	88.89	83.33
	Kappa	0.791	0.791	0.697	0.813	0.791	0.813	0.719	0.836	0.791	0.860	0.790
	Precision	0.857	0.857	0.792	0.874	0.857	0.869	0.822	0.883	0.857	0.905	0.857
	Recall	0.833	0.833	0.759	0.852	0.833	0.852	0.778	0.870	0.833	0.889	0.833
	Specificity	0.954	0.954	0.926	0.958	0.954	0.958	0.938	0.963	0.954	0.971	0.953
	AUROC	0.952	0.949	0.938	0.956	0.951	0.951	0.942	0.950	0.954	0.939	0.948

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	70	20	0	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	49	0	1	0
Mediation	0	0	0	47	0	3
SEA	0	0	0	0	81	19
Negotiation	0	0	0	0	47	143

Considering results from MLP experiments, it can be seen that the MLP ECC algorithm outperformed other MLP classifiers in every performance measure except the AUROC value for resolution method selection. Thus, it can be said that the best performing MLP classifier is obtained from the ECC technique that achieved ‘83.33%’ average classification accuracy. In addition, MLP ECC algorithm is superior to kNN algorithms experimented previously for resolution method selection. However, the performance of the MLP ECC algorithm in multiclass classification of resolution method selection is outperformed by the J48 ECC and Naïve Bayes ECC algorithms.

5.3.1.13. The Polynomial Kernel SVM and its Configuration in WEKA

The SVM algorithm cannot naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. Therefore, both multiclass classification problems must be solved by decomposing them into several binary problems using the class 'weka.classifiers.meta.MultiClassClassifier'. Configuration details for the polynomial kernel SVM algorithm using decomposition techniques in WEKA can be seen in Figure 5.23.

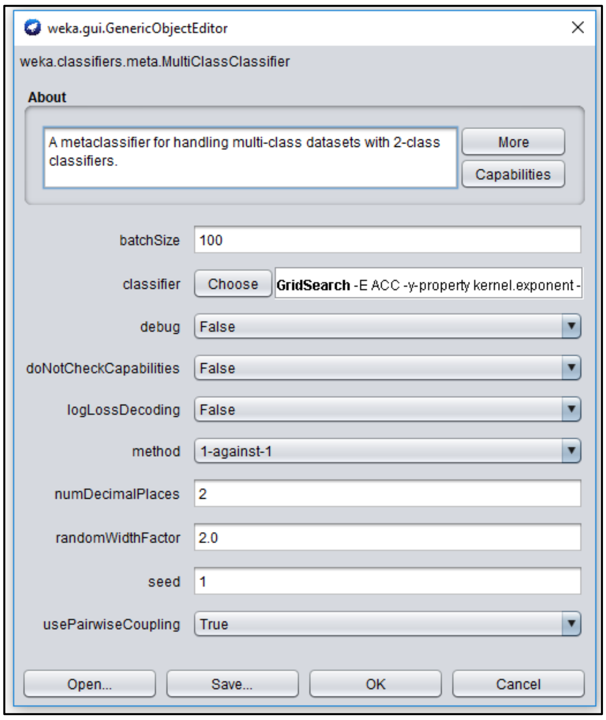


Figure 5.23. The Multiclass Polynomial Kernel SVM Classifier Configuration in WEKA

To select the polynomial kernel SVM algorithm, the 'classifier' setting should be set to SMO with polynomial kernel function using the configuration in binary classification task ('GridSearch') (Figure 5.10 and Figure 5.11). The 'method' setting allows the user to select the decomposition technique. The SMO algorithm is tested using all the available methods, which are OVO, OVA, RCC, and ECC. Default values in WEKA are used for remaining settings.

5.3.1.14. Results from the Polynomial Kernel SVM for Potential Compensation Prediction

As mentioned earlier, the SVM algorithm is not capable of solving multiclass classification problems without decomposing them into binary classification problems. Thus, there will be no natural multiclass solution for the polynomial kernel SVM algorithm. Instead, the OVO technique with pairwise coupling will be utilized in the first experiment with the polynomial kernel SVM. According to 10-fold cross-validation results with 10 repeats (Table 5.52), polynomial kernel SVM classifiers using OVO technique with pairwise coupling have an average classification accuracy of '73.54%' with lower and upper bounds (71.24% - 75.83%) within 95% CI. In other words, the polynomial kernel SVM OVO algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '73.54%'.

Table 5.52. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number 1	2	3	4	5	6	7	8	9	10	Avg.
Poly. Kernel SVM OVO Pairwise Coupling	Accuracy(%)	76.83	73.17	79.27	74.39	74.39	67.07	73.17	73.17	71.95	71.95	73.54
	Kappa	0.623	0.576	0.674	0.584	0.595	0.478	0.569	0.563	0.552	0.556	0.577
	Precision	0.690	0.680	0.762	0.649	0.693	0.614	0.657	0.618	0.657	0.671	0.669
	Recall	0.768	0.732	0.793	0.744	0.744	0.671	0.732	0.732	0.720	0.720	0.736
	Specificity	0.866	0.864	0.902	0.854	0.879	0.843	0.862	0.849	0.855	0.866	0.864
	AUROC	0.865	0.867	0.899	0.853	0.853	0.801	0.845	0.851	0.844	0.823	0.850

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	20	72	0	28
Cost Comp. Only	41	329	6	4
Time Comp. Only	3	2	0	45
Cost & Time Comp.	8	0	8	254

The average for Kappa statistic value is '0.577' that shows a moderate agreement. Among polynomial kernel SVM classifiers, the highest weighted average precision value is obtained from this algorithm as '0.669'. The weighted average sensitivity (recall) value is '0.736' that means the success of classifiers in identifying true positive instances is '73.6%'. The highest weighted average specificity value is also obtained from this algorithm as '0.864' showing the algorithm achieved '86.4%' success in identifying true negative instances. In addition, the weighted average AUROC value is '0.850', which is the highest AUROC value among polynomial kernel SVM classifiers for potential compensation prediction.

In the second experiment with the polynomial kernel SVM algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.53), polynomial kernel SVM classifiers using OVA technique have an average classification accuracy of '72.20%' with lower and upper bounds (70.91% - 73.48%) within 95% CI. In other words, the polynomial kernel SVM OVA classifier predicts the potential compensation type that can be acquired in a dispute with an average success rate of '72.20%'.

Table 5.53. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM OVA	Accuracy(%)	70.73	74.39	70.73	70.73	70.73	70.73	71.95	75.61	73.17	73.17	72.20
	Kappa	0.518	0.581	0.516	0.518	0.518	0.529	0.533	0.601	0.563	0.564	0.544
	Precision	0.597	0.638	0.617	0.623	0.628	0.635	0.641	0.665	0.664	0.671	0.638
	Recall	0.707	0.744	0.707	0.707	0.707	0.707	0.720	0.756	0.732	0.732	0.722
	Specificity	0.823	0.847	0.813	0.819	0.813	0.845	0.812	0.856	0.840	0.845	0.831
	AUROC	0.802	0.826	0.806	0.812	0.817	0.805	0.813	0.829	0.819	0.822	0.815

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	10	84	0	26
Cost Comp. Only	27	343	8	2
Time Comp. Only	0	7	1	42
Cost & Time Comp.	6	22	4	238

The average for Kappa statistic value is '0.544' that shows a moderate agreement. The weighted average precision value is '0.638'. The weighted average sensitivity (recall) value is '0.722' that means the success of classifiers in identifying true positive instances is '72.2%'. The weighted average specificity value is '0.831' showing the algorithm achieved '83.1%' success in identifying true negative instances. The weighted average AUROC value is '0.815'.

In the third experiment with the polynomial kernel SVM algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.54), polynomial kernel SVM classifiers using RCC technique have an average classification accuracy of '72.44%' with lower and upper bounds (71.42% - 73.46%). In other words, the polynomial kernel SVM RCC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '72.44%'.

Table 5.54. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM RCC	Accuracy(%)	73.17	73.17	70.73	74.39	71.95	69.51	73.17	71.95	73.17	73.17	72.44
	Kappa	0.560	0.565	0.529	0.583	0.530	0.507	0.561	0.546	0.565	0.568	0.551
	Precision	0.611	0.626	0.636	0.656	0.593	0.622	0.644	0.616	0.665	0.663	0.633
	Recall	0.732	0.732	0.707	0.744	0.720	0.695	0.732	0.720	0.732	0.732	0.725
	Specificity	0.843	0.855	0.850	0.855	0.815	0.831	0.840	0.850	0.855	0.852	0.845
	AUROC	0.800	0.782	0.767	0.833	0.816	0.803	0.790	0.803	0.813	0.814	0.802

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	7	83	3	27
Cost Comp. Only	25	341	5	9
Time Comp. Only	1	3	0	46
Cost & Time Comp.	8	7	9	246

The average for Kappa statistic value is '0.551' that shows a moderate agreement. The weighted average precision value is '0.633'. The weighted average sensitivity (recall) value is '0.725' that means the success of classifiers in identifying true positive instances is ‘72.5%’. Similarly, the weighted average specificity value is ‘0.845’ showing the algorithm achieved ‘84.5%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.802’.

In the final experiment with the polynomial kernel SVM algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.55), polynomial kernel SVM classifiers using ECC technique have an average classification accuracy of ‘74.39%’ with lower and upper bounds (72.97% - 75.81%) within 95% CI. In other words, the polynomial kernel SVM ECC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘74.39%’ that makes it the best polynomial kernel SVM algorithm in terms of classification accuracy for potential compensation prediction.

Table 5.55. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM ECC	Accuracy(%)	74.39	73.17	74.39	74.39	74.39	70.73	74.39	73.17	78.05	76.83	74.39
	Kappa	0.578	0.571	0.581	0.582	0.580	0.529	0.584	0.569	0.641	0.623	0.584
	Precision	0.614	0.677	0.649	0.655	0.647	0.635	0.662	0.653	0.706	0.700	0.660
	Recall	0.744	0.732	0.744	0.744	0.744	0.707	0.744	0.732	0.780	0.768	0.744
	Specificity	0.845	0.857	0.854	0.850	0.849	0.845	0.855	0.861	0.868	0.862	0.855
	AUROC	0.797	0.837	0.836	0.849	0.809	0.830	0.830	0.816	0.832	0.818	0.825

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	12	81	0	27
Cost Comp. Only	30	340	3	7
Time Comp. Only	1	1	1	47
Cost & Time Comp.	7	2	4	257

Among all experimented polynomial kernel SVM classifiers, the highest average value for Kappa statistic is obtained from this algorithm as ‘0.584’ that shows a moderate agreement. The weighted average precision value is ‘0.660’. The highest weighted average sensitivity (recall) value is also obtained from this algorithm as ‘0.744’ that means the success of classifiers in identifying true positive instances is

‘74.4%’. The weighted average specificity value is ‘0.855’ showing the algorithm achieved ‘85.5%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.825’.

In the light of results from polynomial kernel SVM classifiers for potential compensation prediction, it can be seen that performances are very similar. There is no single polynomial kernel SVM classifier that outperforms others in majority of the performance measures. Therefore, the primary evaluation criterion, which is the average classification accuracy, is used to determine the best polynomial kernel SVM classifier. According to this, it can be said that the best performing polynomial kernel SVM classifier is obtained from using ECC technique. Polynomial kernel SVM ECC algorithm achieved ‘74.39%’ average prediction accuracy for potential compensation prediction. However, this classifier is outperformed by the Naïve Bayes OVA, multiclass kNN, and multiclass J48 algorithms.

5.3.1.15. Results from the Polynomial Kernel SVM for Resolution Method Selection

In the first experiment with the polynomial kernel SVM algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.56), polynomial kernel SVM classifiers using OVO technique with pairwise coupling have an average classification accuracy of ‘78.15%’ with lower and upper bounds (76.10% - 80.20%) within 95% CI. In other words, the polynomial kernel SVM OVO algorithm predicts the resolution method to be used in construction projects with an average success rate of ‘78.15%’.

The average for Kappa statistic value is ‘0.716’ that shows a substantial agreement. The weighted average precision value is ‘0.796’. The weighted average sensitivity (recall) value is ‘0.781’ that means the success of classifiers in identifying true positive instances is ‘78.1%’. Similarly, the weighted average specificity value is ‘0.913’ showing the algorithm achieved ‘91.3%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.912’.

Table 5.56. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using OVO Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM OVO Pairwise Coupling	Accuracy(%)	75.93	79.63	83.33	79.63	74.07	77.78	79.63	74.07	77.78	79.63	78.15
	Kappa	0.684	0.737	0.785	0.738	0.660	0.709	0.734	0.666	0.710	0.736	0.716
	Precision	0.772	0.813	0.838	0.803	0.762	0.787	0.820	0.769	0.777	0.814	0.796
	Recall	0.759	0.796	0.833	0.796	0.741	0.778	0.796	0.741	0.778	0.796	0.781
	Specificity	0.900	0.926	0.933	0.925	0.896	0.903	0.920	0.900	0.903	0.921	0.913
	AUROC	0.903	0.924	0.909	0.918	0.883	0.900	0.939	0.893	0.904	0.943	0.912

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	89	1	0	0	0	0
Arbitration	7	53	0	0	0	0
DRB	8	0	33	0	4	5
Mediation	0	0	0	41	0	9
SEA	0	0	0	0	48	52
Negotiation	0	0	0	0	32	158

In the second experiment with the polynomial kernel SVM algorithm, the OVA technique is utilized. According to the 10-fold cross-validation results with 10 repeats (Table 5.57), polynomial kernel SVM classifiers using OVA technique have an average classification accuracy of ‘75.19%’ with lower and upper bounds (72.92% - 77.45%) within 95% CI. In other words, the polynomial kernel SVM OVA algorithm predicts the resolution method to be used in construction projects with an average success rate of ‘75.19%’.

The average for Kappa statistic value is ‘0.672’ that shows a substantial agreement. The weighted average precision value is ‘0.784’. The weighted average sensitivity (recall) value is ‘0.752’ that means the success of classifiers in identifying true positive instances is ‘75.2%’. Similarly, the weighted average specificity value is ‘0.888’ showing the algorithm achieved ‘88.8%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.896’.

Table 5.57. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using OVA Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM OVA	Accuracy(%)	79.63	77.78	72.22	77.78	75.93	68.52	74.07	75.93	75.93	74.07	75.19
	Kappa	0.734	0.704	0.635	0.706	0.686	0.580	0.656	0.681	0.680	0.656	0.672
	Precision	0.819	0.816	0.749	0.810	0.774	0.728	0.777	0.793	0.797	0.772	0.784
	Recall	0.796	0.778	0.722	0.778	0.759	0.685	0.741	0.759	0.759	0.741	0.752
	Specificity	0.909	0.892	0.879	0.899	0.901	0.861	0.885	0.887	0.889	0.877	0.888
	AUROC	0.906	0.910	0.871	0.919	0.897	0.854	0.884	0.889	0.916	0.916	0.896

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	56	6	0	0	0	28
Arbitration	3	55	0	0	0	2
DRB	6	0	30	0	0	14
Mediation	0	0	0	47	0	3
SEA	0	0	0	0	49	51
Negotiation	0	0	0	0	21	169

In the third experiment with the polynomial kernel SVM algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.58), polynomial kernel SVM classifier using RCC technique have an average classification accuracy of '74.63%' with lower and upper bounds (71.38% - 77.88%). In other words, the polynomial kernel SVM RCC algorithm predicts the resolution method to be used in construction projects with an average success rate of '74.63%'.

The average for Kappa statistic value is '0.681' that shows a substantial agreement. The weighted average precision value is '0.764'. The weighted average sensitivity (recall) value is '0.746' that means the success of classifiers in identifying true positive instances is '74.6%'. Similarly, the weighted average specificity value is '0.933' showing the algorithm achieved '93.3%' success in identifying true negative instances. The weighted average AUROC value is '0.904'.

Table 5.58. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using RCC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM RCC	Accuracy(%)	77.78	77.78	68.52	81.48	74.07	72.22	66.67	75.93	74.07	77.78	74.63
	Kappa	0.717	0.721	0.605	0.766	0.676	0.650	0.580	0.698	0.679	0.716	0.681
	Precision	0.789	0.794	0.709	0.832	0.757	0.751	0.678	0.775	0.773	0.779	0.764
	Recall	0.778	0.778	0.685	0.815	0.741	0.722	0.667	0.759	0.741	0.778	0.746
	Specificity	0.924	0.943	0.918	0.947	0.933	0.929	0.919	0.935	0.952	0.932	0.933
	AUROC	0.894	0.884	0.900	0.922	0.897	0.893	0.874	0.927	0.936	0.908	0.904

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	70	14	3	1	1	1
Arbitration	1	59	0	0	0	0
DRB	5	1	39	2	1	2
Mediation	4	0	3	39	0	4
SEA	5	3	2	3	60	27
Negotiation	2	7	4	4	37	136

In the final experiment with the polynomial kernel SVM algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.59), polynomial kernel SVM classifiers using ECC technique have an average classification accuracy of '82.04%' with lower and upper bounds (79.17% - 84.90%). In other words, the polynomial kernel SVM ECC algorithm predicts the resolution method to be used in construction projects with an average success rate of '82.04%' that makes it the best performing polynomial kernel SVM algorithm for resolution method selection in terms of accuracy.

Among all experimented polynomial kernel SVM algorithms, the highest average for Kappa statistic value ('0.773'), the highest weighted average precision value ('0.839'), the highest weighted average sensitivity (recall) value ('0.820'), the highest weighted average specificity value ('0.944'), and the highest AUROC value ('0.945') is obtained from this algorithm. Thus, the polynomial kernel SVM ECC algorithm achieved '82.0%' success in identifying true positive instances and '94.4%' success in identifying true negative instances.

Table 5.59. 10-Times 10-Fold Cross-Validation Results of the Poly. Kernel SVM Algorithm Using ECC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Poly. Kernel SVM ECC	Accuracy(%)	83.33	77.78	74.07	85.19	85.19	77.78	85.19	83.33	83.33	85.19	82.04
	Kappa	0.790	0.721	0.672	0.813	0.814	0.715	0.814	0.790	0.791	0.812	0.773
	Precision	0.845	0.815	0.769	0.872	0.869	0.786	0.869	0.845	0.858	0.858	0.839
	Recall	0.833	0.778	0.741	0.852	0.852	0.778	0.852	0.833	0.833	0.852	0.820
	Specificity	0.946	0.934	0.916	0.957	0.957	0.919	0.957	0.946	0.952	0.951	0.944
	AUROC	0.950	0.943	0.927	0.951	0.955	0.938	0.946	0.943	0.955	0.944	0.945

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	79	11	0	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	47	0	2	1
Mediation	0	0	0	48	0	2
SEA	0	0	0	0	72	28
Negotiation	0	0	1	0	52	137

Considering results from the polynomial kernel SVM algorithm, it can be said that the best performing polynomial kernel SVM classifier is obtained from the ECC technique that achieved '82.04%' average prediction accuracy for resolution method selection. In addition, the polynomial kernel SVM ECC algorithm is superior to others in all performance measures. However, considering previous algorithms, polynomial kernel SVM is outperformed by the Naïve Bayes ECC, J48 ECC, and MLP ECC algorithms.

5.3.1.16. The RBF Kernel SVM and its Configuration in WEKA

As mentioned earlier, the SVM algorithm cannot naturally solve multiclass classification problems of potential compensation prediction and resolution method selection. Therefore, both multiclass classification problems should be solved by decomposing them into several binary problems using the class 'weka.classifiers.meta.MultiClassClassifier'. Configuration details for the Gaussian RBF kernel SVM algorithm using decomposition techniques in WEKA can be seen in Figure 5.24.

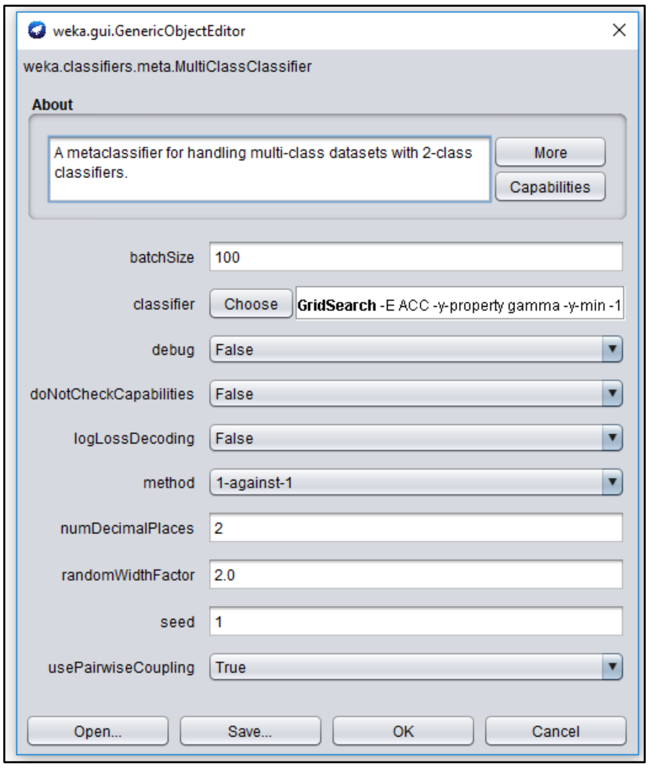


Figure 5.24. The Multiclass Gaussian RBF Kernel SVM Classifier Configuration

To select the Gaussian RBF kernel SVM algorithm, the 'classifier' setting should be set to LibSVM with Gaussian RBF kernel function using the configuration in binary classification task ('GridSearch') (Figure 5.12 and Figure 5.13). The 'method' setting allows the user to select the decomposition technique. The LibSVM algorithm is tested using all available methods, which are OVO, OVA, RCC, and ECC. Default values in WEKA are used for remaining settings.

5.3.1.17. Results from the Gaussian RBF Kernel SVM for Potential Compensation Prediction

In the first experiment with the Gaussian RBF kernel SVM algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.60), Gaussian RBF kernel SVM classifiers using OVO technique with pairwise coupling have an average classification accuracy of '73.05%' with lower and upper bounds (71.28% - 74.82%) within 95% CI. In other words, the

Gaussian RBF kernel SVM OVO classifier predicts the potential compensation type that can be acquired in a dispute with an average success rate of '73.05%'.

Among all experimented RBF kernel SVM algorithms, the highest average for Kappa statistic value ('0.568') and the highest weighted average precision value ('0.659') is obtained from this algorithm. The weighted average sensitivity (recall) value is '0.731' that means the success of classifiers in identifying true positive instances is '73.1%'. The highest weighted average specificity value is also obtained from this algorithm as '0.859' showing the algorithm achieved '85.9%' success in identifying true negative instances. The weighted average AUROC value is '0.849', which is the highest among RBF kernel SVM algorithms experimented for potential compensation prediction.

Table 5.60. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using OVO Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM OVO Pairwise Coupling	Accuracy(%)	74.39	73.17	78.05	71.95	73.17	68.29	71.95	74.39	71.95	73.17	73.05
	Kappa	0.584	0.572	0.653	0.550	0.571	0.496	0.550	0.582	0.555	0.566	0.568
	Precision	0.649	0.662	0.748	0.636	0.662	0.624	0.647	0.649	0.659	0.654	0.659
	Recall	0.744	0.732	0.780	0.720	0.732	0.683	0.720	0.744	0.720	0.732	0.731
	Specificity	0.854	0.860	0.896	0.849	0.860	0.844	0.856	0.854	0.859	0.854	0.859
	AUROC	0.848	0.858	0.892	0.825	0.851	0.807	0.852	0.869	0.847	0.841	0.849

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	17	73	0	30
Cost Comp. Only	45	325	1	9
Time Comp. Only	0	0	0	50
Cost & Time Comp.	6	0	7	257

In the second experiment with the Gaussian RBF kernel SVM algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.61), Gaussian RBF kernel SVM classifiers using OVA technique have an average classification accuracy of '71.55%' with lower and upper bounds (69.63% - 73.47%) within 95% CI. In other words, the Gaussian RBF kernel SVM OVA algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘71.55%’.

The average for Kappa statistic value is ‘0.527’ that shows a moderate agreement. The weighted average precision value is ‘0.610’. The weighted average sensitivity (recall) value is ‘0.715’ that means the success of classifiers in identifying true positive instances is ‘71.5%’. Similarly, the weighted average specificity value is ‘0.817’ showing the algorithm achieved ‘81.7%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.813’.

Table 5.61. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using OVA Technique for Potential Compensation Prediction

Classifier	**Performance Measure**	**Run Number**										**Avg.**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
RBF Kernel SVM OVA	Accuracy(%)	71.95	74.39	71.95	69.51	70.73	68.29	71.95	76.83	67.97	71.95	71.55
	Kappa	0.543	0.579	0.526	0.497	0.509	0.480	0.532	0.609	0.454	0.543	0.527
	Precision	0.636	0.640	0.591	0.588	0.586	0.585	0.629	0.617	0.587	0.644	0.610
	Recall	0.720	0.744	0.720	0.695	0.707	0.683	0.720	0.768	0.671	0.720	0.715
	Specificity	0.838	0.842	0.799	0.812	0.800	0.810	0.810	0.840	0.786	0.834	0.817
	AUROC	0.795	0.815	0.838	0.805	0.810	0.804	0.811	0.827	0.800	0.822	0.813

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	3	91	0	26
Cost Comp. Only	25	347	2	6
Time Comp. Only	0	13	2	35
Cost & Time Comp.	6	25	5	234

In the third experiment with the Gaussian RBF kernel SVM algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.62), Gaussian RBF kernel SVM classifiers using RCC technique have an average classification accuracy of ‘72.44%’ with lower and upper bounds (71.26% - 73.62%). In other words, the Gaussian RBF kernel SVM RCC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of ‘72.44%’.

The average for Kappa statistic value is '0.552' that shows a moderate agreement. The weighted average precision value is '0.641'. The weighted average sensitivity (recall) value is '0.725' that means the success of classifiers in identifying true positive instances is '72.5%'. Similarly, the weighted average specificity value is '0.847' showing the algorithm achieved '84.7%' success in identifying true negative instances. The weighted average AUROC value is '0.805'.

Table 5.62. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using RCC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM RCC	Accuracy(%)	73.17	74.39	70.73	73.17	69.51	71.95	73.17	70.73	73.17	74.39	72.44
	Kappa	0.559	0.579	0.532	0.566	0.499	0.541	0.563	0.530	0.570	0.584	0.552
	Precision	0.610	0.620	0.642	0.649	0.596	0.627	0.652	0.615	0.696	0.706	0.641
	Recall	0.732	0.744	0.707	0.732	0.695	0.720	0.732	0.707	0.732	0.744	0.725
	Specificity	0.843	0.847	0.858	0.857	0.816	0.832	0.845	0.854	0.865	0.849	0.847
	AUROC	0.808	0.791	0.766	0.826	0.817	0.797	0.796	0.814	0.820	0.816	0.805

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	7	81	3	29
Cost Comp. Only	25	342	4	9
Time Comp. Only	1	5	1	43
Cost & Time Comp.	11	5	10	244

In the final experiment with the Gaussian RBF kernel SVM algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.63), Gaussian RBF kernel SVM classifiers using ECC technique have an average classification accuracy of '73.41%' with lower and upper bounds (72.34% - 74.49%) within 95% CI. In other words, the Gaussian RBF kernel SVM ECC algorithm predicts the potential compensation type that can be acquired in a dispute with an average success rate of '73.41%' that makes it the best one among experimented Gaussian RBF kernel algorithms in terms of accuracy for potential compensation prediction.

The average Kappa statistic value is '0.564' that shows a moderate agreement. The weighted average precision value is '0.626'. Among all experimented RBF kernel SVM algorithms, the highest weighted average sensitivity (recall) value is obtained from this algorithm as '0.733' that means the success of classifiers in identifying true positive instances is '73.3%'. The weighted average specificity value is '0.845' showing the algorithm achieved '84.5%' success in identifying true negative instances. The weighted average AUROC value is '0.827'.

Table 5.63. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using ECC Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM ECC	Accuracy(%)	73.17	74.39	73.17	73.17	70.73	71.95	75.61	73.17	75.61	73.17	73.41
	Kappa	0.560	0.578	0.560	0.564	0.520	0.543	0.599	0.565	0.600	0.557	0.564
	Precision	0.611	0.614	0.611	0.645	0.598	0.609	0.657	0.642	0.659	0.610	0.626
	Recall	0.732	0.733	0.732	0.732	0.707	0.720	0.756	0.732	0.756	0.732	0.733
	Specificity	0.843	0.845	0.843	0.848	0.830	0.841	0.851	0.851	0.856	0.838	0.845
	AUROC	0.813	0.850	0.843	0.817	0.816	0.822	0.827	0.834	0.823	0.829	0.827

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	4	87	0	29
Cost Comp. Only	30	341	0	9
Time Comp. Only	1	0	0	49
Cost & Time Comp.	9	3	1	257

Considering that performances of Gaussian RBF kernel SVM algorithms are close to each other, the primary evaluation criterion (classification accuracy) will be used for determining the best RBF kernel SVM classifier for potential compensation prediction. Thus, it can be said that the best performing Gaussian RBF kernel SVM classifier is obtained from the ECC technique that achieved '73.41%' average prediction accuracy. However, considering previous algorithms, Gaussian RBF kernel SVM is outperformed by the Naïve Bayes OVA, multiclass kNN, multiclass J48, and polynomial kernel SVM ECC algorithms.

5.3.1.18. Results from the Gaussian RBF Kernel SVM for Resolution Method Selection

In the first experiment with the Gaussian RBF kernel SVM algorithm, the OVO technique with pairwise coupling is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.64), Gaussian RBF kernel SVM classifiers using OVO technique with pairwise coupling have an average classification accuracy of '78.33%' with lower and upper bounds (74.43% - 82.24%) within 95% CI. In other words, the Gaussian RBF kernel SVM OVO algorithm predicts the resolution method to be used in construction projects with an average success rate of '78.33%'.

The average for Kappa statistic value is '0.720' that shows a substantial agreement. The weighted average precision value is '0.798'. The weighted average sensivity (recall) value is '0.783' that means the success of classifiers in identifying true positive instances is '78.3%'. Similarly, the weighted average specificity value is '0.917' showing the algorithm achieved '91.7%' success in identifying true negative instances. The weighted average AUROC value is '0.918'.

Table 5.64. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using OVO Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM OVO Pairwise Coupling	Accuracy(%)	77.78	81.48	74.07	83.33	66.67	79.63	85.19	74.07	81.48	79.63	78.33
	Kappa	0.711	0.761	0.674	0.785	0.561	0.733	0.810	0.663	0.760	0.739	0.720
	Precision	0.785	0.817	0.790	0.834	0.701	0.813	0.866	0.746	0.822	0.804	0.798
	Recall	0.778	0.815	0.741	0.833	0.667	0.796	0.852	0.741	0.815	0.796	0.783
	Specificity	0.903	0.929	0.925	0.933	0.874	0.907	0.949	0.896	0.923	0.926	0.917
	AUROC	0.910	0.925	0.901	0.933	0.878	0.933	0.955	0.899	0.926	0.919	0.918

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	88	2	0	0	0	0
Arbitration	7	53	0	0	0	0
DRB	1	0	37	0	8	4
Mediation	0	0	0	44	0	6
SEA	0	0	0	0	50	50
Negotiation	0	0	0	0	39	151

In the second experiment with the Gaussian RBF kernel SVM algorithm, the OVA technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.65), Gaussian RBF kernel SVM classifiers using OVA technique have an average classification accuracy of '74.07%' with lower and upper bounds (71.66% - 76.49%) within 95% CI. In other words, the Gaussian RBF kernel SVM OVA algorithm predicts the resolution method to be used in construction projects with an average success rate of '74.07%'.

The average for Kappa statistic value is '0.659' that shows a substantial agreement. The weighted average precision value is '0.784'. The weighted average sensitivity (recall) value is '0.741' that means the success of classifiers in identifying true positive instances is '74.1%'. The weighted average specificity value is '0.884' showing the algorithm achieved '88.4%' success in identifying true negative instances. The weighted average AUROC value is '0.878'.

Table 5.65. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using OVA Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM OVA	Accuracy(%)	72.22	72.22	72.22	75.93	74.07	66.67	77.78	75.93	77.78	75.93	74.07
	Kappa	0.636	0.627	0.634	0.680	0.662	0.560	0.710	0.687	0.705	0.687	0.659
	Precision	0.761	0.786	0.769	0.809	0.770	0.719	0.811	0.794	0.828	0.794	0.784
	Recall	0.722	0.722	0.722	0.759	0.741	0.667	0.778	0.759	0.778	0.759	0.741
	Specificity	0.880	0.861	0.873	0.881	0.890	0.856	0.905	0.900	0.893	0.900	0.884
	AUROC	0.866	0.870	0.870	0.892	0.884	0.851	0.883	0.880	0.908	0.880	0.878

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	50	7	0	0	0	33
Arbitration	0	60	0	0	0	0
DRB	0	0	34	0	0	16
Mediation	0	0	0	45	0	5
SEA	0	0	0	0	54	46
Negotiation	0	0	0	0	33	157

In the third experiment with the Gaussian RBF kernel SVM algorithm, the RCC technique is utilized. According to 10-fold cross-validation results with 10 repeats

(Table 5.66), Gaussian RBF kernel SVM classifiers using RCC technique have an average classification accuracy of '75.37%' with lower and upper bounds (77.54% - 73.20%) within 95% CI. In other words, the Gaussian RBF kernel SVM RCC algorithm predicts the resolution method to be used in construction projects with an average success rate of '75.37%'.

The average for Kappa statistic value is '0.690' that shows a substantial agreement. The weighted average precision value is '0.767'. The weighted average sensitivity (recall) value is '0.754' that means the success of classifiers in identifying true positive instances is '75.4%'. Similarly, the weighted average specificity value is '0.933' showing the algorithm achieved '93.3%' success in identifying true negative instances. The weighted average AUROC value is '0.897'.

Table 5.66. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using RCC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM RCC	Accuracy(%)	79.63	77.78	72.22	75.93	77.78	70.37	72.22	77.78	74.07	75.93	75.37
	Kappa	0.744	0.720	0.650	0.694	0.720	0.633	0.646	0.721	0.675	0.694	0.690
	Precision	0.820	0.782	0.736	0.767	0.796	0.725	0.728	0.801	0.763	0.753	0.767
	Recall	0.796	0.778	0.722	0.759	0.778	0.704	0.722	0.778	0.741	0.759	0.754
	Specificity	0.940	0.937	0.924	0.930	0.935	0.931	0.927	0.940	0.940	0.930	0.933
	AUROC	0.896	0.878	0.900	0.883	0.900	0.907	0.868	0.901	0.927	0.905	0.897

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	69	16	3	1	1	0
Arbitration	0	60	0	0	0	0
DRB	4	1	41	0	3	1
Mediation	3	0	2	39	0	6
SEA	1	4	2	4	62	27
Negotiation	5	2	4	3	40	136

In the final experiment with the Gaussian RBF kernel SVM algorithm, the ECC technique is utilized. According to 10-fold cross-validation results with 10 repeats (Table 5.67), Gaussian RBF kernel SVM classifiers using ECC technique have an average classification accuracy of '80.93%' with lower and upper bounds (79.84% -

82.02%). In other words, the Gaussian RBF kernel SVM ECC algorithm predicts the resolution method to be used in construction projects with an average success rate of ‘80.93%’ that makes it the best one among Gaussian RBF kernel classifiers.

Among all experimented Gaussian RBF kernel SVM algorithms, the highest average Kappa statistic value (‘0.760’), the highest weighted average precision value (‘0.833’), the highest weighted average sensitivity (recall) value (‘0.809’), the highest weighted average specificity value (‘0.943’), and the highest weighted average AUROC value (‘0.944’) is obtained from this algorithm.

Table 5.67. 10-Times 10-Fold Cross-Validation Results of the RBF Kernel SVM Algorithm Using ECC Techniques for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
RBF Kernel SVM ECC	Accuracy(%)	81.48	81.48	79.63	77.78	81.48	81.48	81.48	79.63	83.33	81.48	80.93
	Kappa	0.768	0.767	0.742	0.719	0.767	0.766	0.765	0.743	0.791	0.767	0.760
	Precision	0.826	0.851	0.816	0.794	0.851	0.833	0.843	0.826	0.857	0.836	0.833
	Recall	0.815	0.815	0.796	0.778	0.815	0.815	0.815	0.796	0.833	0.815	0.809
	Specificity	0.945	0.948	0.932	0.932	0.948	0.944	0.942	0.938	0.954	0.949	0.943
	AUROC	0.947	0.940	0.941	0.948	0.950	0.938	0.951	0.944	0.946	0.938	0.944

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	76	13	1	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	45	0	5	0
Mediation	0	0	0	48	0	2
SEA	1	0	0	0	73	26
Negotiation	0	0	1	0	54	135

Considering results from the Gaussian RBF kernel SVM algorithm, it can be said that the best performing Gaussian RBF kernel SVM classifier is obtained from the ECC technique that achieved ‘80.93%’ average prediction accuracy for resolution method selection. In addition, the Gaussian RBF kernel SVM ECC algorithm is superior to remaining Gaussian RBF kernel SVM classifiers in all other performance measures. However, it is outperformed by the Naïve Bayes ECC, J48 ECC, MLP ECC, and polynomial kernel SVM ECC algorithms.

5.3.2. Comparison of Results from Single Classifiers for Potential Compensation Prediction

Table 5.68 shows the 10-times 10-fold cross-validation results of single classifiers with their best parameter settings. The best version of each algorithm is considered (results from the best decomposition technique). This table is used for comparing performances of single classifiers with each other.

Table 5.68. Comparison of Single Classifiers for Potential Compensation Prediction

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Weigh. Avg. Precision	Weigh Avg. Recall (TPR)	Weight. Avg. Specificity	Weigh. Avg. AUROC	Rank
Naïve Bayes OVA	**80.61**	**[80.11 – 81.10]**	**0.691**	**0.774**	**0.806**	**0.899**	**0.916**	**1**
KNN Multiclass	78.66	[77.05 – 80.26]	0.661	0.737	0.787	0.893	0.912	**2**
J48 Multiclass	76.95	[75.99 – 77.91]	0.616	0.632	0.769	0.852	0.811	**3**
MLP OVA	68.63	[67.03 – 70.23]	0.512	0.666	0.686	0.864	0.842	6
Poly. Kernel SVM ECC	74.39	[72.97 – 75.81]	0.584	0.660	0.744	0.855	0.825	4
RBF Kernel SVM ECC	73.41	[72.34 – 74.49]	0.564	0.626	0.733	0.845	0.827	5

The best average classification accuracy is obtained from the Naïve Bayes algorithm using OVA technique that achieved '80.61%' average classification accuracy. It is followed by the multiclass kNN (no decomposition, natural solution) algorithm that achieved '78.66%' average classification accuracy. The third place belongs to multiclass J48 classifiers with '76.95%' average classification accuracy. The average classification accuracy of single classifiers within 95% CI can be seen in Figure 5.25.

Besides the average accuracy measure, the Naïve Bayes OVA algorithm generated the best results in all performance measures.

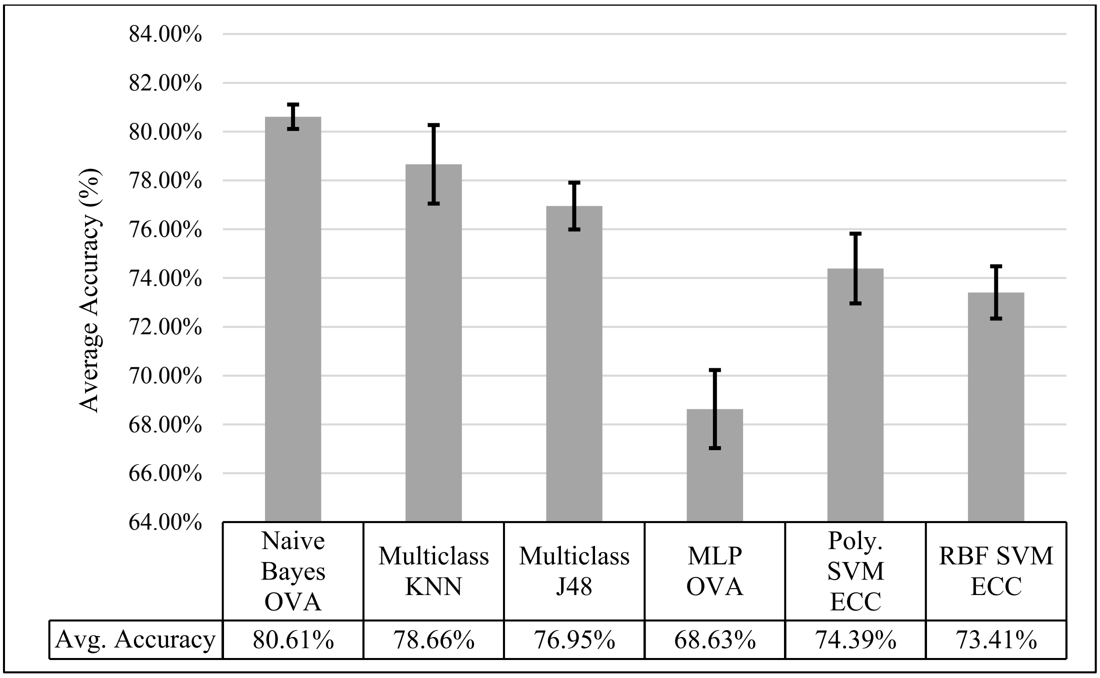


Figure 5.25. Average Classification Accuracies of Single Classifiers within 95% CI for Potential Compensation Prediction

In the light of these comparisons, it is observed that the Naïve Bayes OVA algorithm is superior to others in terms of average classification accuracy, average Kappa, weighted average precision, weighted average true positive rate (recall), weighted average specificity, and weighted average AUROC measures.

5.3.3. Comparison of Results from Single Classifiers for Resolution Method Selection

Table 5.69 shows the 10-times 10-fold cross-validation results of single classifiers with their best parameter settings. The best version of each algorithm is considered (results from the best decomposition technique). This table is used for comparing performances of single classifiers with each other.

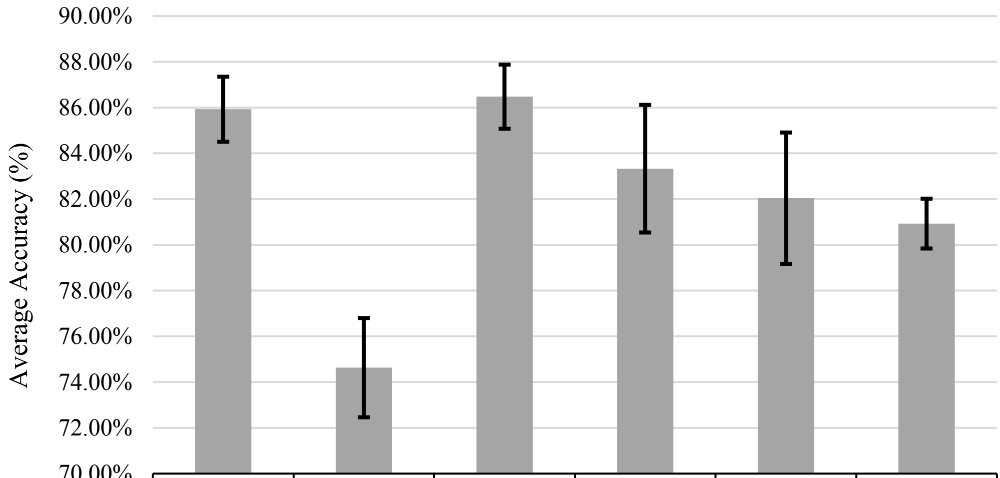
Table 5.69. Comparison of Single Classifiers for Resolution Method Selection

Algorithm	**Avg. Accuracy (%)**	**%95 CI Accuracy (%)**	**Avg. Kappa**	**Weigh. Avg. Precision**	**Weigh Avg. Recall (TPR)**	**Weigh. Avg. Specificity**	**Weigh. Avg. AUROC**	**Rank**
Naïve Bayes ECC	85.93	[84.50 – 87.35]	0.817	0.875	0.859	0.935	0.961	**2**
KNN ECC	74.63	[72.46 – 76.80]	0.674	0.769	0.746	0.910	0.908	6
J48 ECC	**86.48**	**[85.08 – 87.88]**	**0.830**	**0.879**	**0.865**	**0.967**	**0.964**	**1**
MLP ECC	83.33	[80.54 – 86.13]	0.790	0.857	0.833	0.953	0.948	**3**
Poly. Kernel SVM ECC	82.04	[79.17 – 84.90]	0.773	0.839	0.820	0.944	0.945	4
RBF Kernel SVM ECC	80.93	[79.84 – 82.02]	0.760	0.833	0.809	0.943	0.944	5

The best average classification accuracy is obtained from the J48 algorithm using ECC technique that achieved ‘86.48%’ average classification accuracy. It is followed by the Naïve Bayes algorithm using ECC technique that achieved ‘85.93%’ average classification accuracy. The third place belongs to MLP classifiers using ECC technique with ‘83.33%’ average classification accuracy. The average classification accuracy of single classifiers within 95% CI can be seen in Figure 5.26.

Besides the average accuracy measure, the J48 ECC algorithm generated the best results in all performance measures.

In the light of these comparisons, it is observed that the Naïve Bayes OVA algorithm is superior to others in terms of average classification accuracy, average Kappa, weighted average precision, weighted average true positive rate (recall), weighted average specificity, and weighted average AUROC measures.



	Naive Bayes ECC	KNN ECC	J48 ECC	MLP ECC	Poly. SVM ECC	RBF SVM ECC
Avg. Accuracy	85.93%	74.63%	86.48%	83.33%	82.04%	80.93%

Figure 5.26. Average Classification Accuracies of Single Classifiers within 95% CI for Resolution Method Selection

5.3.4. Multiclass Classification Using Ensemble ML Algorithms

WEKA configuration details of each ensemble ML algorithm and obtained multiclass classification results are given in this section starting with the voting technique, which will be followed by the stacked generalization technique and the AdaBoost algorithm, in the given order.

5.3.4.1. The Voting Technique and its Configuration in WEKA

In this research, results of the top three base classifiers in terms of multiclass classification accuracy are considered during voting. For potential compensation prediction, these algorithms are (1) Naïve Bayes using OVA technique, (2) multiclass kNN, and (3) multiclass J48 algorithms. For resolution method selection, these algorithms are (1) J48 using ECC technique, (2) Naïve Bayes using ECC technique, and (3) MLP algorithm using ECC technique.

In order to define the mentioned algorithms, firstly, the voting technique is selected from the class 'weka.classifiers.meta.Vote' class; secondly, all three algorithms are selected one by one using the class 'weka.classifiers.meta.MultiClassClassifier' with their original configurations in their single versions. Figure 5.27 shows configuration details of the voting technique in WEKA workbench.

Among various voting strategies in the literature, the majority voting and the average of probabilities techniques are experimented in this research. The voting strategy can be selected by adjusting the 'combinationRule' setting. Default values of WEKA are used for remaining settings.

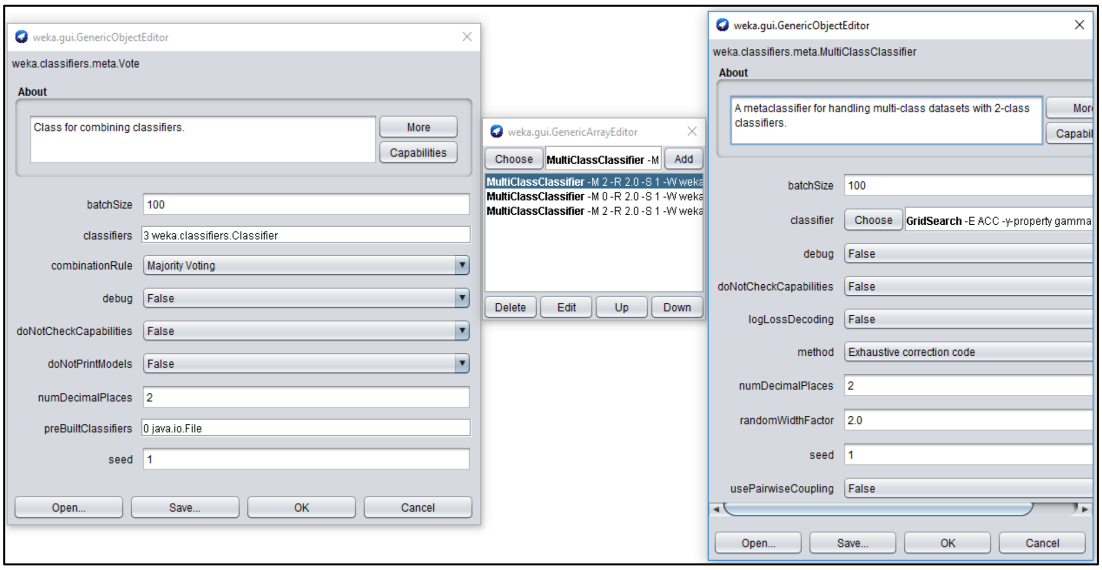


Figure 5.27. The Voting Technique Configuration in WEKA for Multiclass Classification Problems

5.3.4.2. Results from Voting Technique for Potential Compensation Prediction

10-fold cross-validation results with 10 repeats obtained from the majority voting technique are given in Table 5.70. Ensemble classifiers obtained from majority voting have an average classification accuracy of '80.61%' with lower and upper bounds (79.57% - 81.65%) within 95% CI. In other words, ensemble classifiers predict the potential compensation type that can be acquired in a dispute with an average success rate of '80.61%'.

The average for Kappa statistic value is '0.688' that shows a substantial agreement. The weighted average precision value is '0.755'. The weighted average sensitivity (recall) value is '0.806' that means the algorithm achieved '80.6%' success in identifying true positive instances. Similarly, the weighted average specificity value is '0.894' showing the algorithm achieved '89.4%' success in identifying true negative instances. The weighted average AUROC value is '0.850'.

Table 5.70. 10-Times 10-Fold Cross-Validation Results of Majority Voting Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Majority Voting	Accuracy(%)	81.71	80.49	80.49	79.27	78.05	82.93	81.71	79.27	81.71	80.49	80.61
	Kappa	0.706	0.685	0.688	0.667	0.645	0.724	0.706	0.667	0.705	0.689	0.688
	Precision	0.767	0.750	0.765	0.735	0.714	0.786	0.767	0.748	0.764	0.753	0.755
	Recall	0.817	0.805	0.805	0.793	0.780	0.829	0.817	0.793	0.817	0.805	0.806
	Specificity	0.900	0.889	0.899	0.887	0.877	0.902	0.900	0.888	0.895	0.898	0.894
	AUROC	0.858	0.847	0.852	0.840	0.829	0.866	0.858	0.841	0.856	0.851	0.850

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	45	45	0	30
Cost Comp. Only	22	348	0	10
Time Comp. Only	0	1	0	49
Cost & Time Comp.	0	0	2	268

10-fold cross-validation results with 10 repeats obtained from average of probabilities voting technique are given in Table 5.71. Ensemble classifiers obtained from the average of probabilities voting technique have an average classification accuracy of '77.07%' with lower and upper bounds (75.72% - 78.42%) within 95% CI. In other words, ensemble classifiers predict the potential compensation type that can be acquired in a dispute with an average success rate of '77.07%'.

The average for Kappa statistic value is '0.625' that shows a substantial agreement. The weighted average precision value is '0.624'. The weighted average sensitivity (recall) value is '0.771' that means the algorithm achieved '77.1%' success in identifying true positive instances. Similarly, the weighted average specificity value is '0.864' showing the algorithm achieved '86.4%' success in identifying true negative instances. The weighted average AUROC value is '0.915'.

Table 5.71. 10-Times 10-Fold Cross-Validation Results of Average of Probabilities Voting Technique for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Average of Prob. Voting	Accuracy(%)	76.83	78.05	79.27	75.61	73.17	78.05	79.27	78.05	76.83	75.61	77.07
	Kappa	0.618	0.641	0.663	0.601	0.561	0.641	0.659	0.645	0.618	0.605	0.625
	Precision	0.673	0.706	0.752	0.659	0.611	0.706	0.732	0.728	0.673	0.687	0.693
	Recall	0.768	0.780	0.793	0.756	0.732	0.780	0.793	0.780	0.768	0.756	0.771
	Specificity	0.858	0.868	0.880	0.856	0.843	0.868	0.870	0.878	0.858	0.865	0.864
	AUROC	0.922	0.923	0.913	0.915	0.903	0.908	0.922	0.915	0.913	0.917	0.915

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	17	73	0	30
Cost Comp. Only	22	348	0	10
Time Comp. Only	0	0	0	50
Cost & Time Comp.	0	0	3	267

In the light of these results, it can be seen that majority voting technique performed better than the average of probabilities voting technique for potential compensation prediction. However, the average classification accuracy obtained from majority voting technique is same as the best single classifier, which is the Naïve Bayes OVA classifier. In addition, the single Naïve Bayes OVA algorithm has better performance in remaining measures compared to the ensemble classifier obtained from the majority voting technique.

5.3.4.3. Results from the Voting Technique for Resolution Method Selection

10-fold cross-validation results with 10 repeats obtained from the majority voting technique are given in Table 5.72. Ensemble classifiers obtained from majority voting have an average classification accuracy of '89.44%' with lower and upper bounds (87.37% - 91.52%) within 95% CI. In other words, ensemble classifiers predict the resolution method to be used in construction projects with an average success rate of ‘89.44%’.

The average for Kappa statistic value is ‘0.866’ that shows a perfect agreement. The weighted average precision value is ‘0.900’. The weighted average sensitivity (recall) value is ‘0.894’ that means the success of the algorithm in identifying true positive instances is ‘89.4%’. Similarly, the weighted average specificity value is ‘0.965’ showing the algorithm achieved ‘96.5%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.930’.

Table 5.72. 10-Times 10-Fold Cross-Validation Results of Majority Voting Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Majority Voting	Accuracy(%)	90.74	88.89	83.33	88.89	90.74	90.74	87.04	90.74	88.89	94.44	89.44
	Kappa	0.882	0.859	0.788	0.859	0.882	0.882	0.834	0.882	0.859	0.930	0.866
	Precision	0.910	0.895	0.845	0.899	0.910	0.910	0.879	0.910	0.895	0.947	0.900
	Recall	0.907	0.889	0.833	0.889	0.907	0.907	0.870	0.907	0.889	0.944	0.894
	Specificity	0.969	0.965	0.939	0.965	0.969	0.969	0.957	0.970	0.965	0.982	0.965
	AUROC	0.938	0.927	0.886	0.927	0.938	0.938	0.914	0.938	0.927	0.963	0.930

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	81	6	3	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	50	0	0	0
Mediation	0	0	0	47	0	3
SEA	0	0	0	0	81	19
Negotiation	0	0	0	0	26	164

10-fold cross-validation results with 10 repeats obtained from average of probabilities voting technique are given in Table 5.73. Ensemble classifiers obtained from average of probabilities voting have an average classification accuracy of ‘88.33%’ with lower and upper bounds (87.24% - 89.42%) within 95% CI. In other words, ensemble classifiers predict the resolution method to be used in construction projects with an average success rate of ‘88.33%’.

The average for Kappa statistic value is '0.853' that shows a perfect agreement. The weighted average precision value is '0.899'. The weighted average sensitivity (recall) value is '0.883' that means the success of the algorithm in identifying true positive instances is '88.3%'. Similarly, the weighted average specificity value is '0.970' showing the algorithm achieved '97.0%' success in identifying true negative instances. The weighted average AUROC value is '0.969'.

Table 5.73. 10-Times 10-Fold Cross-Validation Results of Average of Probabilities Voting Technique for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Average of Prob. Voting	Accuracy(%)	88.89	87.04	85.19	88.89	88.89	90.74	87.04	88.89	88.89	88.89	88.33
	Kappa	0.860	0.837	0.814	0.859	0.860	0.883	0.837	0.860	0.860	0.860	0.853
	Precision	0.905	0.893	0.869	0.898	0.905	0.917	0.890	0.902	0.902	0.905	0.899
	Recall	0.889	0.870	0.852	0.889	0.889	0.907	0.870	0.889	0.889	0.889	0.883
	Specificity	0.973	0.968	0.957	0.967	0.973	0.975	0.971	0.973	0.973	0.971	0.970
	AUROC	0.971	0.965	0.964	0.968	0.970	0.972	0.969	0.970	0.973	0.971	0.969

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	72	15	3	0	0	0
Arbitration	0	60	0	0	0	0
DRB	0	0	50	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	88	12
Negotiation	0	0	0	0	33	157

In the light of these results, it can be observed that the majority voting technique performed better than the average of probabilities voting technique for resolution method selection. In addition, the average classification accuracy obtained from majority voting technique outperformed the single ML algorithms it contains. The ensemble classifier improved the average classification accuracy of MLP ECC classifiers by '+6.11%', Naïve Bayes ECC classifiers by '+3.51%', and J48 ECC classifiers by '+2.96%'. Thus, majority voting technique contributed to overall performance significantly.

5.3.4.4. The Stacked Generalization and its Configuration in WEKA

In this thesis study, the top three single classifiers are combined with remaining experimented ML algorithms during stacking. For potential compensation prediction, the top three algorithms are (1) Naïve Bayes algorithm using OVO technique, (2) multiclass kNN algorithm, and (3) multiclass J48 algorithm. In stacking, same algorithms should not be stacked together. Therefore, each algorithm is combined with the remaining five algorithms so that ‘15’ ensemble classifiers are obtained for potential compensation prediction. For resolution method selection, the top three algorithms are (1) J48 algorithm using ECC technique, (2) Naïve Bayes algorithm using ECC technique, and (3) MLP algorithm using ECC technique. Similar to the potential compensation prediction, there will be ‘15’ stacked ensemble classifiers obtained for resolution method selection.

In order to apply stacked generalization to multiclass classification problems of potential compensation prediction and resolution method selection, firstly, the stacked generalization class in Weka is selected as ‘weka.classifiers.meta.Stacking’. In stacked generalization, the primary algorithm, which is the base-learner, is defined in the ‘classifiers’ setting by selecting the relevant algorithm. The secondary algorithm, which is the meta-learner, is defined in the ‘metaClassifier’ setting by selecting the relevant algorithm. Figure 5.28 shows an example configuration of the ensemble algorithm obtained by combining the kNN algorithm as base-learner and the Naïve Bayes algorithm as meta-learner. While defining the kNN algorithm as the base-learner, the ECC multiclass decomposition technique is used by selecting the ‘weka.classifiers.meta.MultiClassClassifier’ class. Similarly, while defining the Naïve Bayes algorithm as the meta-learner, the ECC multiclass decomposition is used.

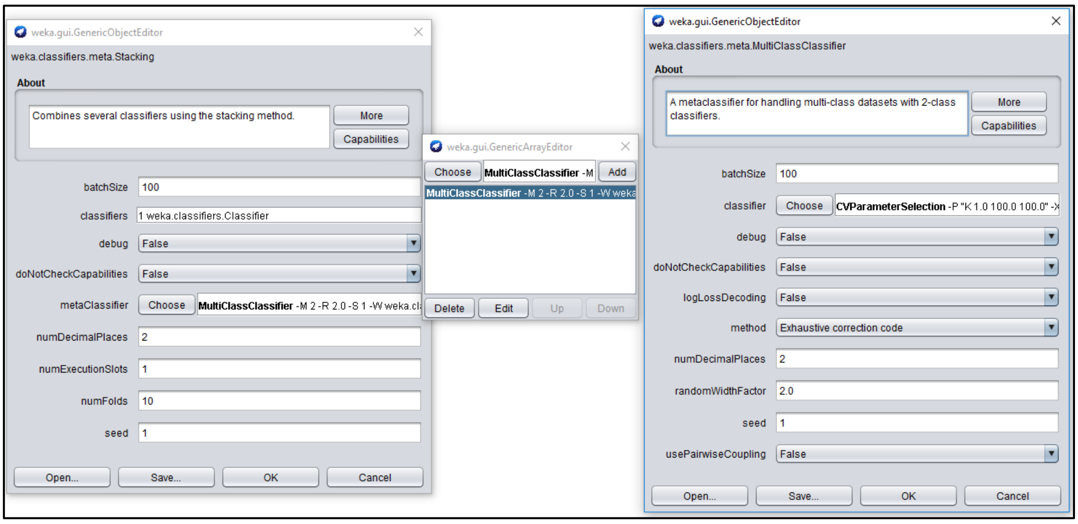


Figure 5.28. The Stacked Generalization Configuration in WEKA for Multiclass Classification Problems

5.3.4.5. Results from Stacking Technique for Potential Compensation Prediction

It was experienced in the binary data classification problem that not all stacked ensemble classifiers achieved better classification accuracies than the single ones they contain. A similar situation also exist for the multiclass classification problem of potential compensation prediction. For this purpose, the classification accuracy of the ensemble model is compared with accuracies of both of the single classifiers they contain.

When the base-learner is the Naïve Bayes algorithm using OVA technique (best performing single Naïve Bayes algorithm for potential compensation prediction), none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models ‘Naïve Bayes OVA + multiclass kNN’, ‘Naïve Bayes OVA + multiclass J48’, ‘Naïve Bayes OVA + MLP OVA’, ‘Naïve Bayes OVA + Polynomial kernel SVM ECC’, and ‘Naïve Bayes OVA + Gaussian RBF kernel SVM ECC’ are not considered.

Similarly, when the base learner is the multiclass kNN algorithm (best performing single kNN algorithm for potential compensation prediction), none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models containing multiclass kNN algorithm as base-learner are not considered.

When the base-learner is multiclass J48 algorithm (best performing single J48 algorithm for potential compensation prediction), two of the ensemble models achieved better accuracies than single algorithms they contain. These are ‘multiclass J48 + MLP OVA’ and ‘multiclass J48 + Gaussian RBF kernel SVM ECC’.

10-fold cross-validation results with 10 repeats obtained by combining the multiclass J48 and MLP OVA algorithms are given in Table 5.74. These ensemble classifiers have an average classification accuracy of ‘77.56%’ with lower and upper bounds (76.95% - 78.17%) within 95% CI. In other words, ensemble classifiers predict the potential compensation type that can be acquired in a dispute with an average success rate of ‘77.56%’. Thus, the stacked classifier enhanced the average classification accuracy of the base-learner (multiclass J48) by ‘0.61%’ and the meta-learner (MLP OVA) by ‘8.93%’.

Table 5.74. 10-Times 10-Fold Cross-Validation Results of the ‘Multiclass J48 + MLP OVA’ Stacked Classifier for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Stacking Multi. J48 + MLP OVA	Accuracy(%)	76.83	78.05	78.05	78.05	78.05	78.05	78.05	78.05	76.83	75.61	77.56
	Kappa	0.613	0.631	0.631	0.631	0.631	0.631	0.631	0.631	0.613	0.593	0.624
	Precision	0.618	0.620	0.620	0.620	0.620	0.620	0.620	0.620	0.613	0.603	0.617
	Recall	0.768	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.768	0.756	0.775
	Specificity	0.849	0.851	0.851	0.851	0.851	0.851	0.851	0.851	0.850	0.839	0.850
	AUROC	0.827	0.796	0.788	0.812	0.801	0.809	0.805	0.802	0.794	0.793	0.803

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	0	89	0	31
Cost Comp. Only	1	366	0	13
Time Comp. Only	0	0	0	50
Cost & Time Comp.	0	0	0	270

The average for Kappa statistic value is ‘0.624’ that shows a substantial agreement. The weighted average precision value is ‘0.617’. The weighted average sensitivity (recall) value is ‘0.775’ that means the algorithm achieved ‘77.5%’ success in identifying true positive instances. Similarly, the weighted average specificity value is ‘0.850’ showing the algorithm achieved ‘85.0%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.803’.

10-fold cross-validation results with 10 repeats obtained by combining the multiclass J48 and RBF kernel SVM ECC algorithms are given in Table 5.75. These ensemble classifiers have an average classification accuracy of ‘77.20%’ with lower and upper bounds (76.37% - 78.02%) within 95% CI. In other words, ensemble classifiers predict the potential compensation type that can be acquired in a dispute with an average success rate of ‘77.20%’. Thus, the stacked classifier enhanced the average classification accuracy of the base-learner (multiclass J48) by ‘0.25%’ and the meta-learner (RBF kernel SVM ECC) by ‘3.79%’.

Table 5.75. 10-Times 10-Fold Cross-Validation Results of the ‘Multiclass J48 + RBF Kernel SVM ECC’ Stacked Classifier for Potential Compensation Prediction

Classifier	**Performance Measure**	**Run Number**										**Avg.**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Stacking Multi. J48 + RBF Kernel SVM ECC	Accuracy(%)	75.61	78.05	78.05	75.61	78.05	75.61	78.05	78.05	76.83	78.05	77.20
	Kappa	0.595	0.631	0.631	0.595	0.631	0.594	0.631	0.631	0.613	0.631	0.619
	Precision	0.615	0.620	0.620	0.615	0.620	0.615	0.620	0.620	0.613	0.620	0.618
	Recall	0.756	0.780	0.780	0.756	0.780	0.756	0.780	0.780	0.768	0.780	0.772
	Specificity	0.847	0.851	0.851	0.850	0.851	0.848	0.851	0.851	0.850	0.851	0.850
	AUROC	0.803	0.812	0.814	0.806	0.810	0.808	0.820	0.811	0.813	0.806	0.810

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	0	89	0	31
Cost Comp. Only	1	368	0	11
Time Comp. Only	0	0	0	50
Cost & Time Comp.	2	0	3	265

The average for Kappa statistic value is ‘0.619’ that shows a substantial agreement. The weighted average precision value is ‘0.618’. The weighted average sensitivity (recall) value is ‘0.772’ that means the algorithm achieved ‘77.2%’ success in identifying true positive instances. Similarly, the weighted average specificity value is ‘0.850’ showing the algorithm achieved ‘85.0%’ success in identifying true negative instances. The weighted average AUROC value is ‘0.810’.

5.3.4.6. Results from Stacking Technique for Resolution Method Selection

It was experienced in previous problems that not all stacked ensemble classifiers achieved better classification accuracies than the single ones they contain. A similar situation also exist for the multiclass classification problem of resolution method selection. For this purpose, the classification accuracy of the ensemble model is compared with accuracies of both of the single classifiers they contain.

When the base-learner is the Naïve Bayes algorithm using ECC technique (best performing single Naïve Bayes algorithm for resolution method selection), none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models ‘Naïve Bayes ECC + kNN ECC’, ‘Naïve Bayes ECC + J48 ECC’, ‘Naïve Bayes ECC + MLP ECC, ‘Naïve Bayes ECC + Polynomial kernel SVM ECC’, and ‘Naïve Bayes ECC + Gaussian RBF kernel SVM ECC’ are not considered.

Similarly, when the base learner is the MLP ECC algorithm (best performing single MLP algorithm for resolution method selection), none of the ensemble classifiers achieved better classification accuracy than single algorithms they contain. Therefore, classification results of ensemble models containing MLP ECC algorithm as base-learner are not considered.

When the base-learner is J48 ECC algorithm (best performing single J48 algorithm for resolution method selection), only one ensemble model achieved better accuracy than single algorithms they contain. This is ‘J48 ECC + Naïve Bayes ECC’ combination.

10-fold cross-validation results with 10 repeats obtained by combining the J48 ECC and Naïve Bayes ECC algorithms are given in Table 5.76. These ensemble classifiers have an average classification accuracy of '86.67%' with lower and upper bounds (85.04% - 88.30%) within 95% CI. In other words, ensemble classifiers predict the resolution method to be used in construction projects with an average success rate of '86.67%'. Thus, the stacked classifier enhanced the average classification accuracy of the base-learner (J48 ECC) by '0.19%' and the meta-learner (Naïve Bayes ECC) by '0.74%'.

Table 5.76. 10-Times 10-Fold Cross-Validation Results of the 'J48 ECC+ Naïve Bayes ECC' Stacked Classifier for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
Stacking J48 ECC + Naïve Bayes ECC	Accuracy(%)	85.19	87.04	88.89	83.33	88.89	85.19	87.04	88.89	88.89	83.33	86.67
	Kappa	0.813	0.836	0.860	0.791	0.860	0.813	0.837	0.860	0.860	0.789	0.832
	Precision	0.860	0.880	0.904	0.850	0.904	0.869	0.887	0.904	0.904	0.845	0.881
	Recall	0.852	0.870	0.889	0.833	0.889	0.852	0.870	0.889	0.889	0.833	0.867
	Specificity	0.957	0.959	0.969	0.953	0.969	0.964	0.967	0.969	0.969	0.954	0.963
	AUROC	0.965	0.950	0.967	0.961	0.958	0.949	0.959	0.957	0.963	0.961	0.959

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	86	1	3	0	0	0
Arbitration	3	57	0	0	0	0
DRB	9	0	41	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	86	14
Negotiation	0	0	0	1	41	148

The average for Kappa statistic value is '0.832' that shows a perfect agreement. The weighted average precision value is '0.881'. The weighted average sensitivity (recall) value is '0.867' that means the algorithm achieved '86.7%' success in identifying true positive instances. Similarly, the weighted average specificity value is '0.963' showing the algorithm achieved '96.3%' success in identifying true negative instances. The weighted average AUROC value is '0.959', which is almost an ideal AUROC value.

5.3.4.7. The AdaBoost Algorithm and its Configuration in WEKA

In order to use the AdaBoost algorithm with single multiclass classifiers, the class 'weka.classifiers.meta.AdaBoostM1' should be selected first. Then, the 'classifier' setting should be set to 'weka.classifiers.meta.MultiClassClassifier' class. Inside the multiclass classifier, each single ML algorithm should be defined with its corresponding configuration one by one as weak learner. The corresponding decomposition technique for each single algorithm is adjusted by the 'method' setting in the multiclass classifier. Similar to the binary case, using resampling technique did not improve the performance. Default values of WEKA are used for remaining settings. Figure 5.29 shows the configuration for the AdaBoost algorithm.

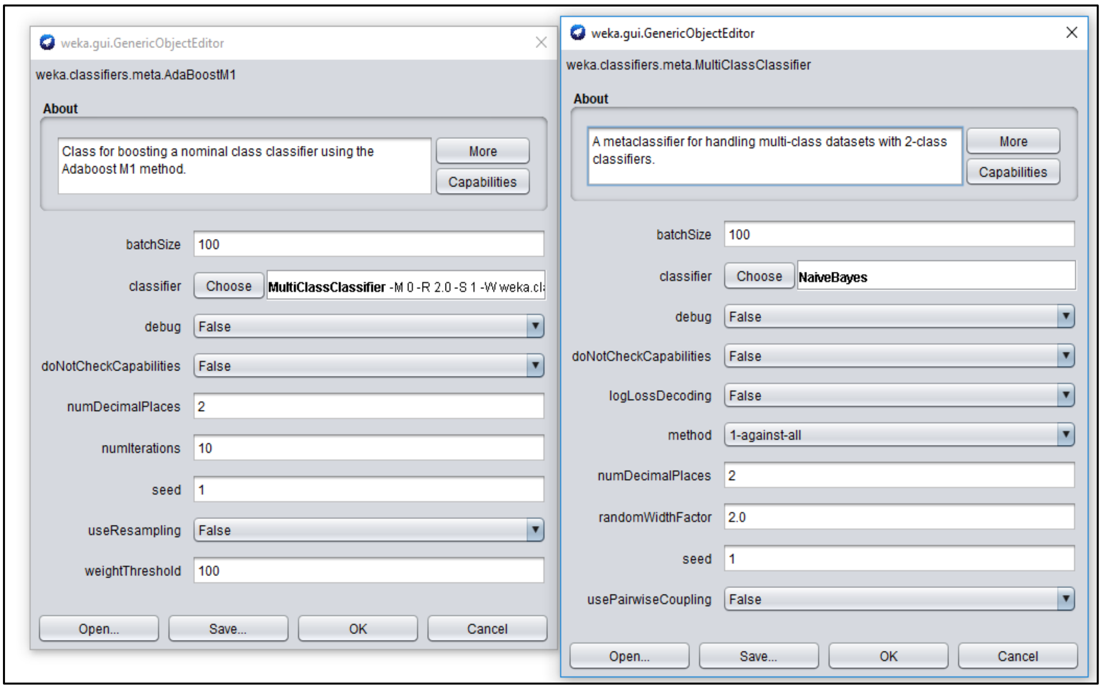


Figure 5.29. The AdaBoost Algorithm Configuration in WEKA for Multiclass Classification Problems

5.3.4.8. Results from the AdaBoost Algorithm for Potential Compensation Prediction

In this thesis study, all six single ML algorithms are boosted by the AdaBoost algorithm for multiclass classification problem of potential compensation prediction. During boosting, each single algorithm is used with the decomposition technique and parameter setting that gave the best performance. Thus, the Naïve Bayes algorithm using OVA technique, multiclass kNN algorithm, multiclass J48 algorithm, MLP algorithm using OVA technique, polynomial kernel SVM using ECC technique, and Gaussian RBF kernel SVM using ECC technique are used as weak learners one by one. Among all boosting experiments, the only enhancement is achieved in boosting of MLP OVA algorithm. In other experiments, the single ML algorithm outperformed the boosted version.

10-fold cross-validation results with 10 repeats obtained from the AdaBoost algorithm that combined MLP OVA classifiers to form an ensemble classifier are given in Table 5.77. These ensemble classifiers have an average classification accuracy of '71.10%' with lower and upper bounds (69.40% - 72.80%) within 95% CI. In other words, ensemble classifiers predict the potential compensation type that can be acquired in a dispute with an average success rate of '71.10%'. Thus, the boosted classifier enhanced the average classification accuracy of the single algorithm (MLP OVA) by '2.47%'.

The average for Kappa statistic value is '0.552' that shows a moderate agreement. The weighted average precision value is '0.689'. The weighted average sensitivity (recall) value is '0.711' that means the algorithm achieved '71.1%' success in identifying true positive instances. Similarly, the weighted average specificity value is '0.876' showing the algorithm achieved '87.6%' success in identifying true negative instances. The weighted average AUROC value is '0.853'.

Table 5.77. 10-Times 10-Fold Cross-Validation Results of the AdaBoost Algorithm with Ensemble MLP OVA Classifiers for Potential Compensation Prediction

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
AdaBoost MLP OVA	Accuracy(%)	69.51	68.29	74.39	70.73	70.73	71.95	75.61	70.73	68.29	70.73	71.10
	Kappa	0.521	0.506	0.607	0.543	0.550	0.572	0.618	0.546	0.504	0.554	0.552
	Precision	0.660	0.652	0.730	0.677	0.697	0.720	0.734	0.680	0.638	0.705	0.689
	Recall	0.695	0.683	0.744	0.707	0.707	0.720	0.756	0.707	0.683	0.707	0.711
	Specificity	0.858	0.859	0.894	0.867	0.877	0.894	0.887	0.875	0.855	0.890	0.876
	AUROC	0.875	0.860	0.872	0.853	0.842	0.832	0.862	0.851	0.843	0.841	0.853

Confusion Matrix

	Predicted			
Actual	No Comp.	Cost Comp. Only	Time Comp. Only	Cost & Time Comp.
No Comp.	30	61	5	24
Cost Comp. Only	69	306	5	0
Time Comp. Only	0	1	4	45
Cost & Time Comp.	3	0	24	243

5.3.4.9. Results from the AdaBoost Algorithm for Resolution Method Selection

In this thesis study, all six single ML algorithms are boosted by the AdaBoost algorithm for multiclass classification problem of resolution method selection. During boosting, each single algorithm is used with the decomposition technique and parameter setting that gave the best performance. Thus, the Naïve Bayes algorithm using ECC technique, kNN algorithm using ECC technique, J48 algorithm using ECC technique, MLP algorithm using ECC technique, polynomial kernel SVM using ECC technique, and Gaussian RBF kernel SVM using ECC technique are used as weak learners one by one. Among all boosting experiments, the only enhancement is achieved in boosting of J48 ECC algorithm. In other experiments, the single ML algorithm outperformed the boosted version.

10-fold cross-validation results with 10 repeats obtained from the AdaBoost algorithm that combined J48 ECC classifiers to form an ensemble classifier are given in Table 5.78. These ensemble classifiers have an average classification accuracy of ‘88.15%’ with lower and upper bounds (85.34% - 90.95%) within 95% CI. In other words, ensemble classifiers predict the resolution method to be used in construction projects with an average success rate of '88.15%'. Thus, the boosted classifier enhanced the average classification accuracy of the single algorithm (J48 ECC) by '1.67%'.

Table 5.78. 10-Times 10-Fold Cross-Validation Results of the AdaBoost Algorithm with Ensemble J48 ECC Classifiers for Resolution Method Selection

Classifier	Performance Measure	Run Number										Avg.
		1	2	3	4	5	6	7	8	9	10	
AdaBoost J48 ECC	Accuracy(%)	92.59	88.89	90.74	85.19	87.04	81.48	92.59	85.19	92.59	85.19	88.15
	Kappa	0.906	0.859	0.883	0.813	0.836	0.767	0.906	0.814	0.906	0.814	0.850
	Precision	0.930	0.899	0.921	0.860	0.877	0.830	0.931	0.879	0.931	0.876	0.893
	Recall	0.926	0.889	0.907	0.852	0.870	0.815	0.926	0.852	0.926	0.852	0.882
	Specificity	0.978	0.971	0.973	0.957	0.971	0.951	0.979	0.967	0.979	0.963	0.969
	AUROC	0.968	0.956	0.968	0.950	0.966	0.948	0.982	0.972	0.963	0.967	0.964

Confusion Matrix

	Predicted					
Actual	Litigation	Arbitration	DRB	Mediation	SEA	Negotiation
Litigation	77	9	4	0	0	0
Arbitration	1	59	0	0	0	0
DRB	5	0	45	0	0	0
Mediation	0	0	0	50	0	0
SEA	0	0	0	0	88	12
Negotiation	0	0	0	0	33	157

The average for Kappa statistic value is '0.850' that shows a perfect agreement. The weighted average precision value is '0.893'. The weighted average sensitivity (recall) value is '0.882' that means the algorithm achieved '88.2%' success in identifying true positive instances. Similarly, the weighted average specificity value is '0.969' showing the algorithm achieved '96.9%' success in identifying true negative instances. The weighted average AUROC value is '0.964'.

5.3.5. Comparison of Results from Ensemble Classifiers for Potential Compensation Prediction

The Table 5.79 shows the 10-times 10-fold cross-validation results of ensemble classifiers that performed better than their single counterparts for potential compensation prediction. This table is used for comparing performances of ensemble classifiers with each other.

Table 5.79. 10-Times 10-Fold Cross-Validation Performance of Ensemble Classifiers for Potential Compensation Prediction

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Weig. Avg. Prec.	Weig. Avg. Recall (TPR)	Weig. Avg. Spec.	Weigh. Avg. AUROC	Improve Base Learner Accuracy	Improve Meta Learner Accuracy
Majority Voting	**80.61**	**[79.57-81.65]**	**0.688**	**0.755**	**0.806**	**0.894**	**0.850**	-	NA
Stacking Multi. J48 + MLP OVA	77.56	[76.95-78.17]	0.624	0.617	0.775	0.850	0.803	+0.61%	+8.93%
Stacking Multi. J48 + RBF SVM ECC	77.20	[76.37-78.02]	0.619	0.618	0.772	0.850	0.810	+0.25%	+3.79%
AdaBoost MLP OVA	71.10	[69.40-72.80]	0.552	0.689	0.711	0.876	0.853	+2.47%	NA

The best average classification accuracy is obtained from the majority voting technique that combined prediction decisions from Naïve Bayes algorithm using OVA technique, multiclass kNN algorithm, and multiclass J48 algorithm. However, it is outperformed by the single Naïve Bayes OVA algorithm.

Considering that even the best ensemble classifier generated a lower performance compared to single ML algorithms, it can be said that ensemble classifiers are outperformed by single counterparts for potential compensation prediction.

5.3.6. Comparison of Results from Ensemble Classifiers for Resolution Method Selection

The Table 5.80 shows the 10-times 10-fold cross-validation results of ensemble classifiers that performed better than their single counterparts for resolution method selection. This table is used for comparing performances of ensemble classifiers with each other.

Table 5.80. 10-Times 10-Fold Cross-Validation Performance of Ensemble Classifiers for Resolution Method Selection

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Weig. Avg. Prec.	Weig. Avg. Recall (TPR)	Weig. Avg. Spec.	Weig. Avg. AUROC	Improve Base Learner Accuracy	Improve Meta Learner Accuracy
Majority Voting	**89.44**	**[87.37-91.52]**	**0.866**	**0.900**	**0.894**	0.965	0.930	+2.96% to best base learner	NA
Stacking J48 ECC + Naïve Bayes ECC	86.67	[85.04-88.30]	0.832	0.881	0.867	0.963	0.959	+0.19%	+0.74%
AdaBoost J48 ECC	88.15	[85.34-90.95]	0.850	0.893	0.882	**0.969**	**0.964**	+1.67%	NA

The best average classification accuracy (89.44%) is obtained from the majority voting technique that combined prediction decisions from J48 algorithm using ECC technique, Naïve Bayes algorithm using ECC technique, and MLP algorithm using ECC technique. The improvement in average classification accuracy by using the majority voting technique is '+2.96%' to the best single classifier (J48 ECC), '+3.51%'to the second best single classifier (Naïve Bayes ECC), and '+6.11%' to third best single classifier (MLP ECC).

The improvement in average classification accuracy by using the stacked generalization method is '+0.19%' to base-learner (J48 ECC) and '+0.74%' to meta-learner (Naïve Bayes ECC). However, ensemble classifiers obtained by stacking have lower performance compared to ensemble classifiers obtained from the majority voting technique and the AdaBoost algorithm.

Finally, the improvement in average classification accuracy by using the AdaBoost algorithm is '+1.67%' to the base-learner (J48 ECC). Although this classifier performed better than the stacked classifier, it is outperformed by the majority voting technique.

5.3.7. Comparison of All Classifiers for Compensation Prediction

Table 5.81 shows the 10-times 10-fold cross-validation results of all classifiers (single and ensemble) for potential compensation prediction together for comparison.

Table 5.81. Comparison of All Potential Compensation Prediction Classifiers

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Weigh. Avg. Precision	Weigh. Avg. Recall (TPR)	Weigh. Avg. Specificity	Weigh. Avg. AUROC	Rank
Naïve Bayes OVA	**80.61**	**[80.11 – 81.10]**	**0.691**	**0.774**	**0.806**	**0.899**	**0.916**	1
Majority Voting	80.61	[79.57-81.65]	0.688	0.755	**0.806**	0.894	0.850	2
KNN Multiclass	78.66	[77.05 – 80.26]	0.661	0.737	0.787	0.893	0.912	3
Stacking Multi. J48 + MLP OVA	77.56	[76.95-78.17]	0.624	0.617	0.775	0.850	0.803	4
Stacking Multi. J48 + RBF SVM ECC	77.20	[76.37-78.02]	0.619	0.618	0.772	0.850	0.810	5
J48 Multiclass	76.95	[75.99 – 77.91]	0.616	0.632	0.769	0.852	0.811	6
Poly Kernel SVM ECC	74.39	[72.97 – 75.81]	0.584	0.660	0.744	0.855	0.825	7
RBF Kernel SVM ECC	73.41	[72.34 – 74.49]	0.564	0.626	0.733	0.845	0.827	8
AdaBoost MLP OVA	71.10	[69.40-72.80]	0.552	0.689	0.711	0.876	0.853	9
MLP OVA	68.63	[67.03 – 70.23]	0.512	0.666	0.686	0.864	0.842	10

As it can be observed from Table 5.81, the best performance is obtained from the single Naïve Bayes algorithm using OVA technique. An average classification accuracy of '80.61%' is obtained from Naïve Bayes OVA classifiers. Although the same average classification accuracy is also generated by ensemble classifiers obtained from the majority voting technique, in every other performance measure the Naïve Bayes OVA algorithm is superior to remaining algorithms. The average classification accuracy of all classifiers for potential compensation prediction within 95% CI can be seen in Figure 5.30.

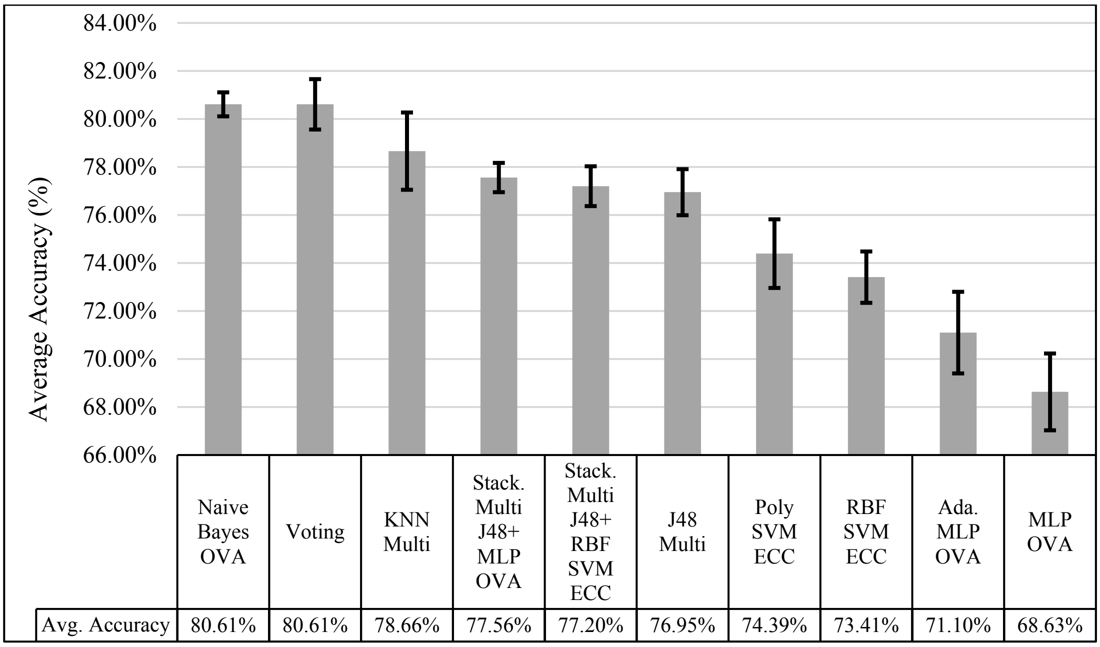


Figure 5.30. Average Classification Accuracies of All Classifiers within 95% CI for Potential Compensation Prediction

In the light of these, the final model for potential compensation prediction is the single classifier obtained from the Naïve Bayes algorithm using OVA technique.

5.3.8. Comparison of All Classifiers for Resolution Method Selection

Table 5.82 shows the 10-times 10-fold cross-validation results of all classifiers (single and ensemble) for resolution method selection together for comparison.

Table 5.82. Comparison of All Dispute Occurrence Prediction Classifiers

Algorithm	Avg. Accuracy (%)	%95 CI Accuracy (%)	Avg. Kappa	Weigh. Avg. Precision	Weigh. Avg. Recall (TPR)	Weigh. Avg. Specificity	Weigh. Avg. AUROC	Rank
Majority Voting	**89.44**	**[87.37-91.52]**	**0.866**	**0.900**	**0.894**	0.965	0.930	1
AdaBoost J48 ECC	88.15	[85.34-90.95]	0.850	0.893	0.882	**0.969**	**0.964**	2
Stacking J48 ECC + Naïve Bayes ECC	86.67	[85.04-88.30]	0.832	0.881	0.867	0.963	0.959	3
J48 ECC	86.48	[85.08 – 87.88]	0.830	0.879	0.865	0.967	0.964	4
Naïve Bayes ECC	85.93	[84.50 – 87.35]	0.817	0.875	0.859	0.935	0.961	5
MLP ECC	83.33	[80.54 – 86.13]	0.790	0.857	0.833	0.953	0.948	6
Poly. Kernel SVM ECC	82.04	[79.17 – 84.90]	0.773	0.839	0.820	0.944	0.945	7
RBF Kernel SVM ECC	80.93	[79.84 – 82.02]	0.760	0.833	0.809	0.943	0.944	8
KNN ECC	74.63	[72.46 – 76.80]	0.674	0.769	0.746	0.910	0.908	9

As it can be seen from Table 5.82, ensemble classifiers outperformed single classifiers in terms of average prediction accuracy. The first three best performing classifiers are ensemble classifiers. The best single classifier (J48 algorithm using ECC technique) has the fourth rank in overall comparison. The average classification accuracy of all classifiers for resolution method selection within 95% CI can be seen in Figure 5.31. In addition, the best classifier, which is the ensemble classifier obtained from the majority voting technique, gave the best performance in every measure other than specificity and AUROC. The best weighted average specificity ('0.969') and the best weighted average AUROC value ('0.964') are generated by ensemble classifiers obtained from the AdaBoost algorithm on J48 ECC classifiers.

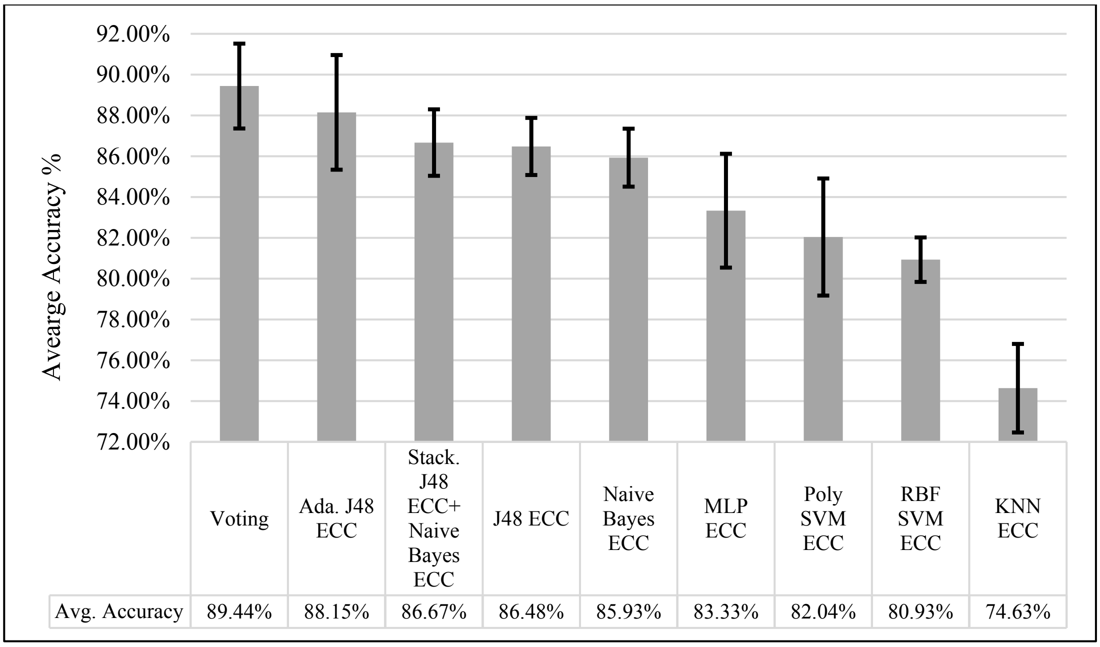


Figure 5.31. Average Classification Accuracies of All Classifiers within 95% CI for Resolution Method Selection

In the light of these, the final model for resolution method selection is the ensemble classifier obtained from the majority voting method that combined J48 algorithm using ECC technique, Naïve Bayes algorithm using ECC technique, and MLP algorithm using ECC technique.

In short, final classifiers for dispute occurrence and potential compensation prediction models as well as the final classifier for resolution method selection are determined in this chapter based on averaged 10 times 10-fold cross-validated classification results of various ML algorithms. The next chapter will include the overview of the research, conclusions, and contributions to literature and industry along with research limitations and recommendations for future works.

CHAPTER 6

CONCLUSIONS

In this chapter, the research overview will be given and findings of the thesis study will be highlighted. The process for selecting final classifiers for binary data classification problem of dispute occurrence prediction as well as multiclass data classification problems of potential compensation prediction and resolution method selection will be summarized. Strengths and weaknesses of this research to similar research and contributions to the literature will be mentioned along with potential benefits to the construction industry and dispute management domain. Limitations of the research and potential enhancements to final classifiers will be discussed and further recommendations will be given for future research.

As mentioned in the introduction chapter, detrimental effects of disputes in the construction industry is well understood and documented however, the industry still struggles to find methods to resolve disputes effectively. In the current state, the construction industry has acquired a bad reputation for being contentious and is overwhelmed by the increasing number and severity of disputes (Arditi et al., 1998; Cheng et al., 2009). This is a clear proof that current practices are insufficient in avoiding disputes. Therefore, an early-warning of dispute occurrence will be beneficial for the management personnel so that necessary actions to avoid disputes can be taken. This can be achieved by prediction (Fenn, 2007). Machine learning (ML) algorithms present necessary tools for dispute occurrence prediction and when the output variable is a categorical variable, prediction problems become data classification problems (Chou and Lin, 2002). In the case of dispute occurrence prediction, there is a binary data classification problem that can be solved by various ML algorithms.

Another aspect to be considered during dispute prediction is the potential compensation(s) that can be acquired out of a disputed case. If parties know whether they can acquire any compensation or not in a disputed case with some certainty, their decisions and strategies might change. An accurate compensation prediction may result in backing down on claims so that disputes can be avoided. In the case of potential compensation prediction, there is a multiclass data classification problem that can be solved systematically by ML algorithms.

Upon inevitable occurrence, disputes should be resolved by using the most appropriate method available with the best management efforts. However, successful dispute management is dependent on making complex and challenging decisions and the current tendency in the industry is to make these decisions intuitively based on experience of the decision-maker with limited available information of questionable quality (Chou et al., 2013b). Thus, there is a need for new decision support technologies that is based on systematical selection of resolution methods instead of a subjective decision-making process. This subjectivity can be minimized by utilizing ML techniques in systematical selection of resolution methods (Cheung et. al, 2004a). In the case of resolution method selection, there is a multiclass data classification problem that can be solved systematically by ML techniques.

In short, the construction industry requires development and employment of adequate decision support technologies in order to avoid disputes and upon inevitable occurrence, to forestall and mitigate disputes via appropriate resolution methods. However, the disputes literature lacks such supporting models or systems (İlter and Dikbaş, 2008). As mentioned in the literature review on dispute prediction (Section 2.1.3), the available limited studies lack consideration of numerous complex and interrelated factors related to disputes. Due to their limited capabilities in discovering complex and interrelated factors between input variables, it can be claimed that prediction studies that do not utilize ML techniques are insufficient. Considering the multitude of participants, various sources of uncertainties, and numerous variables in construction industry, the utilization of ML techniques in construction dispute domain is a necessity. Another problem of dispute management models and systems is the level of representation. Dispute management models and systems are mainly based on local industries and they are not capable of representing the construction industry as a whole. Moreover, instead of a general approach, it is observed that the main preference is to conduct the research for specific project types such as public projects only or PPP projects only. Despite the fact that there are various parties from many domains in a construction project, most of the studies also fail to represent or target these various professions since they merely review certain groups. In addition, previous studies on dispute prediction using ML techniques generally focused on specific change order disputes or on conventional contracting projects, which means ignoring variations in the project environment and its characteristics (Chou et al., 2013b). Similarly, based on the literature review on decision support systems for resolution method selection (Section 2.2.4), it is observed that studies have limitations such as being industry, project type, dispute type, and contracting strategy specific. Global-scaled models that consider these variations during resolution method selection do not exist in the literature.

In the light of the foregoing considerations, the research at the core of this thesis study addresses the need for sound decision support technologies for dispute prediction (occurrence and potential compensation prediction) and resolution method selection that are capable of representing the industry on a global-scale with a general approach that can be benefited by various project participants. For this purpose, various real construction project data is collected and processed to establish a disputes database. The collected database is classified by models derived from utilizing ML algorithms. Numerous experiments are performed on the collected data using several ML algorithms with the aim of presenting the best classifier in terms of classification (prediction) accuracy for dispute occurrence and potential compensation prediction along with resolution method selection in order to fill the mentioned gaps in the construction dispute literature.

Before starting the efforts for data collection, an extensive analysis of literature on construction conflicts, claims, disputes, and resolution methods with the aim of synthesizing findings of the previous research is conducted. As a result of this literature review, frameworks for dispute occurrence, potential compensation, and resolution method are established. These frameworks involve input variables (attributes) that may impact outputs of proposed models. In the light of these frameworks, a conceptual model is developed that includes all attributes identified in the literature review. Currently, there is no consensus in the literature on variables that affect dispute development and decision-making in dispute management strategies. Such a conceptual model that synthesized findings of previous research will contribute to the literature by creating a common ground for dispute prediction and resolution method selection research. Moreover, the conceptual model will eliminate the confusion in dispute management terminology due to overlapping concepts.

Within the scope of this thesis study, by using the conceptual model, a questionnaire is designed for collecting empirical data via face-to-face and online meetings with authorized project participants. As mentioned in Section 3.2, in order to reflect variations in the construction industry, the collected dataset is composed of 108 construction projects executed in 19 different countries. In addition, these projects are obtained from 78 individuals from 6 different nationalities representing 75 different construction companies. This complete dataset is used in dispute occurrence prediction. Among these 108 projects, 82 disputed cases are obtained. These 82 cases are used in potential compensation prediction. Finally, among these 82 disputed cases, the satisfactorily resolved 54 cases are used in resolution method selection model. Considering that there are limited studies focusing on interrelations between disputes and various project characteristics based on empirical data, this thesis study will have another contribution to the literature with its data dependent nature.

The collected data is initially analyzed to discover the profile of participants (Section 3.2.1). The profile of projects with respect to output variables (Section 3.2.2) and input variables (Section 3.2.3) are also analyzed. These analyses highlighted several important findings, some of which can be listed as:

i. The construction industry is dominated by disputed projects with the dataset containing 65% of disputed projects.

ii. The dispute management decision-making is performed by engineers mostly (75.9%). Considering that dispute management decisions in construction industry requires both technical and legal backgrounds, decision support technologies for resolution method selection proposed in this thesis study can said to be beneficial for the management personnel from engineering domain who lack legal expertise.

iii. The highest dispute occurrence rate is observed when the dispute management decision-making is performed by legal representatives (80%). This is mainly due to considering the disputed issue only from the legal perspective, not from the technical perspective. Thus, this thesis study is envisaged to be beneficial for legal representatives who lack technical expertise.

iv. With the increasing amount in contract values, dispute occurrence rates also increase. In addition, with the increasing planned project duration, dispute occurrence rates also increase.

v. Dispute occurrence in international projects is more likely (82%) compared to domestic projects (59%).

vi. When there are time extensions, more disputes are encountered. Moreover, with the increasing amount of time extensions, dispute occurrence rates also increase.

vii. The most preferred resolution method is the negotiation technique (35%). Considering advantages of negotiation (Section 2.2.2.4), it is beneficial for the construction industry to utilize negotiation processes in most of the cases. In accordance with this, the second most resorted resolution method is senior executive appraisal (SEA) (19%). However, the third most resorted resolution method is litigation (17%). Considering the claim that litigation should be avoided even with the best outcomes, the situation in the dataset poses a contradiction.

viii. Construction professionals are least familiar with the dispute review board (DRB) method and most familiar with the negotiation and SEA methods. The familiarity with SEA is surprising however; the reason of this is the tendency of professionals to perceive SEA as a form of negotiation that is performed with the top-level management and owners.

ix. In terms of resolution costs, it is seen that arbitration is the most expensive method of resolution. Although litigation is generally associated with high resolution costs, it is the second most expensive resolution method behind arbitration. The DRB technique has lower costs compared to conventional resolution techniques and mediation is even less expensive than DRB. Thus, it might be better to resolve disputes by alternative dispute resolution (ADR) techniques instead of resorting to conventional methods of arbitration and litigation considering resolution costs.

x. In terms of resolution duration, it is observed that litigation is the longest procedure. Although arbitration is initially considered as a fast alternative to litigation, arbitral proceedings have the second longest duration. Mediation is the most effective resolution technique in terms of resolution duration. Negotiation, SEA, and DRB also have comparable resolution duration to mediation. Thus, it might be better to resolve disputes by ADR techniques instead of resorting to conventional methods of arbitration and litigation considering resolution durations.

As mentioned in Section 4.1.5, the performance of ML algorithms is generally affected negatively by the irrelevant or insignificant attributes (Pulket and Arditi, 2009b). Elimination of insignificant attributes and selection of the ones affecting the model outcomes improve generalization performance of ML algorithms (Sönmez and Sözgen, 2017). Therefore, following the analysis to reveal the initial findings, the Chi-

Square tests of association are performed for evaluating the statistical significance of association between input and output variables in datasets. According to the results of the Chi-Square tests, insignificant attributes on outcomes are eliminated from the conceptual model and final prediction models are developed by using the remaining significant attributes. What follows provides details on the selected attributes for the proposed prediction models.

The conceptual model involved 25 attributes associated with the dispute occurrence. According to results of the Chi-Square tests (Section 3.3.2.1) on the complete dataset (108 projects), only 14 attributes are found to be significantly associated with dispute occurrence. Thus, the finalized prediction model showed that dispute occurrence is affected by project characteristics, skills of parties, changes, and delays.

The conceptual model involved 32 attributes associated with the potential compensation type. According to results of the Chi-Square tests (Section 3.3.2.2) on disputed cases (82 projects), only 9 attributes are found to be significantly associated with compensations. Therefore, the finalized prediction model showed that potential compensation that can be acquired in a disputed case is affected by project characteristics, changes, delays, and dispute characteristics.

Finally, the conceptual model involved 55 attributes associated with the resolution method selection. According to the results of the Chi-Square tests (Section 3.3.2.3) on satisfactorily resolved cases (54 projects), only 7 attributes are found to be significantly associated with the resolution method selection. Thus, the finalized selection model showed that resolution method selection decision-making is affected by project characteristics, changes, dispute characteristics, resolution method characteristics, and level of knowledge on resolution methods.

Subsequent to finalizing the prediction models, data classification was performed by using the ML algorithms. Based on the literature review on data classification using ML techniques (Chapter 4), (1) Naïve Bayes, (2) kNN, (3) C4.5 decision trees (J48 algorithm), (4) MLP, (5) polynomial kernel SVM, and (6) Gaussian RBF kernel SVM algorithms are the selected ML algorithms for developing single classifiers. These ML algorithms have various parameter settings that affect their generalization performances. For this reason, a process called parameter tuning is performed on these algorithms in order to obtain optimum parameter settings that maximize algorithm performance. In short, all single ML algorithms are experimented by using their optimum parameter settings.

Apart from single ML algorithms, ensemble algorithms are also experimented in pursuit of enhancing classification performances. The selected techniques for developing ensemble classifiers are (1) voting technique, (2) stacked generalization, and (3) AdaBoost algorithm.

Problems reviewed in this thesis study require different approaches during classification. The dispute occurrence prediction problem has only two classes as disputed projects and undisputed projects. Therefore, the dispute occurrence prediction is basically a binary data classification problem. On the other hand, potential compensation prediction problem has four classes and resolution method selection problem has six classes. Hence, they are both multiclass data classification problems. Multiclass classification problems can be solved by two different approaches. In the first approach, the ML algorithm can naturally solve both binary and multiclass classification problems. However, algorithms like SVM cannot naturally solve multiclass classification problems. In such cases, multiclass problems are decomposed into several binary problems and each problem is solved separately. Thus, the second approach is decomposing multiclass problem into several binary problems. The decomposition techniques utilized in this thesis study are (1) OVO, (2) OVA, (3) RCC, and (4) ECC techniques (Section 5.3). Both natural and decomposition approaches are experimented whenever it is possible.

In the experiments, stratified 10-fold cross-validation technique is used. Moreover, in order to decrease the variance associated with ML algorithms, all experiments are repeated 10 times and the average values (within 95% confidence intervals (CI)) are considered as the final performance measures. In evaluation of the classifier performance, (1) classification accuracy, (2) Kappa statistic, (3) precision, (4) sensitivity (recall) (TP rate), (5) specificity, and (6) AUROC measures are used (Section 4.1.4). Using these measures, performances of all single and ensemble classifiers for dispute occurrence and potential compensation prediction are compared with each other. Similarly, performances of all single and ensemble classifiers for resolution method selection are compared with each other.

The best performance for dispute occurrence prediction is obtained from the ensemble classifier generated by using majority voting technique that combined classification decisions of Gaussian RBF kernel SVM, polynomial kernel SVM, and J48 decision trees. This ensemble classifier achieved '91.11%' average classification accuracy. In other words, the dispute occurrence prediction model achieved '91.11%' success.

Among limited research on dispute occurrence prediction based on empirical data, Chou and Lin (2012) predicted dispute occurrence in PPP projects undertaken by TPCC using single and ensemble ML techniques. According to 10-fold cross-validation performances, the highest dispute occurrence prediction accuracy was '84.33%' obtained from the ensemble classifier that combined SVM, ANN, and C5.0 algorithms. They also achieved '85.60%' precision, '95.26%' sensitivity, '48.82%' specificity, and '0.7229' AUROC values. The dispute occurrence prediction classifier developed in this thesis study not only outperforms this classifier in terms of classification accuracy, but also it is capable of generating better precision ('93.70%'), sensitivity ('92.60%'), specificity ('88.40%'), and AUROC ('0.905') values; although it should be noted that compared studies used different datasets and attributes in classification. Chou et al. (2014) developed a GA based SVM model by using the same dataset that achieved '89.30%' dispute occurrence prediction success with '94.67%' precision, '74.24%' sensitivity, '93.64%' specificity, and '0.8364' AUROC values. This classifier is also outperformed by the classifier developed in this thesis study. Although precision and specificity values seem higher in Chou et al. (2014), these values are obtained from the best classifier with only a single trial and the variance in ML algorithms is not considered. On the other hand, this thesis study presents average results obtained from repeating each test 10 times. In another study that considers the variance in ML algorithms by repeating classification tests 10 times, Chou et al. (2016) achieved an average 10-fold cross-validation classification accuracy of ‘83.92%’ using C5.0 algorithm for dispute occurrence prediction in PPP projects using the same dataset in their previous attempts. Similar to dispute occurrence prediction studies, there are studies predicting the litigation likelihood of disputes. For instance, Chen and Hsu (2007) developed a hybrid model combining MLP and CBR techniques to classify construction projects with change orders according to their litigation likelihood, which achieved ‘84.61%’ classification accuracy. Based on the same dataset, Chen (2008) achieved ‘84.38%’ classification accuracy using a kNN-based model. However, compared to the model proposed herein both studies generate lower accuracies and they only focus on change order related disputes.

The best performance for potential compensation prediction is obtained from the single Naïve Bayes classifiers using OVA decomposition technique that achieved ‘80.61%’ average classification accuracy. In other words, the potential compensation prediction model achieved ‘80.61%’ success. Although there are studies on claim quantification in the literature, it is not possible to quantify a claim precisely even with the best information available (Ren et al., 2001). However, parties may benefit from a decision support that helps them understand whether they can acquire any compensation or not and in what aspect (time or cost or both) depending on the dispute source. Unfortunately, the disputes literature lacks such studies. In a similar search for predicting the dispute types, Chou et al. (2016) achieved an average 10-fold cross-validation classification accuracy of ‘77.00%’ using C5.0 algorithm, which is lower than that of the proposed model.

Finally, the best performance for resolution method selection is obtained from the ensemble classifier generated by using majority voting technique that combined classification decisions of J48 decision trees using ECC decomposition technique, Naïve Bayes algorithm using ECC decomposition technique, and MLP algorithm using ECC decomposition technique. This ensemble classifier achieved '89.44%' average classification accuracy. In other words, the resolution method selection model achieved '89.44%' success.

Among limited research on resolution method selection based on empirical data, Chou (2012) achieved '84.65%' classification accuracy on test set using an ensemble model combining QUEST, Exhaustive CHAID, and C5.0 algorithms during project initiation phase. However, the test set classification accuracy dropped to '69.05%' for dispute occurred phase using an ensemble model combining CART, Exhaustive CHAID, and SVM. In another attempt, Chou et al. (2013b) achieved '77.04% classification accuracy using fmGA based SVM model with fuzzy logic for resolution method selection. Finally, Chou et al. (2016) achieved an average 10-fold cross-validation accuracy of '81.12%' using SVM algorithm for resolution method selection. Therefore, it can be observed that the resolution method selection model developed in this thesis study outperformed the mentioned studies.

In the light of foregoing observations, it can firmly be concluded that the results obtained by the three models developed in this thesis study are promising. Potential contributions of these models to construction industry are also encouraging. Prediction of potential disputes (before occurrence) using the proposed dispute occurrence prediction model will be valuable for management personnel as it will avail early planning for taking necessary precautions. This, in turn, may reduce the effort, time, and cost of dispute management actions considerably. In addition, if parties know whether they can acquire any compensation or not and in what aspects in a disputed case with some certainty, their decisions and strategies might change. The potential compensation prediction model proposed in this thesis study can avoid inconclusive disputes along with waste of scarce resources that will be spent on resolution of these disputes. Finally, the resolution method selection model presented in this thesis study can help decision-makers in making informed and logical decisions during dispute resolution. Considering that dispute resolution decision-making requires consideration of various interrelated and complex factors along with legal and technical expertise, the presented resolution method selection model can provide assistance to management personnel who lack legal or technical expertise in cases where rationalizing these complex and interrelated factors are difficult.

Data specific nature of this research is regarded as its main limitation since the established models are data dependent. In other words, they are based on data from a finite number of construction projects. Although collected datasets can said to be quite representative, the number and variety of projects are still limited due to limitations on access to such information, research duration, and budget. The number and the variety of projects can be increased in the future so that the level of representation of the construction industry and generalization capabilities of presented models will be improved. Another future work includes converting these models to an integrated decision-support system. Further research can be performed to establish a combined decision-support system utilizing presented models via a user interface. In addition, the underlying rule sets in classification decisions of these models can be discovered so that knowledge can be extracted from association rules for more solid support.

REFERENCES

American Arbitration Association (AAA) (2018, May 22). *2017 Annual Report and Financial Statements.* Cornell University ILR School Scheinman Institute on Conflict Resolution, New York, USA. Retrieved from https://www.adr.org/sites/default/files/document_repository/AAA_AnnualReport_Financials_2018.pdf (Last access: 03 December 2019).

American Arbitration Association (AAA) (2019a). *Arbitration.* Retrieved from https://www.adr.org/Arbitration (Last access: 03 December 2019).

American Arbitration Association (AAA) (2019b). *Mediation.* Retrieved from https://www.adr.org/Mediation (Last access: 03 December 2019).

Acharya, N. K., Dai Lee, Y., & Man Im, H. (2006). Conflicting Factors in Construction Projects: Korean Perspective. *Engineering, Construction and Architectural Management*, *13*(6), 543–566. https://doi.org/10.1108/09699980610712364

Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons, New York.

Alkass, S., Mazerolle, M., & Harris, F. (1993). An Integrated System to Minimize the Cost of Analyzing Construction Claims. *Computing Systems in Engineering*, *4*(2–3), 271–280. https://doi.org/10.1016/0956-0521(93)90051-W

Alpaydın, E. (2010). *Introduction to Machine Learning by Ethem Alpaydin* (2nd ed.). Retrieved from https://www.cambridge.org/core/product/identifier/S0269888906220745/type/journal_article (Last access: 03 December 2019).

Alshawi, M. A., & Hope, A. E. (1989). Expert Systems and Contractual Disputes: Extension of Time under JCT 80. *Construction Management and Economics*, *7*(1), 65–74. https://doi.org/10.1080/01446198900000007

Altun, H., & Polat, G. (2008). On the Comparison of Classifiers' Performance in Emotion Classification: Critiques and Suggestions. *IEEE 16th Signal Processing, Communication and Applications Conference*, 1–4. https://doi.org/10.1109/siu.2008.4632592

Aly, M. (2005). Survey on Multiclass Classification Methods. *Neural Netw, 19*, 1–9.

Amin, N., & Habib, A. (2015). Comparison of Different Classification Techniques Using WEKA for Hematological Data. *American Journal of Engineering Research (AJER)*, *4*(3), 55–61.

An, S., Park, U., Kang, K., Cho, M.-Y., & Cho, H. (2007). Application of Support Vector Machines in Assessing Conceptual Cost Estimates. *Journal of Computing in Civil Engineering*, *21*(4), 259–264. https://doi.org/10.1061/(asce)0887-3801(2007)21:4(259)

Arditi, D., Oksay, F. E., & Tokdemir, O. B. (1998). Predicting the Outcome of Construction Litigation using Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, *13*(2), 75–81. https://doi.org/10.1111/0885-9507.00087

Arditi, D., & Tokdemir, O. B. (1999). Using Case-Based Reasoning to Predict the Outcome of Construction Litigation. *Computer-Aided Civil and Infrastructure Engineering*, *14*(6), 385–393. https://doi.org/10.1111/0885-9507.00157

Arditi, D., & Pulket, T. (2005). Predicting the Outcome of Construction Litigation Using Boosted Decision Trees. *Journal of Computing in Civil Engineering*, *19*(4), 387–393. https://doi.org/10.1061/(asce)0887-3801(2005)19:4(387)

Arditi, D., & Pattanakitchamroon, T. (2006). Selecting a Delay Analysis Method in Resolving Construction Claims. *International Journal of Project Management*, *24*(2), 145–155. https://doi.org/10.1016/j.ijproman.2005.08.005

Arditi, D., & Pulket, T. (2009). Predicting the Outcome of Construction Litigation Using an Integrated Artificial Intelligence Model. *Journal of Computing in Civil Engineering*, *24*(1), 73–80. https://doi.org/10.1061/(asce)0887-3801(2010)24:1(73)

Arıcı, Y. (2012). *İnşaat Sektöründe ADR (Alternatif Uyuşmazlık Çözüm Yolları) Kullanımı ve Seçim Kriterlerinin Kamu ve Özel Sektör Açısından Kıyaslanması.* Master's thesis. İstanbul Technical University.

Bal, C., Er, F., & Sönmez, H. (2009). A Review of Statistical Techniques for 2x2 and RxC Categorical Data Tables in SPSS. *Journal of Pediatric Sciences*, *1*(1), 1–10. Retrieved from http://dergipark.gov.tr/download/article-file/178758 (Last access: 03 December 2019).

Belousov, A. I., Verzakov, S. A., & Von Frese, J. (2002a). A Flexible Classification Approach with Optimal Generalisation Performance: Support Vector Machines. *Chemometrics and Intelligent Laboratory Systems*, *64*(1), 15–25. https://doi.org/10.1016/S0169-7439(02)00046-1

Belousov, A. I., Verzakov, S. A., & Von Frese, J. (2002b). Applicational Aspects of Support Vector Machines. *Journal of Chemometrics*, *16*(8–10), 482–489. https://doi.org/10.1002/cem.744

Bristow, D. I, & Vasilopoulos, R. (1995). The New CCDC 2: Facilitating Dispute Resolution of Construction Projects. *Construction Law Journal, 11*, 95-95.

Brown, H. & Marriot, A. (1993). *ADR Principles and Practice* (1st ed.), Sweet and Maxwell, London.

Brownlee, J. (2018a). *Machine Learning Mastery with WEKA* (1.5).

Brownlee, J. (2018b). *Master Machine Learning Algorithms* (1.13).

Bruce, R. R., Macmillan, R., Ellam, T. S. J., Intven, H., & Miedema, T. (2004). *Dispute Resolution in the Telecommunications Sector*. Retrieved from http://www.itu.int/ITU-D/treg/publications/ITU_WB_Dispute_Res-E.pdf (Last access: 03 December 2019).

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Chau, K. W. (2007). Application of a PSO-Based Neural Network in Analysis of Outcomes of Construction Claims. *Automation in Construction*, *16*(5), 642–646. https://doi.org/10.1016/j.autcon.2006.11.008

Chen, J. H., & Hsu, S. C. (2007). Hybrid ANN-CBR Model for Disputed Change Orders in Construction Projects. *Automation in Construction*, *17*(1), 56–64. https://doi.org/10.1016/j.autcon.2007.03.003

Chen, J. H. (2008). KNN Based Knowledge-Sharing Model for Severe Change Order Disputes in Construction. *Automation in Construction*, *17*(6), 773–779. https://doi.org/10.1016/j.autcon.2008.02.005

Cheng, M. Y., Tsai, H. C., & Chiu, Y. H. (2009). Fuzzy Case-Based Reasoning for Coping with Construction Disputes. *Expert Systems with Applications*, *36*(2 PART 2), 4106–4113. https://doi.org/10.1016/j.eswa.2008.03.025

Cheng, M. Y., & Wu, Y. W. (2009). Evolutionary Support Vector Machine Inference System for Construction Management. *Automation in Construction*, *18*(5), 597–604. https://doi.org/10.1016/j.autcon.2008.12.002

Cheung, S. O., & Yeung, Y. W. (1998). The Effectiveness of the Dispute Resolution Advisor System: A Critical Appraisal. *International Journal of Project Management*, *16*(6), 367–374. https://doi.org/10.1016/S0263-7863(98)00010-6

Cheung, S. O. (1999). Critical Factors Affecting the Use of Alternative Dispute Resolution Processes in Construction. *International Journal of Project Management*, *17*(3), 189–194. https://doi.org/10.1016/S0263-7863(98)00027-1

Cheung, S. O., Tam, C. M., Ndekugri, I., & Harris, F. C. (2000). Factors Affecting Clients' Project Dispute Resolution Satisfaction in Hong Kong. *Construction Management & Economics*, *18*(3), 281-294.

Cheung, S. O., & Suen, H. C. H. (2002). A Multi-Attribute Utility Model for Dispute Resolution Strategy Selection. *Construction Management and Economics*, *20*(7), 557–568. https://doi.org/10.1080/01446190210157568

Cheung, S. O., Suen, H. C. H., & Lam, T. (2002). Fundamentals of Alternative Dispute Resolution Processes in Construction. *Journal of Construction Engineering and Management*, *128*(5), 409–417. https://doi.org/10.1061/(asce)0733-9364(2002)128:5(409)

Cheung, S. O., Tam, C. M., & Harris, F. C. (2002). Project Dispute Resolution Satisfaction Classification through Neural Network. *Journal of Management in Engineering*, *16*(1), 70–79. https://doi.org/10.1061/(asce)0742-597x(2000)16:1(70)

Cheung, S. O., Au-yeung, R. F., & Wong, V. W. K. (2004a). A CBR Based Dispute Resolution Process Selection System. *International Journal of IT in Architecture, Engineering and Construction*, *2*(2), 129–145.

Cheung, S. O., Cheung, K. K. W., & Suen, H. C. H. (2004b). E-valuate-Med: Prototype Web-based Mediator Assessment System for Role-Plays . *Journal of Professional Issues in Engineering Education and Practice*, *130*(2), 142–151. https://doi.org/10.1061/(asce)1052-3928(2004)130:2(142)

Cheung, S. O., Yiu, T. W., & Yeung, S. F. (2006). A Study of Styles and Outcomes in Construction Dispute Negotiation. *Journal of Construction Engineering and Management*, *132*(8), 805–814. https://doi.org/10.1061/(asce)0733-9364(2006)132:8(805)

Cheung, Sai On, Yiu, T. W., & Chan, H. W. (2010). Exploring the Potential for Predicting Project Dispute Resolution Satisfaction Using Logistic Regression. *Journal of Construction Engineering and Management*, *136*(5), 508–517. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000157

Cheung, S. O., & Pang, K. H. Y. (2013). Anatomy of Construction Disputes. *Journal of Construction Engineering and Management*, *139*(1), 15–23. https://doi.org/10.1061/(asce)co.1943-7862.0000532

Chong, H. Y., & Zin, R. M. (2012). Selection of Dispute Resolution Methods: Factor Analysis Approach. *Engineering, Construction and Architectural Management. 19*(4), 428-443.

Chou, J. S. (2012). Comparison of Multilabel Classification Models to Forecast Project Dispute Resolutions. *Expert Systems with Applications*, *39*(11), 10202–10211. https://doi.org/10.1016/j.eswa.2012.02.103

Chou, J. S., & Lin, C. (2012). Predicting Disputes in Public-Private Partnership Projects: Classification and Ensemble Models. *Journal of Computing in Civil Engineering*, *27*(1), 51–60. https://doi.org/10.1061/(asce)cp.1943-5487.0000197

Chou, J. S., Tsai, C., & Lu, Y. (2013a). Project Dispute Prediction by Hybrid Machine Learning Techniques. *Journal of Civil Engineering and Management*, *19*(4), 505–517. https://doi.org/10.3846/13923730.2013.768544

Chou, J. S., Cheng, M. Y., & Wu, Y. W. (2013b). Improving Classification Accuracy of Project Dispute Resolution Using Hybrid Artificial Intelligence and Support Vector Machine Models. *Expert Systems with Applications*, *40*(6), 2263–2274. https://doi.org/10.1016/j.eswa.2012.10.036

Chou, J. S., Cheng, M. Y., Wu, Y. W., & Pham, A. D. (2014). Optimizing Parameters of Support Vector Machine Using Fast Messy Genetic Algorithm for Dispute Classification. *Expert Systems with Applications*, *41*(8), 3955–3964. https://doi.org/10.1016/j.eswa.2013.12.035

Chou, J. S., Hsu, S. C., Lin, C. W., & Chang, Y. C. (2016). Classifying Influential Information for Project Information to Discover Rule Sets for Project Disputes and Possible Resolutions. *International Journal of Project Management*, *34*(8), 1706–1716. https://doi.org/10.1016/j.ijproman.2016.10.001

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297.

Çevikbaş, M., & Köksal, A. (2018). An Investigation of Litigation Process in Construction Industry in Turkey. *Teknik Dergi*, 8715–8730. https://doi.org/10.18400/tekderg.389757

Dalton, D., & Shehadeh, N. (2003). Statistical Modelling of Claims Procedures and Construction Conflicts. *Construction Conflict Management and Resolution,* 2, 275-285.

Dibike, Y. B., Velickov, S., Solomatine, D., & Abbott, M. B. (2001). Model Induction with Support Vector Machines: Introduction and Applications. *Journal of Computing in Civil Engineering*, *15*(3), 208–216. https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208)

Diekmann, J. E., & Girard, M. J. (1995). Are Contract Disputes Predictable? *Journal of Construction Engineering and Management*, *121*(4), 355–363. https://doi.org/10.1061/(ASCE)0733-9364(1995)121:4(355)

Dietterich, T. G., & Bakiri, G. (1994). Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research,* 2, 263-286.

Drazin, S., & Montag, M. (2012). Decision Tree Analysis using Weka. *Machine Learning-Project II*, 1–3.

Drucker, H., Donghui Wu, & Vapnik, V. N. (1999). Support Vector Machines for Spam Categorization. *IEEE Transactions on Neural Networks*, *10*(5), 1048–1054. https://doi.org/10.1109/72.788645

El-adaway, I. H., & Kandil, A. A. (2009). Multiagent System for Construction Dispute Resolution (MAS-COR). *Journal of Construction Engineering and Management*, *136*(3), 303–315. https://doi.org/10.1061/(asce)co.1943-7862.0000144

Ellis, F., & Baiden, B. K. (2008). A Conceptual Model for Conflict Management in Construction Firms. *RICS & CIB International Research Conference (COBRA)*, 67–74. https://doi.org/10.1097/DAD.0b013e318213f80d

Engineering News-Record (ENR) (2018). ENR's Top 250 International Contractors. Retrieved from https://www.enr.com/toplists/2018-Top-250-International-Contractors-1 (Last access: 03 December 2019).

Eugenio, B. Di, & Glass, M. (2004). The Kappa Statistic: A Second Look. *Computational Linguistics*, *30*(1), 95–101. https://doi.org/10.1162/089120104773633402

Fan, H., & Li, H. (2013). Retrieving Similar Cases for Alternative Dispute Resolution in Construction Accidents Using Text Mining Techniques. *Automation in Construction*, *34*, 85–91. https://doi.org/10.1016/j.autcon.2012.10.014

Farid, D., Zhang, L., Mofizur, C., Hossain, M. A., & Strachan, R. (2014). Hybrid Decision Tree and Naïve Bayes Classifiers for Multi-Class Classification Tasks. *Expert Systems With Applications*, *41*(4), 1937–1946. https://doi.org/10.1016/j.eswa.2013.08.089

Fenn, P., Lowe, D., & Speck, C. (1997). Conflict and Dispute in Construction. *Construction Management and Economics*, *15*(6), 513–518. https://doi.org/10.1080/014461997372719

Fenn, P. (2007). Predicting Construction Disputes: An Aetiological Approach. *Proceedings of the Institution of Civil Engineers - Management, Procurement and Law*, *160*(2), 69–73. https://doi.org/10.1680/mpal.2007.160.2.69

Frank, E., Mark, H. A., & Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (4th ed.), Morgan Kaufmann.

Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. *13th International Conference on Machine Learning*, 148–156. Retrieved from http://www.research.att.com/orgs/ssr/people/fyoav,schapireg/ (Last access: 03 December 2019).

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139.

Friedrichs, F., & Igel, C. (2005). Evolutionary Tuning of Multiple SVM Parameters. *Neurocomputing*, *64*(1-4 SPEC. ISS.), 107–117. https://doi.org/10.1016/j.neucom.2004.11.022

Gebken, R. J., & Gibson, G. E. (2006). Quantification of Costs for Dispute Resolution Procedures in the Construction Industry. *Journal of Professional Issues in Engineering Education and Practice*, *132*(3), 264–271. https://doi.org/10.1061/(asce)1052-3928(2006)132:3(264)

Gencer, H. (2005). Kamu İhale Kanununa Göre Yapılan Yapım İşleri İhalelerindeki Uyuşmazlıklar Ve Uyuşmazlık Kararları Üzerine Bir Araştırma. *TMMOB İnşaat Mühendisleri Odası 3. Yapı İşletmesi Kongresi*, 453–464.

Girja Sharma, N. B., Bhargava, R., & Mathuria, M. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(6), 1114–1119. Retrieved from https://www.academia.edu/4375403/Decision_Tree_Analysis_on_J48_Algorithm_for_Data_Mining (Last access: 03 December 2019).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*, 1157–1182. Retrieved from http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf (Last access: 03 December 2019).

Günay, G. & Birgönül, M. T. (2001). *Türk İnşaat Sektöründe Hukuksal Anlaşmazlıkların Oluşumu ve Çözüm Yolları. Dispute Creation and Resolution Mechanisms of the Turkish Construction Sector in Public Projects.* (Turkish and English versions appear in the same book). Türkiye Müteahhitler Birliği, p. 206.

Harmon, K. M. J. (2003). Resolution of Construction Disputes: A Review of Current Methodologies. *Leadership and Management in Engineering*, *3*(4), 187–201. https://doi.org/10.1061/(ASCE)1532-6748(2003)3:4(187)

Harmon, K. M. J. (2009). Case Study as to the Effectiveness of Dispute Review Boards on the Central Artery/Tunnel Project. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *1*(1), 18–31. https://doi.org/10.1061/(asce)1943-4162(2009)1:1(18)

Heath, B., Hills, B., & Berry, M. (1994). The Nature and Origin of Conflict Within the Construction Process. *Proceedings of the First Plenary Meeting of TG15,* Publication 171, CIB Report, The Netherlands.

Hong Kong International Arbitration Center (HKIAC) (2009). HKIAC 2009 Annual Report. Retrieved from https://www.hkiac.org/sites/default/files/annual_report/2009_Annual_Report.pdf (Last access: 03 December 2019).

Hong Kong International Arbitration Center (HKIAC) (2018). HKIAC 2018 Annual Report. Retrieved from https://www.hkiac.org/sites/default/files/annual_report/annual%20report%203463-7390-6190%20v.4.pdf (Last access: 03 December 2019).

Howard, W. E., Bell, L. C., & McCormick, R. E. (1997). Economic Principles of Contractor Compensation. *Journal of Management in Engineering*, *13*(5), 81–89. https://doi.org/10.1061/(ASCE)0742-597X(1997)13:5(81)

Hsu, C. W., & Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425. https://doi.org/10.1109/72.991427

Hsu, C. W., Chang, C.-C., & Lin, C. J. (2003). *A Practical Guide to Support Vector Classification*. Technical Report, Department of Computer Science, National Taiwan University, 1–16. Retrieved from http://www.csie.ntu.edu.tw/~cjlin (Last access: 03 December 2019).

Ilter, D., Lees, M., & Dikbaş, A. (2007). Alternative Dispute Resolution: Suggestions for Application in Turkish Construction Industry. *Salford University Postgraduate Annual Conference (SPARC) Proceedings,* ISBN 978-1-905732-42-5, Manchester, UK.

Ilter, D., & Dikbas, A. (2008). The Use of Key Attributes in Alternative Dispute Resolution (ADR) Process Design. *ARCOM*, (September), 455–464.

Ilter, D., & Dikbas, A. (2009). An Investigation of the Factors Influencing Dispute Frequency in Construction Projects. In *Proceedings of RICS International Research Conference (COBRA 2009), Cape Town, South Africa*, (September), 1496–1504.

Ilter, D. (2010a). *İnşaat Projelerinde Uyuşmazlık Çözüm Yöntemi Seçimi İçin Çok Kriterli Karar Verme Modeli.* Doctoral dissertation, İstanbul Technical University.

Ilter, D. (2010b). Opinions of Legal Professionals Regarding the Selection of Appropriate Resolution Method in Construction Disputes. In *Proceedings of RICS & CIB International Research Conference* (*COBRA 2010), Paris*, 587-594. Retrieved from http://irbnet.de/daten/iconda/CIB20092.pdf (Last access: 03 December 2019).

Ilter, D. (2012). Identification of the Relations between Dispute Factors and Dispute Categories in Construction Projects. *International Journal of Law in the Built Environment*, *4*(1), 45–59. https://doi.org/10.1108/17561451211211732

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1398*, 137–142. https://doi.org/10.1007/s13928716

Jones, D. (2006). Construction Project Dispute Resolution: Options for Effective Dispute Avoidance and Management. *Journal of Professional Issues in Engineering Education and Practice*, *132*(3), 225–235. https://doi.org/10.1061/(asce)1052-3928(2006)132:3(225)

Kassab, M., Hipel, K., & Hegazy, T. (2006). Conflict Resolution in Construction Disputes Using the Graph Model. *Journal of Construction Engineering and Management*, *132*(10), 1043–1052. https://doi.org/10.1061/(asce)0733-9364(2006)132:10(1043)

Kassab, M., Hegazy, T., & Hipel, K. (2010). Computerized DSS for Construction Conflict Resolution under Uncertainty. *Journal of Construction Engineering and Management*, *136*(12), 1249–1257. https://doi.org/10.1061/(asce)co.1943-7862.0000239

Kilian, J. J., & Gibson, G. E. (2005). Construction Litigation for the U.S. Naval Facilities Engineering Command, 1982–2002. *Journal of Construction Engineering and Management*, *131*(9), 945–952. https://doi.org/10.1061/(asce)0733-9364(2005)131:9(945)

Kittler, J., Hatef, M., & Duin, R. P. W. (1996). Combining Classifiers. *13th International Conference on Pattern Recognition*, 897–901. IEEE.

Kumaraswamy, M. M. (1997). Conflicts, Claims and Disputes in Construction. *Engineering, Construction and Architectural Management*, *4*(2), 95–111. https://doi.org/10.1108/eb021042

Lavesson, N., & Davidsson, P. (2006). Quantifying the Impact of Learning Algorithm Parameter Tuning. *American Association for Artificial Intelligence*, *1*(1), 395–400.

Lee, C. K., Yiu, T. W., & Cheung, S. O. (2016). Selection and Use of Alternative Dispute Resolution (ADR) in Construction Projects - Past and Future Research. *International Journal of Project Management, 34*(3), 494-507.

Li, T., Zhang, C., & Ogihara, M. (2004). A Comparative Study of Feature Selection and Multiclass Classification Methods for Tissue Classification Based on Gene Expression. *Bioinformatics*, *20*(15), 2429–2437. https://doi.org/10.1093/bioinformatics/bth267

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data Mining Techniques and Applications - A Decade Review from 2000 to 2011. *Expert Systems with Applications*, *39*, pp. 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063

Lin, H. T., & Lin, C. J. (2003). A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-Type Methods. *Neural Computation*, *3*, 1–32. https://doi.org/200609271228537896664

Lin, S. W., Lee, Z. J., Chen, S. C., & Tseng, T. Y. (2008). Parameter Determination of Support Vector Machine and Feature Selection Using Simulated Annealing Approach. *Applied Soft Computing Journal*, *8*(4), 1505–1512. https://doi.org/10.1016/j.asoc.2007.10.012

Lipsky, D., & Seeber, R. (1998). The Appropriate Resolution of Corporate Disputes: A Report on the Growing Use of ADR by US Corporations. *Scheinman Institute on Conflict Resolution*, (January), 1–40.

Love, P. E. D., Davis, P. R., Ellis, J., & Cheung, S. O. (2010). Dispute Causation: Identification of Pathogenic Influences in Construction. *Engineering, Construction and Architectural Management*, *17*(4), 404–423. https://doi.org/10.1108/09699981011056592

Love, P. E. D., Davis, P. R., Cheung, S. O., & Irani, Z. (2011). Causal Discovery and Inference of Project Disputes. *IEEE Transactions on Engineering Management*, *58*(3), 400–411. https://doi.org/10.1109/TEM.2010.2048907

Mackie, C., South, J., Marsh, W., Alexander, N., Lorenz, O., & Barbieri, G. L. (2011). *Alternative Dispute Resolution Guidelines*. Retrieved from http://siteresources.worldbank.org/INTECA/Resources/15322_ADRG_Web.pdf (Last access: 03 December 2019).

Mahfouz, T., & Kandil, A. (2011). Litigation Outcome Prediction of Differing Site Condition Disputes through Machine Learning Models. *Journal of Computing in Civil Engineering*, *26*(3), 298–308. https://doi.org/10.1061/(asce)cp.1943-5487.0000148

Mantovani, R. G., Rossi, A. L. D., Vanschoren, J., Bischl, B., & Carvalho, A. C. P. L. F. (2015). To Tune or Not To Tune : Recommending When To Adjust SVM Hyper-Parameters via Meta-Learning. *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2015.7280644

Marzouk, M., El-Mesteckawi, L., & El-Said, M. (2011). Dispute Resolution Aided Tool for Construction Projects in Egypt. *Journal of Civil Engineering and Management*, *17*(1), 63–71. https://doi.org/10.3846/13923730.2011.554165

Mayoraz, E., & Alpaydin, E. (1999). Support Vector Machines for Multi-Class Classification. In *Efficient Learning Machines* (pp. 833–842). https://doi.org/10.1007/BFb0100551

McGeorge, D., London, K., Love, P., Davis, P., Jefferies, M., Ward, P., & Chesworth, B. (2007). Dispute Avoidance and Resolution: A Literature Review. In *CRC Construction Innovation, Rep. 1*.

Mchugh, M. L. (2012). Interrater Reliability : The Kappa Statistic. *Biochemia Medica*, *22*(3), 276–282.

McHugh, M. L. (2013). The Chi-square Test of Independence. *Biochemia Medica*, *23*(2), 143–149. https://doi.org/10.11613/BM.2013.018

Mehta, C. R., & Patel, N. R. (2012). *IBM SPSS Exact Tests*. Cambridge, MA: IBM Corporation.

Menassa, C. C., & Peña Mora, F. (2009). Analysis of Dispute Review Boards Application in U.S. Construction Projects from 1975 to 2007. *Journal of Management in Engineering*, *26*(2), 65–77. https://doi.org/10.1061/(asce)me.1943-5479.0000001

Mendenhall, W. M., & Sinchic, T. L. (2016). *Statistic for Engineering and The Sciences* (6th ed.). Chapman and Hall/CRC.

Mitropoulos, P., & Howell, G. (2002). Model for Understanding, Preventing, and Resolving Project Disputes. *Journal of Construction Engineering and Management*, *127*(3), 223–231. https://doi.org/10.1061/(asce)0733-9364(2001)127:3(223)

Molenaar, K., Washington, S., & Diekmann, J. (2000). Structural Equation Model of Construction Contract Dispute Potential. *Journal of Construction Engineering and Management*, *126*(4), 268–277. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:4(268)

Molina, M. M., Luna, J. M., Romero, C., & Ventura, S. (2012). Meta-Learning Approach for Automatic Parameter Tuning : A Case Study with Educational Datasets. *5th International Conference on Educatiıonal Data Mining*, 180–183. Chania, Greece.

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621–633. https://doi.org/10.1016/j.eswa.2012.07.059

National Alternative Dispute Resolution Advisory Council (NADRAC) (2011). Glossary of ADR Terms. Retrieved from https://treasury.gov.au/sites/default/files/2019-03/NADRAC.pdf (Last access: 03 December 2019).

Office of Government Commerce (2002). *Dispute Resolution Guidance.* Norwich, UK: Office of Government Commerce

Özer, S. (2012). *İnşaat Sözleşmelerinden Doğabilecek Uyuşmazlıkların Alternatif Hukuki Çözüm Yolları.* Adalet Yayınevi, Ankara (ISBN: 978-605-5263-46-1).

Park, S. H., Goo, J. M., & Jo, C. H. (2004). Receiver Operation Characteristics (ROC) Curve: Practical : Review for Radiologist. *Korean Journal of Radiology*, 5 (1), 11-18.

Patel, K., Vala, J., & Pandya, J. (2014). Comparison of Various Classification Algorithms on Iris Datasets using WEKA. *International Journal of Advance Engineering and Research Development (IJAERD)*, *1*(1), 1–7.

Peña-Mora, F., Sosa, C. E., & McCone, D. S. (2003). *Introduction To Construction Dispute Resolution*. MIT-Prentice Hall series on Civil, Environmental, and Systems Engineering, Pearson Education, US.

PMI. (2016). Construction Extension to the PMBOK® Guide. In *Project Management Institute*. https://doi.org/10.1002/pmj

Pulket, T., & Arditi, D. (2009a). Construction Litigation Prediction System Using Ant Colony Optimization. *Construction Management and Economics*, *27*(3), 241–251. https://doi.org/10.1080/01446190802714781

Pulket, T., & Arditi, D. (2009b). Universal Prediction Model for Construction Litigation. *Journal of Computing in Civil Engineering*, *23*(3), 178–187. https://doi.org/10.1061/(asce)0887-3801(2009)23:3(178)

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning.* Morgan Kaufmann, Los Altos, California.

Raziff, A. R. A., Sulaiman, N., Mustapha, N., & Perumal, T. (2017). Single Classifier, OvO, OvA and RCC Multiclass Classification Method in Handheld Based Smartphone Gait Identification. *AIP Conference*, *1891*(1), 020009. https://doi.org/10.1063/1.5005342

Reif, M., Shafait, F., & Dengel, A. (2011). Prediction of Classifier Training Time Including Parameter Optimization. *34th Annual German Conference on Artificial Intelligence KI11*, 260–271. Berlin, Germany.

Ren, Z., Anumba, G. J., & Ugwu, O. O. (2001). Construction Claims Management: Towards an Agent-Based Approach. *Engineering, Construction and Architectural Management*, *8*(3), 185–197. https://doi.org/10.1108/eb021181

Revay, S. G. (1993). Can Construction Claims Be Avoided?. *Building Research & Information*, *21*(1), 56–58. https://doi.org/10.1080/09613219308727256

Rhys-Jones, S. (1994). How Constructive Is Construction Law?. *Construction Law Journal, 10*(1)*,* 28-38.

Rubin, R. A., & Quintas, B. V. (2003). Alternative Dispute Resolution in U.S. Public Works: Proposed Model. *Journal of Professional Issues in Engineering Education and Practice*, *129*(2), 80–83. https://doi.org/10.1061/(asce)1052-3928(2003)129:2(80)

Sobhana, M. S. (2014). *Novel Approaches to Machine Recognition of Handwritten Tamil Documents*. Doctoral dissertation. Mahatma Gandhi University.

Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sönmez, R., & Sözgen, B. (2017). A Support Vector Machine Method for Bid/No Bid Decision Making. *Journal of Civil Engineering and Management*, *23*(5), 641–649. https://doi.org/10.3846/13923730.2017.1281836

Sözgen, B. (2009). *Neural Network and Regression Models to Decide Whether or Not to Bid for a Tender in Offshore Petroleum Platform Fabrication Industry*. Master's thesis. Middle East Technical University.

Thakkar, M., Bhatt, M., & Bhensdadia, C. K. (2011). Performance Evaluation of Classification Techniques for Computer Vision based Cashew Grading System. *International Journal of Computer Applications*, *18*(6), 9–12.

Thompson, R. M., Vorster, M. C., & Groton, J. P. (2000). Innovations to Manage Disputes: DRB and NEC. *Journal of Management in Engineering*, *16*(5), 51–59. https://doi.org/10.1061/(ASCE)0742-597X(2000)16:5(51)

Vanwinckelen, G., & Blockeel, H. (2012). On Estimating Model Accuracy with Repeated Cross-Validation. *21st Belgian-Dutch Conference on Machine Learning*, 39–44. Ghent, Belgium.

Wang, W., Xu, Z., Lu, W., & Zhang, X. (2003). Determination of the Spread Parameter in the Gaussian Kernel for Classification and Regression. *Neurocomputing*, *55*(3–4), 643–663. https://doi.org/10.1016/S0925-2312(02)00632-X

Watts, V. M., & Scrivener, J. C. (1993). Review of Australian Building Disputes Settled by Litigation. *Building Research & Information*, *21*(1), 59–63. https://doi.org/10.1080/09613219308727257

Weisburd, D., & Britt, C. (2007). Chapter 13 Measures of Association for Nominal and Ordinal Variables. In *Statistics in Criminal Justice* (pp. 335–380). https://doi.org/10.1007/978-0-387-34113-2_13

Witten, H. W., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann, Burlington, MA.

Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, *5*(2), 241–259. Retrieved from https://www.sciencedirect.com/science/article/pii/S0893608005800231%0Apapers://5e3e5e59-48a2-47c1-b6b1-a778137d3ec1/Paper/p2022 (Last access: 03 December 2019).

Yates, D. (2003). Can Claims and Disputes (in Construction Contracts) Be Prevented or Reduced?. *Building Journal Hongkong China*, (June).

Yılmaz, İ. C., & Dikbaş, A. (2013). Türk Kamu İnşaat Projelerinde Yaşanan Uyuşmazlıklara Yönelik Bir Veri Madenciliği Yaklaşımı. *AJIT-e: Online Academic Journal of Information Technology,* 4(13), 79-86. Retrieved from http://www.acarindex.com/dosyalar/makale/acarindex-1423867586.pdf (Last access: 03 December 2019).

Yiu, T. W., Cheung, S. O., & Mok, F. M. (2006). Logistic Likelihood Analysis of Mediation Outcomes. *Journal of Construction Engineering and Management*, *132*(10), 1026–1036. https://doi.org/10.1061/(asce)0733-9364(2006)132:10(1026)

Yiu, T. W., & Cheung, S. O. (2007). Behavioral Transition: A Framework for the Construction Conflict-Tension Relationship. *IEEE Transactions on Engineering Management, 54*(3), 498-505.

York, S. D. (1996). *Practical ADR.* Pearson Professional Limited, London, UK.

Younis, G., Wood, G., & Malak, M. A. A. (2008). Minimizing Construction Disputes: the Relationship between Risk Allocation and Behavioural Attitudes. *Proceedings of CIB International Conference on Building Education & Research BEAR2008, Sri Lanka*, 134–135. Retrieved from https://www.irbnet.de/daten/iconda/CIB11538.pdf (Last access: 03 December 2019).

Yousefi, V., Yakhchali, S. H., Khanzadi, M., Mehrabanfar, E., & Šaparauskas, J. (2016). Proposing a Neural Network Model to Predict Time and Cost Claims in Construction Projects. *Journal of Civil Engineering and Management*, *22*(7), 967–978. https://doi.org/10.3846/13923730.2016.1205510

APPENDICES

A. THE CONSTRUCTION PROJECT DATA QUESTIONNAIRE

This study is a part of a PhD dissertation being prepared at METU Civil Engineering Department that aims to predict dispute occurrence in construction projects, potential compensations that can be acquired out of a dispute, and the resolution method to be used.

If you kindly agree to participate in the research, a questionnaire, which is expected to take one hour, will be conducted. The questionnaire is designed to collect empirical data related to a past project where you were the authorized decision-maker related to project management strategies in technical and legal aspects.

Participation is voluntary and the participant can leave the research at any time. The questionnaire does not contain any questions that may cause personal discomfort. However, if you feel discomfort due to the questions or any other reason, you can stop participating and leave. In such a case, it is sufficient to tell the researcher that you have decided not to complete the survey. You may refuse or stop participating in the survey without any sanctions or penalties. You can stop participating during the survey or continue answering at another time.

The collected data will be kept confidential. No personal information will be disclosed to a third party. Information obtained from the participants will be evaluated collectively and used in scientific publications only.

For any questions, remarks or suggestions, please contact the researcher. Thank you for your contribution to the research.

Kind regards,

Research Assistant Murat AYHAN
Gazi University Faculty of Engineering
Civil Engineering Deparment Room No: 255
e-mail: muratayhan@gazi.edu.tr
Phone: +90-505-319-19-88

SECTION 1: INFORMATION ABOUT THE PARTICIPANT

Q1. Please indicate your occupation:

☐ Lawyer ☐ Architect ☐ Engineer

☐ Other: …………………………

Q2. Please indicate your specific role in the project:

☐ Project Manager ☐ Legal Advisor

☐ Contract Manager ☐ Claim Adviser / Specialist

☐ Project Consultant ☐ Site Manager

☐ Project Engineer ☐ Other: …………………………

Q3. Represented party in the project

☐ Owner / Employer ☐ Contractor ☐ Other: ………………

Q4. Counter party in the project (evaluated party):

☐ Owner / Employer ☐ Contractor ☐ Other: ………………

Q5. Professional experience in the construction industry: …………… year(s)

Q6. Experience in the current role: …………… year(s)

SECTION 2: INFORMATION ABOUT THE PROJECT AND CONTRACT CHARACTERISTICS

Q7. Project Location: …………………………

Q8. Cost of construction or contract value: ………………………… **$**

Q9. Start and end dates of the contract: …………………………

Q9a. If the project is not completed, planned project duration: ………… days

Q9b. If the project is completed, the real duration of the project: ……… days

Q10. Type of construction:

☐ Housing ☐ Commercial

☐ Public Service & Medical Facilities

☐ Sports & Educational & Cultural Facilities

☐ Industrial Facilities ☐ Transportation Facilities

☐ Treatment Facilities ☐ Water Supply & Reservoirs

☐ Power Plants & Lines ☐ Soil Works

☐ Other: …………………………

Q11. Type of contractor:

☐ Single Company ☐ Joint Venture ☐ Consortium

Q12. Type of employer / owner:

☐ Public ☐ Private ☐ PPP

Q13. Type of contract:

☐ Private Contracts ☐ Public procurement

☐ FIDIC Red Book ☐ FIDIC Yellow Book

☐ FIDIC Silver Book ☐ Other: …………………………

Q14. Payment method of the contract:

☐ Lump-Sum (Fixed Price) ☐ Unit Price ☐ Other :………

Q15. Project delivery system:

☐ Design-Bid-Build (DBB)

☐ Design-Build (DB)

☐ Engineer-Procure-Construct (EPC)

☐ Other: …………………………

Q16. Please rate the level of design complexity of the project:

(1: lowest level of complexity; and 5: highest level of complexity)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q17. Please rate the level of construction complexity of the project:

(1: lowest level of complexity; and 5: highest level of complexity)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

SECTION 3: INFORMATION RELATED TO CHARACTERISTICS OF PARTIES AND THEIR ORGANIZATIONAL STRUCTURES (SKILLS)

Please rate the following using the values between 1 and 5.
(1: weakest / worst; and 5: strongest / best)

Q18. Relationship between parties / individuals:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q19. Parties' previous experience with each other (i.e. level of satisfaction from each other). If there are no previous works together, then consider the reputation or credibility of the counter party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q20. Appropriateness of dispute avoidance incentives in the project:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q21. Quality of communication between parties:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q22. Working culture and skills (qualifications) of the counter party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q23. Working culture and skills (qualifications) of the represented party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q24. Response rate and communication skills of the counter party (please consider the quality of response structures):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q25. Response rate and communication skills of the represented party (please consider the quality of response structures):

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q26. Level of experience (on the project type) of the counter party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q27. Level of experience (on the project type) of the represented party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q28. Project management and coordination skills of the counter party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q29. Project management and coordination skills of the represented party:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

SECTION 4: INFORMATION RELATED TO CHANGES

Q30a. Is there any change order (variation order) in the project?

☐ Yes ☐ No

Q30b. Is there any unexpected event in the project?

☐ Yes ☐ No

Q.30c Is there any force majeure event in the project?

☐ Yes ☐ No

SECTION 5: INFORMATION ABOUT THE DISPUTE CHARACTERISTICS

Q31. Did any disputes occur during the project?

☐ Yes ☐ No

Q32. Number of disputes: …………………………

Q33. Disputant party:

☐ Owner / Employer ☐ Contractor ☐ Other: ………….

Q34. Phase of occurrence of dispute:

☐ Planning & Design & Tender & Procurement

☐ Construction

☐ Transfer & Repair & Maintenance

☐ Other: …………………………

Q35. Source of dispute:

☐ Argument on payment related to extra works due to change order(s)

☐ Argument on payment and EoT related to extra works due to change order(s)

☐ Argument on measurement & valuation of contracted work(s)

☐ Argument on EoT related extra costs

☐ Delay in site handover & possession

☐ Construction / Design defects, errors, and poor quality

☐ Contractor fails to act as a prudent merchant

☐ Delays in payments

☐ Errors or substantial change(s) in Bill of Quantities

☐ Inadequate site / soil investigation

☐ Differences in interpretation of contract clauses

☐ Other: …………………………

Q36. Is there any stoppage / suspension / interruption of works due to disputes?

☐ Yes ☐ No

Q37. Disputed amount: ………………………… $

Q38. Settled amount: ………………………… $

Q39. Success Rate: ………………………… %

Q40. Is there any EoT claim associated with the dispute?

☐ Yes ☐ No

Q41. Disputed EoT amount: ………………………… days

Q42. Settled EoT amount: ………………………… days

Q43. Success Rate: ………………………… %

SECTION 6: INFORMATION ABOUT THE DELAY(S) IN THE PROJECT

Q44a. Time extensions (total amount): …………… days

Q44b. Ratio of extensions to planned project duration: ………………… %

SECTION 7: INFORMATION ABOUT THE RESOLUTION METHOD CHARACTERISTICS

Q45. Utilized resolution method for the dispute:

- ☐ Litigation
- ☐ Arbitration
- ☐ Dispute Review Boards
- ☐ Mediation
- ☐ Senior Executive Appraisal
- ☐ Negotiation
- ☐ Other (Please indicate): …………………………

Q46. Cost of resolution with the preferred method: ………………………… $

Q47. Duration of resolution with the preferred method: …………………(years)

Q48. Please rate the level of satisfaction from the preferred resolution method:

(1: lowest level of satisfaction; and 5: highest level of satisfaction)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q49. Please order the following resolution method attributes by ranking them from 1 to 10 using the values <u>just once</u> according to their importance during method selection for the disputed case in the project.

(1: most important; and 10: least important)

<u>Resolution Method Attribute</u>	**<u>Ranking (order of importance)</u>**
Q49a. Preserve Relationships	…………………………………….
Q49b. Speed of Resolution	…………………………………….
Q49c. Cost of Resolution	…………………………………….
Q49d. Bindingness	…………………………………….
Q49e. Confidentiality	…………………………………….
Q49f. Fairness	…………………………………….
Q49g. Flexibility	…………………………………….
Q49h. Control Over Process	…………………………………….
Q49i. Reaching Creative or Remedying Solutions	…………………………………….
Q49j. Willingness	…………………………………….

SECTION 8: INFORMATION RELATED TO RESOLUTION METHOD KNOWLEDGE

Please indicate your level of knowledge on the resolution methods below:

(1: lowest level of knowledge; and 5: highest level of knowledge)

Q50. Litigation:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q51. Arbitration:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q52. Dispute Review Boards:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q53. Mediation:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q54. Senior Executive Appraisal:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Q55. Negotiation:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

SECTION 9: FINAL REMARKS

Q56. Would you consider using the "dispute occurrence prediction" model that will be established at the end of this research for decision-support?

☐ Yes ☐ No

Q57. Would you consider using the "potential compensation prediction" model that will be established at the end of this research for decision-support?

☐ Yes ☐ No

Q58. Would you consider using the "resolution method selection" model that will be established at the end of this research for decision-support?

☐ Yes ☐ No

Thank you for precious contributions to our research.

B. APPROVAL FROM APPLIED ETHICS RESEARCH CENTER FOR THE CONDUCTION OF QUESTIONNAIRES

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr

Sayı: 28620816 / 436

21 KASIM 2019

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Prof.Dr. M.Talat BİRGÖNÜL

Danışmanlığını yaptığınız **Murat AYHAN'ın "Development of Dispute Prediction and Resolution Method Selection Models for Construction Disputes"** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **416 ODTU 2019** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

Prof. Dr. Tülin GENÇÖZ

Başkan

Prof. Dr. Tolga CAN

Üye

Doç.Dr. Pınar KAYGAN

Üye

Dr. Öğr. Üyesi Ali Emre TURGUT

Üye

Dr. Öğr. Üyesi Şerife SEVİNÇ

Üye

Dr. Öğr. Üyesi Müge GÜNDÜZ

Üye

Dr. Öğr. Üyesi Süreyya Özcan KABASAKAL

Üye

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	: Ayhan, Murat
Nationality	: Turkish (TC)
Date and Place of Birth	: 2 January 1989, Konya
Phone	: +90 505 319 19 88
E-mail	: muratayhan.ce@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Civil Engineering	2012
BS	METU Civil Engineering	2009
High School	Ankara Atatürk Anadolu Lisesi	2005

WORK EXPERIENCE

Year	Place	Enrollment
2010-Present	Gazi University Civil Engineering	Research Assistant
2009-2010	Ertay Construction Company	Site Engineer

FOREIGN LANGUAGES

(1) Turkish (Native)
(2) English (Advanced)
(3) German (Limited Proficiency)

PUBLICATIONS

1. Ayhan, M., Birgönül, M. T., & Dikmen, İ. (2016). İnşaat Sözleşmelerinde Ortaya Çıkması Olası Uyuşmazlıkların Tahmin Edilmesi Üzerine Alternatif Bir Yaklaşım. *4. Proje ve Yapım Yönetimi Kongresi,* Eskişehir, Turkey, 1055-1066.

2. Ayhan, M. (2012). *A Model-Based Guidance and Vulnerability Assessment Approach for Facilities Under the Threat of Multi-Hazard Emergencies.* Master's thesis, Middle East Technical University, Ankara, Turkey.
3. Ayhan, M., Güven, G., Ergen, E., Erberik, M. A., Kurç, Ö., & Birgönül, M. T. (2012). Requirements of Indoor Navigation During Emergency Response and Disaster Management. *Proceedings of XXXVIII IAHS World Congress on Housing*, İstanbul, Turkey, 45-51.
4. Ayhan, M., Ergan, S., Güven, G., Ergen, E., Erberik, M. A., Kurç, Ö., & Birgönül, M. T. (2012). Bina İçi Yönlendirme Sistemlerinde Kullanılan En Kısa Yol Algoritmalarının Afet ve Acil Durum Yönetimi Açısından Değerlendirilmesi. *2. Proje ve Yapım Yönetimi Kongresi,* İzmir, Turkey, 329-340.
5. Güven, G., Ayhan, M., Ergen, E., Sönmez, T., Kurç, Ö., Erberik, M. A., & Birgönül, M. T. (2012). Representation of Topological Building Information for Creating Shortest Evacuation Paths During Building Emergencies. *10th International Congress on Advances in Civil Engineering,* Middle East Technical University, Ankara, Turkey.
6. Anıl, Ö., Bulut, N., & Ayhan, M. (2012). Strain Distribution Between CFRP Strip and Concrete at Strengthened RC Beam Against Shear. *Structural Engineering and Mechanics,* 41, 509-525.

REFERENCES

References available upon request.