DATA SCIENCE TECHNOLOGY SELECTION: DEVELOPMENT OF A DECISION-MAKING APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

KEREM NAZLIEL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

DECEMBER 2022

Approval of the Thesis:

DATA SCIENCE TECHNOLOGY SELECTION: DEVELOPMENT OF A DECISION-MAKING APPROACH

Submitted by KEREM NAZLIEL in partial fulfillment of the requirements for the degree **Master of Science in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics, METU** _____

Prof. Dr. Altan Koçyiğit
Head of Department, **Information Systems, METU** _____

Assoc. Prof Dr. P. Erhan Eren
Supervisor, **Information Systems, METU** _____

Dr. Kerem Kayabay
Co-Supervisor, **TÜBİTAK** _____

Examining Committee Members:

Prof Dr. Altan Koçyiğit
Information Systems, METU _____

Assoc. Prof Dr. P. Erhan Eren
Information Systems, METU _____

Asst. Prof Dr. Bilgin Avenoğlu
Software Engineering Department, Ankara University _____

Date: *29.12.2022*

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, Last name : Kerem Nazlıel

Signature : _____

ABSTRACT

DATA SCIENCE TECHNOLOGY SELECTION: DEVELOPMENT OF A DECISION-MAKING APPROACH

Nazlıel, Kerem

MSc., Department of Information Systems

Supervisor: Assoc. Prof. Dr. P. Erhan Eren

Co-Supervisor: Dr. Kerem Kayabay

January 2023, 107 pages

Developments in IT, Cloud, Analytics, and related fields have created an abundance of Data Science technologies for practitioners, developers, and organizations to use. This abundance and variety complicate the Data Science technology selection and management processes for the analytics teams. When teams select and use improper tools and technologies, they encounter problems and inefficiencies, also known as technical debt. As a remedy, this thesis proposes a systematic technology selection method considering the analytics technology selection literature and tests it on a case study. This method consists of a survey with open-ended questions to determine the requirements of a given Data Science Workflow, linkage grids to map technologies to these requirements, and multi-criteria-decision-making to rank the technologies according to practitioners' needs and preferences. This method enables decision-makers to compare the technology alternatives and select the most suitable Data Science Technology Stack. While the existing studies in this domain consider the technology selection problem in isolation and investigate a subset of technologies, the proposed method encapsulates the end-to-end Data Science Process and the entire analytics technology landscape considering the key principles for developing industrially relevant strategic technology management toolkits.

Keywords: Data Science, Technology Management, Technology Selection, Multi-Criteria Decision-Making

ÖZ

VERİ BİLİMİ TEKNOLOJİ SEÇİMİ: BİR KARAR VERME YAKLAŞIMININ GELİŞTİRİLMESİ

Nazlıel, Kerem

Yüksek Lisans, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Doç. Dr. P. Erhan Eren

Tez Eş Danışmanı: Dr. Kerem Kayabay

Ocak 2023, 107 sayfa

Bilişim Sistemleri, Bulut, Analitik ve ilgili alanlardaki gelişmeler kullanıcılar ve organizasyonlar için Veri Bilimi alanında çok sayıda teknolojik çeşitliliğe neden olmuştur. Bu çeşitlilik Veri Bilimi teknolojilerinin seçilmesini ve yönetilmesini zorlaştırmaktadır. Ekipler uygun olmayan araçları ve teknolojileri seçip kullandıklarında teknik borç olarak bilinen sorunlarla ve verimsizliklerle karşılaşırlar. Bu soruna çözüm olarak, bu tez, analitik teknoloji seçimi literatürünü göz önünde bulundurarak sistematik bir teknoloji seçim yöntemi önermekte ve bunu bir vaka çalışması üzerinde test etmektedir. Bu yöntem, belirli bir Veri Bilimi İş Akışının gereksinimlerini belirlemek için açık uçlu sorular içeren bir anketten, teknolojileri bu gereksinimlerle eşleştirmek için bağlantı tablolarından ve kullanıcıların ihtiyaç ve tercihlerine göre teknolojileri sıralamak için çok kriterli karar verme yöntemleri kullanmaktadır. Bu yöntem, karar vericilerin teknoloji alternatiflerini karşılaştırmasını ve en uygun Veri Bilimi Teknoloji Mimarisi seçmesini sağlar. Bu alandaki mevcut çalışmalar, teknoloji seçimi problemini tek başına ele alıp teknolojilerin bir alt kümesini araştırırken, önerilen yöntem endüstriyel olarak ilgili stratejik teknoloji yönetimi araç setleri geliştirmek için temel ilkelerine uyarak geliştirilmiştir ve bu yöntem uçtan uca Veri Bilimi Sürecini ve tüm analitik teknolojilerini kapsar.

Anahtar Sözcükler: Veri Bilimi, Teknoloji Yönetimi, Teknoloji Seçimi, Çok Kriterli Karar Verme

To my beloved family

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Assoc. Prof. Dr. P. Erhan Eren for his invaluable guidance, encouragement, and support throughout this research. He introduced me to this field; without his guidance, I would not have known about the research opportunities in this domain. Also, I am grateful for our analytical, fruitful, and heartfelt discussions in developing this study.

I would like to thank my co-advisor, Dr. Kerem Kayabay, for sharing his passion for research and inspiring me with his vision and dedication. I highly treasure our rich brainstorming sessions and genuine conversations. Without his mentorship and encouragement, many of this study's accomplishments would not have been possible.

I am thankful to Assoc. Prof. Dr. Ebru Gökalp and Dr. Mert Onuralp Gökalp for their help, comments, and valuable feedback from the very beginning of this study.

I would like to thank Görkem Polat for his input and insightful comments throughout the case study and for introducing me to the METU Informatics Institute Computer Vision Lab's researchers. I would like to share my most sincere thanks with Ayberk Aydın, Oğuz Hanoğlu, and Ümit Mert Çağlar for their input and contributions to the development of this study.

I would like to thank all my colleagues from METU Informatics Institute. Special thanks to Alp Bayar, Arif Ozan Kızıldağ, Burcu Koç, Ezgi Çavuş, Mert Mecit, and Serkan Özdemir for their support and camaraderie. I am grateful to all academic and administrative staff, including Ayşe Nur Özdere Yüksel, Çetin İnci, Hakan Güler, Nilüfer Şeker, and Sibel Gülnar.

I would like to thank Elif Arıkan for her continuous support through this demanding journey.

Finally, I would like to thank my parents, Bijen and Kemal Nazlıel, and my grandparents Aysel and Kenan Özgediz for their exceptional understanding, encouragement, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AHP	Analytic Hierarchy Process
COPRAS	Complex Proportional Assessment
CRISP-DM	Cross-Industry Standard Process for Data Mining
CV	Computer Vision
DIKW	Data Information Knowledge Wisdom
DF	Differentiating Factor
DL	Deep Learning
DS	Data Science
DSW	Data Science Workflow
ETL	Extract, Transform, Load
FSIR	Fuzzy superiority and Inferiority Ranking
GDM	Group Decision-Making
HPC	High-Performance Computing
IT	Information Technology
KDD	Knowledge Discovery in Databases
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MLR	Multivocal Literature Review
MLOps	Machine Learning Operations

M&A	Mergers and Acquisition
NFR	Non-Functional Requirements
NLP	Natural Language Processing
OS	Operating System
ODISEE	Organization, Data, Infrastructure, Software, Experience, Environment
PUE	Prior User Experience
R&D	Research and Development
SACAM	Software Architecture Comparison Analysis Method
SDK	Software Development Kit
SLR	Systematic Literature Review
SME	Small and Medium-sized Enterprises
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
TDSP	Team Data Science Process
TA	Technology Alternative
TM	Technology Management
TS	Technology Selection

CHAPTER 1

INTRODUCTION

1.1. Research Background & Problem Statement

Developments in cloud & edge computing, computer science, analytics, and artificial intelligence are reshaping industries all over the globe. Organizations that aim to keep up with these advancements and be competitive in the future must use data-driven decision-making in their routines and operations [1]. To do so, organizations need to apply various data science practices in their business functions.

The Data Science (DS) industry is ever-expanding, with an estimated $77 Billion by 2023 [2]. These rapid developments in the industry are creating new job roles responsible for working with various data products, including Data Analyst, Data Scientist, Data Engineer, Machine Learning (ML)/Artificial Intelligence (AI) Engineer, MLOps Engineer, Data Product Owner, and other data integrated roles [3]. Each job definition requires the use of specific models and methods and therefore requires specialized technologies, including software packets, libraries, and frameworks. In addition, data has critical importance in the choice of technology. Big Data plays a vital role in the abundance of open-source and commercial technologies [4]. The emergence of Big Data has opened the door for the development of numerous technologies. These technologies vary according to Big Data's characteristics: volume, velocity, variety, and veracity. Besides, research suggests [5] that Open-Source technologies are preferred by 37% of Data Scientists since these technologies are economical and help avoid vendor lock-in.

DS requires high-performance computing in which complex computations are at the core, thus requiring high-performing technologies both in hardware and software. There are numerous technologies utilized for different steps of data science processes. The steps taken under the DS development framework should be recognized to understand these technologies and processes. Researchers from academia and practitioners from the industry have developed process models like Knowledge Discovery in Databases (KDD) [6] and Cross-Industry Standard Process for Data Mining (CRISP-DM) [7] to unify the data processing steps of DS projects and create DS workflows. A Data Science Workflow (DSW) is an end-to-end process consisting of data processing stages where each step uses different techniques, algorithms, and models to manipulate data and create data products.

This expansion of the DS industry has resulted in an abundance of new technologies for development teams to work with. In addition, DS teams consist of numerous people

with different roles working together by using multiple technologies in an interoperable fashion. Using an improper set of technologies, teams face technical debt in which unexpected costs can occur when they try to change their technology infrastructure or the source codes to run on different platforms [8]. Therefore, integrating diverse analytic techniques into business processes is quite complex. It requires comprehensive planning, technology assessment, and forecasting to streamline these processes [9].

These studies show [10]-[13] that technology selection needs to consider various criteria, such as organizational strategy, pricing, and technological requirements. Literature includes research on Big Data Storage [10], Programming Language [11], Big Data Platform [12], and Big Data Visualization Tool [13] selection. However, these studies only focus on a single technology category with limited scope. In addition, the literature presents that there are seven key principles to developing strategic technology management toolkits [14]. The tools utilized in these studies do not adhere to these principles, thus making them limited and insufficient. Furthermore, no studies cover the selection of multiple technologies for a DSW and no approach covers the entire DS technology landscape.

1.2.Research Aim & Objectives

This interdisciplinary research aims to approach the Data Science Technology Selection problem by blending the Data Science and Technology Management domains with the following objectives:

1) To systematically identify, analyze, categorize and evaluate the technology selection approaches and criteria used in the DS domain.
2) To develop a comprehensive technology management method for technology selection covering all DS workflows and the DS technology landscape.
3) To apply the method on a DS project to validate the method's applicability and usefulness.

1.3.Significance of the Study

This study has three significant contributions to the literature:

1) This study compares the related literature on the technology selection problem for various DS technologies from a socio-technical point of view.

2) The study presents a requirements-based holistic technology selection methodology covering the end-to-end DSWs and the entire DS technology landscape covering infrastructure, platform, and application levels.

3) The study validates the proposed technology selection methodology by applying it on exemplary DSW to choose the most appropriate technologies.

1.4. Research Strategy and Organization of Thesis

The need for a comprehensive method for data science technology selection is recognized following an academia and industry collaboration project. Conducting an extensive literature review helped determine the technology selection methods used in Information Technology (IT) and DS domains. The findings are analyzed, and the proposed method is developed iteratively through regular group meetings and periodic literature reviews. Finally, the designed method is revised and validated in a case study.

The thesis consists of 6 chapters. Following this introductory chapter, Chapter 2 provides an overview of Data Science & Processes, DS Technologies, Technology Management, Decision-Making Approaches, and Technology Selection Studies. Chapter 3 identifies the research questions, the research goals, and the research approach, respectively. Chapter 4 describes the proposed method for DS technology selection. Chapter 5 tests this proposed method in a case study and presents the results. Chapter 6 summarizes the findings, contributions limitations, and future work.

CHAPTER 2

LITERATURE REVIEW

In this section, a relevant literature review is presented. First, DS and its respective technologies are briefly explained. Secondly, Technology Management and its respective subdomains are reviewed. Next, decision-making approaches, including Multi-Criteria-Decision-Making and Group-Decision-Making, are discussed. Finally, the technology selection approaches used in the analytics domain are inspected and evaluated.

2.1 Data Science

DS is one of the rapidly developing areas in the Information Technology world. It is a broad field that can be considered as an intersection point of three areas. These are mathematics and statistics, computer science, and another domain like biology, physics, and finance, as shown in Figure 1 [15]. New Vantage Partners [16] predict that 97.2% of organizations will invest in DS and AI to improve competitiveness. Thanks to its interdisciplinary nature DS is used extensively in various industries, including Retail, Medicine, Banking, Finance, Manufacturing, Construction & Transportation, Energy, and many more [17]. It enables organizations to assess, document, and keep up with key performance indicators for organizational decision-making. Companies can use DS to analyze their processes, customers, and productivity to enhance their performance and profitability [18].

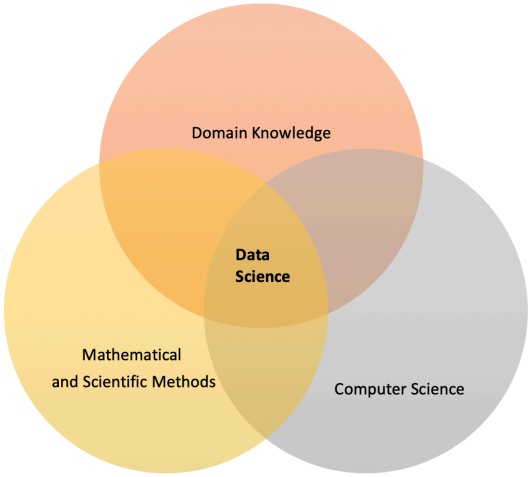


Figure 1: Data Science Venn Diagram, image based on [15]

Data alone is not helpful for organizations. Information gathered from Data helps organizations and people to make insightful decisions. DIKW Pyramid [19] states data, information, knowledge, and wisdom hierarchy. In this pyramid, Data is defined as the collection of raw facts or observations. In the second level of the pyramid, information can be found which is derived from data. Next, refining Information, Knowledge, and Wisdom levels follow. DS works as a facilitator tool to step into higher levels and make sense of the available Data.

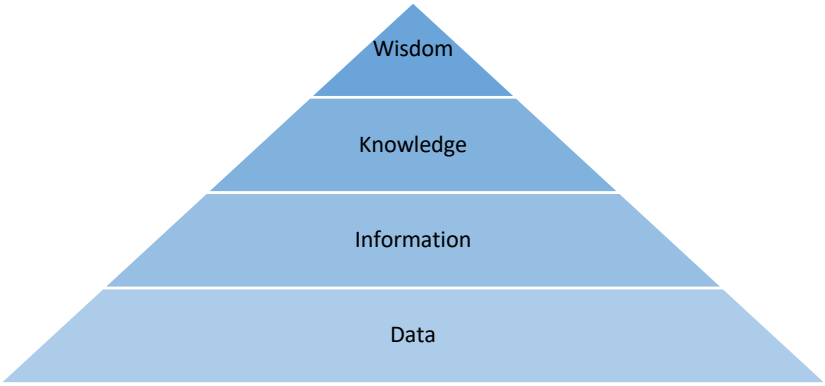


Figure 2: DIKW Pyramid, image based on [19]

Activities listed under DS Lifecycle Management Methods should be understood to understand what is considered a DS processing activity.

2.2. Data Science Lifecycle Management Methods

Researchers and practitioners from academia and industry have developed process models to plan and oversee the DS tasks to streamline data mining and analytics projects. Analytics teams frequently use KDD, CRISP-DM, and Microsoft's Team Data Science Process (TDSP) to plan DS-related workflows among these process models.

KDD [6] was developed by academic experts in the late 1990s. KDD consists of the following steps with possible back steps from the Interpretation/Evaluation step when required. The figure below shows how the KDD process model operates:

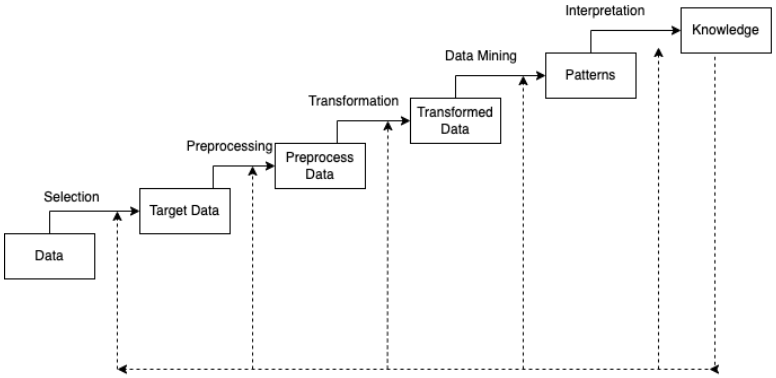


Figure 3: KDD Process, image based on [6]

Cross-Industry Standard Process for Data Mining (CRISP-DM) [7] is a similar approach that takes an iterative DS project management approach. There are loops between business understanding and data understanding, data preparation and modeling, and business understanding and evaluation. The following figure depicts the steps of CRISP-DM:

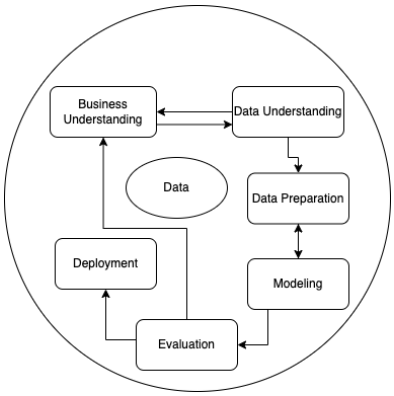


Figure 4: CRISP-DM Process, image based on [7]

Team Data Science Process (TDSP) [20] is the most recent one among the three process models. Microsoft has developed TDSP to advance the prominent and competent CRISP-DM approach. TDSP takes the deployment and modeling steps into account with more depth. KDD and CRISP-DM were developed when the variety of analytics-related workflows was limited to standard database querying or basic data mining techniques. However, TDSP is designed according to Big Data and AI-based workflows and alternative deployment environments like Cloud, Edge, and On-Premises. The following figure depicts how Microsoft TDSP operates:

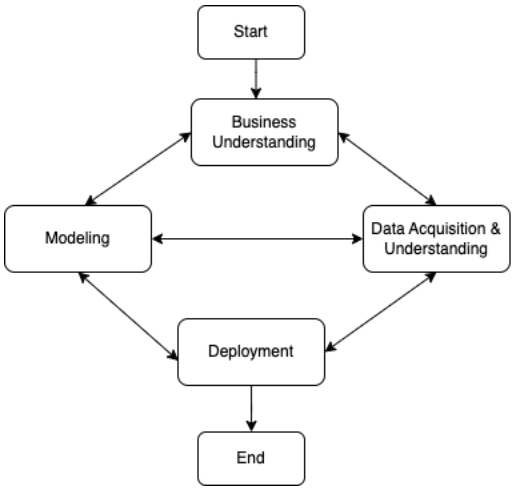


Figure 5: Team Data Science Process, image is based on [20]

DS is a broad field that collaborates with computational domains that work on data-intensive tasks. Areas that converge with DS are listed as the following [21]–[23]:

- Artificial Intelligence
- Machine Learning
- Deep Learning (DL)
- Computer Vision (CV)
- Natural Language Processing (NLP)
- Big Data
- Business Intelligence (BI)
- Data Analytics & Data Mining
- Cloud/Edge Computing
- High-Performance Computing (HPC)

A DSW consists of multiple processing steps, including modeling. This step can include AI models, Rule-Based Systems, Data Mining, and BI techniques. In addition, AI covers ML, DL, CV, and NLP fields, that provide dedicated models and algorithms used in the modeling stage. Next, Big Data has distinctive dimensions, including Volume and Velocity. Volume characteristic of Big Data maps to the storage step of the DS process, whereas Velocity characteristic maps to the Data Acquisition, Data Preparation and Data Processing. Powerful hardware resources are a must to implement state-of-the-art analytics solutions. Cloud, Edge Computing, and HPC technologies can provide the necessary infrastructure to run these solutions.

2.3. Data Science Technologies

DSWs consist of multiple data processing steps, as stated in 2.2. Each of these steps may require different technologies. Also, because of research and development progress in AI & DS, new infrastructure and software technologies are being

developed and released quickly. One other driving force for these new technologies is the large tech companies. These companies develop in-house solutions, but then to expedite their development processes, these software technologies are then transformed into open-source packages so that other software developers and specialists can contribute and utilize these technologies for their use [4]. Some examples of open-source technologies with their organizations and domains include:

Table 1: Open-Source Technologies Developed by the Industry

Organization	Technology	Domain
Twitter	Apache Storm [4][24], Apache Heron [25]	Stream Processing
LinkedIn	Apache Kafka [26]	Stream Processing
Facebook	Prophet[27]	Machine Learning
Facebook	PyTorch [28]	Deep Learning
Google	Tensorflow [29]	Deep Learning
Google	Apache Hadoop [30], Apache Beam [31]	Batch Processing
Google	Kubernetes [32]	Resource Manager & Container Platforms
Intel	OpenCV [33]	Computer Vision

Thanks to both developments in academia and industry, a multitude of technology has emerged. In the Data Science technology landscape, the technologies vary according to different aspects: First, technologies differ according to their purpose or task. These tasks can be considered storage, visualization, data exploration, prediction, image processing, and other data processing tasks. Suppose that chosen task is storage. Next, databases, file systems, or files should be considered. However, in the case of visualization, visualization libraries and dashboarding libraries should be evaluated.

Secondly, the infrastructure where these software technologies are deployed; promotes this technology variety. These technologies can be run on on-premises machines, edge devices, cloud, or High-Performance-Computing infrastructures. In addition, in the case of the cloud, the commercial technologies vary according to the cloud provider since cloud vendors like Amazon Web Services [34], Microsoft Azure [35], and Google Cloud Platform [36] offer commercial solutions for similar problems. Infrastructure includes the platforms where the source codes are run, and the hardware and the processing units inside these machines play an essential role. Industry leaders in computer hardware and semiconductors are developing dedicated processing units

for Data Science tasks. These units are Graphical Processing Units [37], Data Processing Units [38], and Tensor Processing Units [39].

Thirdly, the algorithm or the model to be used also creates a variety of technologies as well. The technologies that work on statistical tests and techniques differ from those of machine learning or deep learning. Furthermore, the developer of the technologies causes them to differentiate. The open-source community develops software technologies, whereas companies develop commercial solutions with similar capabilities. One other aspect to consider is the data. According to the definition of Big Data, it has five dimensions: Volume, Velocity, Variety, Veracity, and Value [40]. Data's characteristics also affect the storage type, such as relational, non-relational, or object database, and the processing type, such as batch or real-time processing. The following section discusses the DOTS Framework, which identifies the critical factors in Analytics processes.

2.4. DOTS Framework

Kayabay et al. [9] propose the DOTS research framework that delineates the factors that influence the use of data science and data-driven transformations. These factors are Data, Organization, Technology, and Strategy. Kayabay et al. use these factors and their subfactors as a template to prepare Data Science Roadmaps. The following table depicts the dimensions of these factors:

Table 2: DOTS Framework [9]

Data	**Organization**	**Technology**	**Strategy**
• Data Governance • Development • Data Collection • Insights • Design • Requirements Collection • Deployment	• Expertise • Organizational Culture • Project Management • Managerial Issues • Organizational Structure • Human Resources Management	• Software Management • Hardware Management • IT Governance	• Business Value • Strategic Objectives • Environment • Transformation

Socio-technical approaches may be required for streamlining analytics process. Thus, it may help using technology management approaches. The following section investigates the technology management and its activities.

2.5. Technology Management

Cetindamar et al. [41] define Technology Management (TM) as "planning, directing, controlling, and coordinating the development and implementation of technological capabilities to shape and accomplish their strategic and operational objectives." TM comprises five activities: acquisition, identification, exploitation, protection, and selection. The following part discusses the activities and their capabilities.

2.5.1. Acquisition

Technology acquisition is how an organization obtains technological resources according to its business goals. Mainly there are two ways of technology acquisition. First, the required technology is obtained through in-house product/service Research and Development (R&D) activities. The other approach is acquiring the needed technology from an external institution by purchasing, outsourcing, licensing, Mergers and Acquisition (M&A), or collaborating with organizations. P&G, one of the prominent companies in the global Fast Moving Consumer Goods industry, utilizes in-house R&D teams worldwide. Also, the company outsources, collaborates, and seeks new joint ventures to expand its technology capabilities.[42]

2.5.2. Exploitation

Cetindamar et al. [41] define technology exploitation as "the utilization of new technology or scientific developments to improve the performance of products, services or manufacturing processes." Exploitation consists of commercialization/marketing, technology transfer, and utilization. Suzlon Energy, one of the industry leaders in Wind Power, has acquired a failing wind turbine producer from Germany and exploited their opportunities by customizing the purchased company's products according to its market needs [41].

2.5.3. Identification

Technology identification activities aim to foresee the technologies and applications that can significantly affect the business. Organizations must regularly search and assess emerging and present technologies and opportunities to identify suitable technologies. Technology identification consists of several activities: technology auditing, forecasting technology and markets, identifying organizational capabilities, and documenting and disseminating the information.

2.5.4. Protection

Technology protection [42] aims to preserve the intellectual assets inside the company. There are two types of intellectual assets: legally protected and intangible assets.

Legally protected assets are considered intellectual property protected by copyright, patent, trademark, and industrial design rights. Employees can be given incentives, or senior management can promote corporate entrepreneurship to preserve intangible assets like employee know-how.

2.5.5. Selection

When more than one technology is suitable in an area, the selection should be made according to a list of criteria relevant to the domain's requirements and the organization's strategic objectives [43]. In addition, Technology Selection (TS) represents a crucial decision in the early stages of a project, which can substantially impact the economic viability [44]. Therefore, this decision requires a disciplined evaluation method to select the most appropriate, efficient, and cost-effective solution. TS is an issue in the Information Technology & Analytics domain. Still, it is also a problem in all areas where technology is at the project's backbone. The following part sheds light on the case studies regarding TS.

The literature includes studies of technology selection from various domains. Prasad & Somasekhara [45] selected telephony technology infrastructure for the Indian government by using AHP. Also, Ahmed & Ahmed [46] investigated choosing a suitable water purifier for the city of Bangladesh. So, the technology selection problem is common and is handled across different domains.

The technology selection problem is explored in the software domain as well. To begin with, Jusoh et al. [47] worked on the issue of open-source software selection. This research focuses on a selection methodology that uses AHP, a Multi-Criteria Decision-Making (MCDM) method, including reliability, usability, performance efficiency, functionality, maintainability, security, responsiveness, assurance, and empathy. Finally, a software toolkit is implemented using the selected functional and non-functional criteria and open-source software alternatives. Also, Şener et al. [48] implemented Multi-Criteria Decision-Making to choose a cloud service. Next, the following section elaborates on Multi-Criteria Decision Making and its use cases accordingly.

2.6. Multi-Criteria Decision-Making

TS is a complicated problem that requires an assessment of technology alternatives based on numerous factors. Due to this characteristic of the problem, the solution requires using Multi-Criteria Decision-Making methods, thus making it an MCDM problem. In addition, MCDM methods determine the best feasible solution according to established criteria with multiple alternatives. There are various methods and techniques used for MCDM problems. Some include Analytic Hierarchy Process (AHP) [49], TOPSIS [50], and Fuzzy methods [51]. The following segment looks into how AHP works.

2.6.1. Analytic Hierarchy Process

AHP [26], a quantitative approach, is one of the most popular MCDM methods as it provides consistency measures. It works by deriving priorities among criteria and alternatives. This process takes in the experts' opinions in numerical values and returns the highest-scoring ones among the options. Here follow the steps of AHP according to Gerdsri et al.[52]:

1) Develop the weights for criteria by
    a. Construct a pairwise comparison matrix for each criterion by assigning values between 1-9
    b. Normalize the resulting criterion matrix
    c. Average the values in each row to get the corresponding rating
    d. Calculate and check the consistency ratio
2) Calculate the weighted average rating for each decision alternative
3) Conduct steps 1 and 2 for each expert and calculate the mean scores of the weighted rating score for each decision alternative
4) Choose the one with the highest score

The figure below represents the architecture of the Analytic Hierarchy Process:

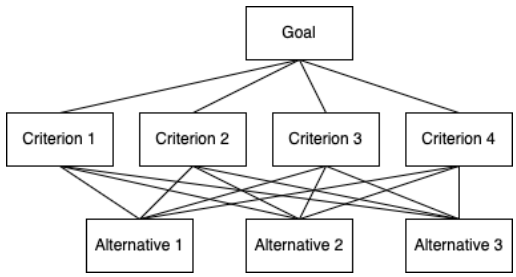


Figure 6: AHP Architecture, image is taken from [52]

Apart from quantitative expert-based methods, there are also qualitative approaches. These are also examined under Group Decision-Making in the following section.

2.7. Group Decision-Making

Group Decision-Making (GDM) techniques are used when people need to select alternative options. There is no limit to the number of people participating in group decision-making; however, it mainly varies between 2-7, but it can differ from one GDM method to another [53]. The composition of these groups can vary according to the nature of the decision. If it is a domain-specific decision, then the group should be chosen from experts, and if it is about a problem that requires input from different demographics, the group should be mixed [54]. There are numerous GDM methods: Brainstorming [55], Dialectical Inquiry [56], Nominal Group Technique [57], and

Delphi Technique. Among these methods, Delphi Technique [58] will be investigated in the following part.

2.7.1. Delphi Technique

The Delphi technique [58] is a group decision-making process used to reach a consensus by surveying experts' panels. This process consists of several rounds of questionnaires, including open-ended and multiple-choice questions in which experts share their answers. At the end of each round, responses are aggregated and shared with the group. Experts can change their responses according to the aggregated result, and after multiple rounds, the goal is to reach a consensus. Not to encounter any expert dominance, the Delphi technique should be done anonymously and individually so that any personal bias does not affect the decision-making process of individuals. This method provides different perspectives on a specific problem that can be shared democratically, and the final decision would be made as a result of the wisdom of the crowds.

Here follows the tentative roadmap for the Delphi technique on a single step of a decision-making process:

1. Step 1: A group of experts is gathered.
2. Step 2: The forecasting/ranking/selection task is identified and shared with the group.
3. Step 3: Experts make their selection/listing/forecast with their justifications. Then these are collected and summarized to provide feedback.
4. Step 4: Feedback is shared, and experts can review their answers and update their choices. Step 3-4 is iterated until a consensus is reached among experts.
5. Step 4: Consensus is reached.

In addition, AHP and Delphi Technique can be used together to decide, as in the study by Abdelanbi [11]. Considering MCDM methods, GDM methods, the DOTS Framework, and Data Science Technologies, the following part analyzes the analytics selection studies in the literature.

2.8. Analytics Technology Selection Literature

Literature includes numerous studies regarding technology selection for analytics technologies. After elaborating on these studies, section 2.9 presents evaluation criteria for these technology selection methodologies. Next, these studies are evaluated in terms of these criteria. Finally, knowledge gaps are determined in section 2.10.

Collier [58], one of the pioneers in analytics software technology selection, worked on selecting Data Mining software for a financial project among numerous options using a Weighted Sum approach. In contrast, Dakić et al. [60] utilize a decision tree-based approach to compare data mining technologies. Altalhi et al. [61] also evaluated and compared open-source data mining and knowledge discovery software suites using weighted sum scoring and checklists. Kachaoui et al. [10] have worked on selecting a Big Data Storage (Data Warehouse, Data Lakehouse) for three different organizations: an R&D Institution, a Governmental Organization, and an SME by using AHP considering technical, social, and cost-related criteria. Similarly, Klein and Gorton [62] have proposed a design assistant for NoSQL technology selection. Furthermore, a Fuzzy-based MCDM approach is used for Data Warehouse selection. Lněnička [12] selected AHP, Rouhani et al. [63] chose Fuzzy MCDM and FSIR, Ilieva [64] used TOPSIS, and Uddin et al. [65] used a Fuzzy TOPSIS approach to select a Big Data Platform among Amazon Kinesis, Apache Hadoop, Apache Spark, Cloudera, MapR, and other alternatives. A study by Singh et al. [66] selected a Big Data Platform for a Hospital in India using a Rembrandt System. In addition, Volk et al. [67] adopted a method that is a combination of AHP and SACAM to compare Big Data Reference Architectures, including Lambda, Bolster, and Kappa, for various scenarios. Next, Grandhi and Wibowo [13] presented a study on Data Visualization technology selection using Intuitionistic Fuzzy MCDM among Power BI, Tableau, Looker, and Aster Discovery. AHP is also used to select a programming language for novice programmers to use in data analytics applications by Abdelnabi [11]. Similarly, a decision tree is used for query programming language selection [68]. Regarding the Business Intelligence Systems, Büyüközkan [69] has chosen Hesitant Fuzzy Linguistic AHP with the COPRAS method to select the most suitable alternative. In contrast, a Hybrid of the AHP & TOPSIS model is used for ETL technology selection [70]. Lehman et al. [71] proposed a Big Data technology stack selection method considering data acquisition, analytics, generation, processing, and storage technologies with the help of checklists and decision trees. Lastly, Volk et al. [72] introduced a multi-staged MCDM method to select a Big Data technology stack.

2.9. Evaluation

According to Kerr et al. [14], a strategic technology management toolkit should have seven characteristics. Also, these characteristics should apply to technology selection methods and tools. These characteristics, abbreviations, and definitions are as follows:

Table 3: Strategic Technology Management Toolkit Characteristics

Characteristic	Definition
Human-centric (H)	A Human-centric approach should enable individuals to participate and engage with others to provide a solution resulting from social interaction and meaningful collaboration.
Workshop-based (W)	A Workshop-based approach should include physical or virtual meetings to have group interaction.
Neutrally facilitated (N)	The workshop where the technology management tool is applied should be facilitated neutrally and objectively.
Lightly processed (L)	The workshops should start in small, dedicated groups and iterate fast.
Modular (M)	Tools should be modular so that they can be integrated with other tools
Scalable (S)	Scalability refers to the hierarchy levels where the tool can be used within an organization, such as by executives and the product development team.
Visual (V)	Tools should have some form of visualization during their application and the outputs.

In Table 4, each technology selection approach discussed in Chapter 2.7 is summarized and analyzed in terms of strategic technology management toolkit characteristics. The character '✓' indicates fulfillment, and '*✓' indicates partial fulfillment.

Table 4: Technology Selection Approaches in the Analytics Domain

Article	Technology Category	Technology Alternatives	Selection Method	Selection Criteria	Validation	H	W	N	L	M	S	V
[59]	Data Mining Software	Knowledge Seeker, Data Mind, Clementine, Darwin	Weighted Sum	Performance, Functionality, Usability, Auxiliary Tasks	Case Study: Data Mining Technology Selection for Finance		✓		✓	✓		
[61]	Data Mining Software	AlphaMiner, CSMR, ELKI, D.ESOM, KEEL, DataMelt, GDatamine, Knime, Rattle, MiningMart, SPMF, Weka, ML-Flex, Tangara, Orange, RapidMiner, V.Waabbiti, Adam, ADAMS	Weighted Sum Scoring, Functionality Mapping	Performance, Functionality, Usability, Auxiliary Tasks	Case Study: Technology capability ranking	✓	✓		✓	✓		

Table 4 continued:

[61]	Data Mining Software	AlphaMiner, CSMR, ELKI, D.ESOM, KEEL, DataMelt, GDatamine, Knime, Rattle, MiningMart, SPMF, Weka, ML-Flex, Tangara, Orange, RapidMiner, V.Waabbiti, Adam, ADAMS	Weighted Sum Scoring, Functionality Mapping	Performance, Functionality, Usability, Auxiliary Tasks	Case Study: Technology capability ranking	✓	✓		✓	✓		
[60]	Data Mining Software	Weka, Azure ML Studio, RapidMiner H20, Apache Spark	Decision Tree	General Characteristics, Data Management, Functionality, Usability	No Validation: Mapping of each technology with functionality				✓	✓		✓

Table 4 continued:

[10]	Big Data Storage	Data Lake, Data Warehouse	AHP	Technical, Organizational, Cost related	3 case studies: Research Institution SME, Public Sector Institution	✓	✓		✓	✓	
[73]	Data Warehouse	Technologies are not shared	Fuzzy MCDM	System-related, Vendor, Organizational	Case Study: Bar Code Implementation Project for Agricultural Products in Taiwan	✓	✓		✓	✓	
[11]	Programming Language	Python, Java, R, SQL, Scala, C	AHP	Functionality, Performance, Cost, Usability	Case Study: Choosing a programming language for a novice data analyst	✓	✓		✓	✓	
[68]	Database Querying Language	Cobol, SQL, Alpha, Rabbit, Taxis	Decision Tree	Usability, Functionality, User Type,	Comparison and Categorization				✓	✓	✓

Table 4 continued:

[13]	Data Visualization	Aster Discovery, Power BI, Tableau, Looker	Intuitionistic Fuzzy MCDM	Cost, Usability, Functionality, Security	Case Study: Dashboard Selection Power and Electrical Equipment Industry	✓	✓	✓	✓
[12]	Big Data Platform	Amazon Kinesis, Apache Hadoop, Apache Spark, Apache Storm, Cloudera, GridGrain, Hortonworks, HPCC, InfoSphere Streams, MapR, Sphere"	AHP	Technical, Organizational, Cost related	3 case studies: Research Institution SME, Public Sector Institution	✓	✓	✓	✓
[64]	Big Data Platform	MapR, Hortonworks, BigInsights, Cloudera, Pivotal HD	Fuzzy MCDM, FSIR method	Performance, Functionality, Usability, Security	Case Study: Big Data Platform Selection for an Iranian IT Company	✓	✓	✓	✓

Table 4 continued:

[65]	Big Data Platform	Cloudera, MapR, Apache Hadoop	Fuzzy TOPSIS	Performance, Data Management Strategy, Usability, Cost	Case Study: Technology capability ranking	✓	✓	✓	✓
[66]	Big Data Platform	Hadoop, MapR, PIG, Hive, JAQL, Zookeeper, HBase, Cassandra, Oozie, Lucene, Avro, Mahout	Rembrandt System	Performance, Usability, Granular Manipulation, Security	Case Study: Selecting Big Data Platform for a Hospital in India	✓	✓	✓	✓
[69]	Business Intelligence Technologies	Power BI, Sisense, Qlik, Tableau	Integrated Hesitant Fuzzy Linguistic AHP & COPRAS	Technical, Functionality, Organization, Cost	Case Study: Research organization to implement BIS	✓	✓	✓	✓
[70]	ETL Technologies	Technologies are not shared	Hybrid of AHP & TOPSIS	Functionality, Vendor, Usability, Reliability, Cost	Case Study: According to chosen criteria and metrics	✓	✓	✓	✓

Table 4 continued:

[71]	Big Data Technologies	Multiple technology alternatives for multiple technology categories	Functionality Technology Mapping, Decision Tree, Compatibility Mappings	Strategy, Data, Time, Analytics capability	Case Study: Shopping Market Big Data Infrastructure Case Study		✓		✓	✓	✓*
[72]	Big Data Technologies	Multiple technology alternatives for multiple technology categories	Multi-staged MCDM, Filtering technologies according to functional requirements	Functionality, Usability, Availability, Cost, Performance, Data Management, Regulations	Case Study: Big Data Client-Server System	✓	✓		✓	✓	✓*

2.10. Analytics Technology Selection Literature

Organizations aiming to be data-driven and planning complete their digital transformation are incorporating Data Science into their business processes. Adopting Data Science principles and technologies is a complex task that requires detailed and holistic planning. One crucial aspect that should be considered in Data Science adoption is the technology dimension. Numerous technologies have emerged thanks to developments in the hardware and software industries. In addition, Data Science processes consist of multiple processing steps. At the same time, each of these processing operations may require different technologies.

Literature offers some studies regarding technology selection approaches in Data Science technologies. Firstly, no model considers the seven principles of developing a strategic technology management kit [14]. In addition, [10], [11], [12], [13], [59], [60], [61], [64], [65], [66], and [69] worked on selecting a single technology category like visualization tool and databases. These studies do not adhere to the end-to-end structure of the Data Science workflows consisting of multiple interoperable technologies. Also, these studies use several approaches, including Weighted Sum, MCDM, Decision Trees, Checklists, and Rembrandt Systems. For validation of the methods, these studies include case studies and selection criteria covering technical, financial, and organizational aspects. Furthermore, the selection literature lacks diversity since the studies mostly cover the same technology categories like Big Data Platforms and Data Mining Software Packages.

The studies [71] and [72] are the only ones that are investigating the selection of data analytics stacks consisting of multiple technologies. Furthermore, these studies' scope is limited to Big Data technologies and does not cover the entire technology categories in the DS technology landscape. [71] provides technology selection for the Acquisition, Storage, Processing, and Analytics steps of a Big Data workflow with the help of a predefined Decision Tree. Thus, this method is limited in the number and variety of technology categories and is likely to become obsolete as new alternatives emerge for each category. [72] uses multi-round MCDM to select a Big Data technology stack. The selection approach here works on selecting each technology category alone and then merging them to make a stack. However, this approach lacks to evaluate the technology stack holistically. This method may provide suitable selections for each technology category, but it may not provide the most suitable stack since it does not present a holistic approach. Lastly, these two studies do not implement a proof-of-concept deployment to test the finalized stacks.

Consequently, this study identifies the following knowledge gaps in the literature:

1. No study in the literature investigates and assesses the Data Science technology selection approaches.

2. There is no comprehensive Data Science technology selection method covering the end-to-end Data Science workflow and the entire technology landscape.

3. No technology selection method meets the seven principles of strategic technology management toolkits.

CHAPTER 3

RESEARCH APPROACH

3.1 Research Questions & Objectives

This study aims to develop a novel holistic method for technology selection for Data Science projects and workflows. Two research questions help identify the scope of this master thesis is:

1. What methods and approaches are used for analytics technology selection in the literature?

2. How can we develop a technology selection method covering the entire Data Science landscape and the end-to-end DSW?

Accordingly, this study has the following objectives considering the research questions together with the knowledge gaps in the literature:

1. To systematically identify, analyze and categorize the technology selection approaches and criteria used in the DS domain
2. To develop a comprehensive technology selection method that covers the entire DSW and DS landscape
3. To apply the method on a DSW to validate the method's applicability and usefulness.

3.2 Research Approach

The research started by realizing and understanding the problems and complexities of technology selection for analytic workflows. Next, a literature review is conducted to determine the commonly used technology selection approaches in analytics software selection and the critical factors that distinguish these technologies.

Then the research progressed based on the principles for developing strategic technology management toolkits [14]. Each study found via the literature review is mapped regarding these principles, technology, alternatives, selection criteria, and selection method. Thus, the characteristics that a novel holistic method should satisfy are determined.

Following this mapping of these studies, an exploratory research approach is followed. The research process is iterative and consists of a literature review and group meetings where the current technology selection methods and approaches are evaluated in terms of their strengths, weaknesses, and scopes. Then, according to these studies, a new method is developed incrementally. After each of these discussions, the literature is revisited in line with the shortcomings of the newly developed model. Subsequently, newly found techniques and methods are added to the list of candidate approaches. Then, a new method is constructed after a few iterations with the help of analytics lifecycle models, the studies' strengths, and alternative solutions to the studies' shortcomings. Therefore, a new technology selection method covers the end-to-end DSW, and the entire DS technology stack is developed. In addition, this method complies with all fundamental principles for developing industrially relevant strategic technology management toolkits. Finally, it is tested in a case study to validate the proposed method. The following diagram depicts the steps of the research approach.

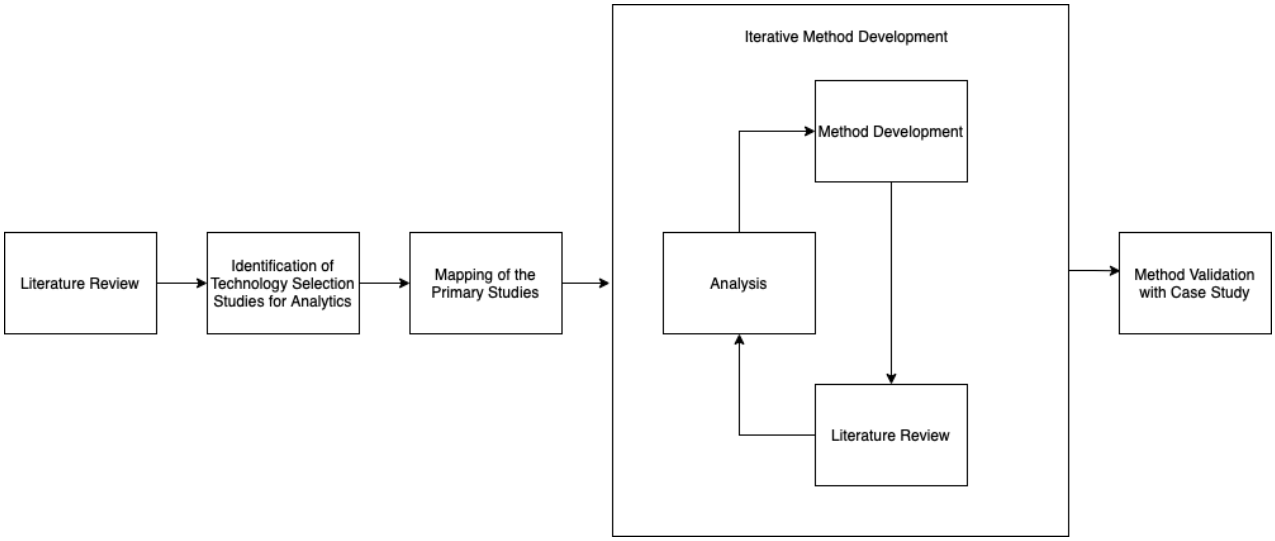


Figure 7: Research Approach

CHAPTER 4

DATA SCIENCE TECHNOLOGY SELECTION METHOD

In this chapter, the proposed Data Science Technology Selection is presented. There are two groups for implementing this method: facilitators who follow the steps and run the method and evaluators who provide input in decision-making. The proposed technology selection method (Figure 9) starts with determining the requirements for the given project. Then, the technology alternatives are mapped and filtered according to these requirements. Following this step, the remaining alternatives are ranked according to the developers' preferences, and thus, the most suitable set of alternatives is determined. Also, during the implementation of this method, many artifacts are created and used.

This chapter consists of three main parts. In section 4.1, the expected requirements and the responsibilities of the participants are described. Next, section 4.2 highlights the artifacts that are developed for this method. Finally, in section 4.3, the proposed method is explained step by step in detail. In chapter 5, this method is demonstrated on an exploratory case study.

4.1 Participants

According to the key principles of developing industrially relevant strategic technology management toolkits, a method should be neutrally facilitated; thus, it should be applied objectively. The proposed method would be run by two parties to comply with the neutrally facilitated principle [14]. These are the Facilitator and the Evaluator teams. Facilitators are the ones who will be applying the method objectively according to the input from the Evaluators, who are the domain experts. Both parties should have experts from either academia or the industry.

The requirements that the experts from both academia and the industry must have as the following:

- Bachelor's degree in Computer science, Electrical & Electronics Engineering, Information Systems, IT, or any other related field

The following are the expected attributes of an expert from academia in addition to the requirements:

- Master's or Ph.D. student with an emphasis on analytics and who is in their thesis phase
- Faculty with Ph.D. in Computer Science, Electrical & Electronics Engineering, Information Systems, IT, or any other related field

Grade student or Faculty with the following research interests:

- Data Science, AI, Machine Learning, Digital Transformation, Computational Social Science, Cloud Technologies, and related fields

The following are the expected attributes of an expert from the analytics industry in addition to the requirements:

- Master's or Ph.D. in Computer Science, Electrical & Electronics Engineering, Information Systems, IT, or any other related field would be preferred but not mandatory
- Minimum of 3 years and more expertise in the analytics domain
- Knowledge of AI & Data Science principles
- Expected job titles: Machine Learning Engineer, Deep Learning Engineer, Data Engineer, Data Scientist, Data Analyst, Analytics Manager, and other related titles

4.2 Artifacts

There are artifacts required throughout the technology selection process. These are the Technology Vision Statement, Data Science Workflow, Functional and Non-Functional Requirements, and Differentiating Features. Finally, the suitable Technology Stack would be determined at the end of the selection process. In the following part, these artifacts are explained.

4.2.1 Technology Vision Statement

The technology vision statement needs to be constructed to understand the technological, organizational, and environmental capabilities and constraints. This artifact is constructed according to the answers to the interview questions in Appendix A. Technology Vision Statement covers the following dimensions: Organization, Data Infrastructure, Software, Experience, Environment (ODISEE).

Table 5: ODISEE Dimensions

Dimension	Details
Organization	Organizational Strategy and Financial Limitations
Data	Data Characteristics and Requirements
Infrastructure	Available On-Premises, Cloud, HPC, Edge, and Other Hardware Capabilities & Resources
Software	Data Processing Steps, Techniques, and Models to be used Throughout the Project
Environment	Legislative or Environmental Limitations, Technological Trends and Drivers, Sectoral Considerations and Practices
Experience	Teams' Experience in Programming Languages, Libraries, Frameworks, Platforms, and Techniques

4.2.1.1 Technology Vision Statement Development

The development team should ask the right questions to construct the technology vision statement. However, it may be difficult and not practical to have a static question set that would help shape the technology vision statement since each DSW may have different requirements regarding the organization, data, software, infrastructure, and other dedicated requirements. Instead, a preliminary question set and guidelines to add and tailor new question sets according to the project should be more applicable and practical.

The preliminary generic question set for each ODISEE dimension is constructed as the following:

Organization:

- What is the type of organization? (Government, Research, Enterprise, SME)
- What is the planned budget for this project?
- Does the organization/team have any open-source or commercial software preferences?
- Does the organization have on-premises or cloud platform preferences?

Data:

- What is the type of dataset? (Tabular, image, time-series)

- What are the size and dimensions of the dataset?
- Does the dataset require frequent updates?
- Does the dataset have any privacy or security requirements?
- Does the dataset have any storage requirements?

Infrastructure:

- Does the team have any on-premises resources at hand? What are the specifications of those resources?
- Does the team have any cloud platform subscriptions? What are the resources and limitations?
- Does the team have any other infrastructure platform or hardware alternative at hand?

Software:

- What statistical methods, algorithms, models, and techniques are required for each step?
- What type of development platform is required?

Experience:

- What are the programming languages, frameworks, libraries, SDKs, and platforms the team has experience with?
- What are the infrastructure resources, OS, and cloud platforms the team has experience with?

Environment:

- Does the domain require any infrastructure limitations and regulations?
- Does the domain require any processing limitations and regulations?
- Does the domain require any storage limitations and regulations?

According to the answers provided by the development teams, additional follow-up questions could be constructed using the DOTS Framework and its dimensions [74]. Higher-level dimensions of the DOTS Framework can be observed in Table 6. These questions can be constructed by analyzing each dimension and its secondary dimensions respectively, considering their definitions in a checklist-based fashion, then addressing questions relevant to the business problem.

Table 6: DOTS Framework [74]

Data	**Organization**	**Technology**	**Strategy**
• Data Governance • Development • Data Collection • Insights • Design • Requirements Collection • Deployment	• Expertise • Organizational Culture • Project Management • Managerial Issues • Organizational Structure • Human Resources Management	• Software Management • Hardware Management • IT Governance	• Business Value • Strategic Objectives • Environment • Transformation

To provide an example of the secondary dimensions, Data Governance's secondary dimensions are presented [74] below:

- Data Quality
- Sensitivity
- Existing Data Integration
- Meta-data management
- Data lifecycle management
- Standards and guidelines
- Availability
- Accessibility
- Data value
- Usability

4.2.2 Data Science Workflow

A DSW describes the steps in a DS project, including querying, modeling, visualization, and other data-integrated processing steps. DS project's task and requirements should be determined to plan a DSW. Input from the Technology Vision Statement's Data, Infrastructure, and Software dimensions will shape the project's requirements. The CRISP-DM lifecycle model will act as a blueprint to synthesize the DSW for a given project. An exemplary DSW can be seen in Figure 3.

4.2.3 Cascading Interoperability Requirements

Cascading Interoperability Requirements are the requirements that emerge from the dependencies between technologies from multiple consecutive DSW steps. If a technology alternative is decided on a single step, then the interoperability of this technology should also be listed as a functional requirement for the other related DSW steps.

4.2.4 Data Science Technology Stack

A technology stack is a combination of infrastructure & software technologies an organization or company uses to develop and run a software application or product [75]. In the case of the Data Science Technology Selection Method, the artifact would be constructed at the end of the selection process.

4.2.5 Differentiating Features

Even though some technologies may satisfy the exact functional requirements, there can be some differentiating factors that are not listed under the functional requirements. These differentiating features can arise due to technical differences between the technologies and other socio-technical factors, such as the developers' long-term experience with a specific technology. Some of these differentiating features can be mapped to functional and non-functional requirements.

4.2.5.1 Functional and Non-Functional Requirements

Functional requirements identify **what** technology or software must do. In the case of non-functional requirements, non-function requirements identify **how** that technology or software package works [76]. Non-functional requirements (NFRs) specify the quality attribute of the software. ISO 25010 standards identify the system and software quality requirements [77]. These software product quality requirements are listed in the figure below:

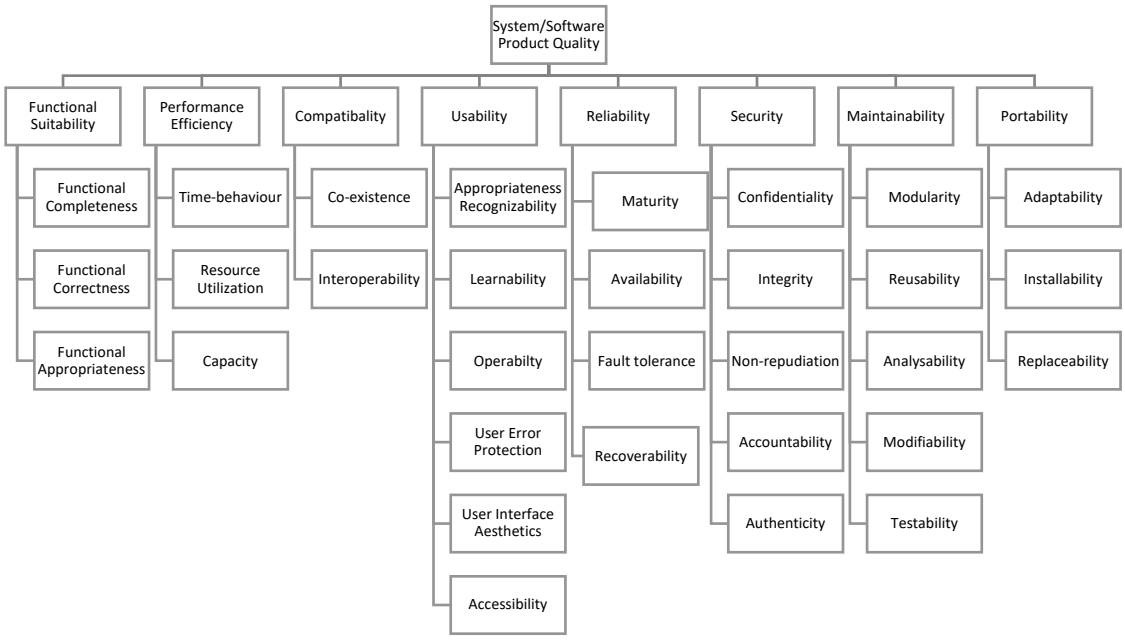


Figure 8: System and Software Quality Requirements, based on [78]

The tables below include the definitions of the non-functional requirements according to ISO 25010 [77]:

Table 7: Definitions of ISO 25010 Factors

Factor/Sub Factor	Definition
Functional Suitability	The degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.
Compatibility	The degree to which a product, system, or component can exchange information with other products, systems, or components, and/or perform its required functions while sharing the same hardware or software environment.
Reliability	The degree to which a system, product, or component performs specified functions under specified conditions for a specified period.
Security	The degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.
Maintainability	The degree of effectiveness and efficiency with which the intended maintainers can modify a product or system
Performance Efficiency	Performance relative to the amount of resources used under stated conditions.
Usability	The degree to which specified users can use a product or system to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context.
Portability	The degree of effectiveness and efficiency with which a system, product, or component can be transferred from one hardware, software, or operational or usage environment to another.

Table 8: Sub-Factors of Functional Suitability & Definitions

Functional Suitability	
Functional Completeness	The degree to which the set of functions covers all the specified tasks and user objectives.
Functional Correctness	The degree to which a product or system provides the correct results with the needed degree of precision.
Functional Appropriateness	The degree to which the functions facilitate the accomplishment of specified tasks and objectives.

Table 9: Sub-Factors of Performance Efficiency & Definitions

Performance Efficiency	
Time-Behavior	The degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements
Resource-utilization	The degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements
Capacity	The degree to which the maximum limits of a product or system parameter meet requirements

Table 10: Sub-Factors of Compatibility & Definitions

Compatibility	
Co-existence	The degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products without detrimental impact on any other product
Interoperability	The degree to which two or more systems, products, or components can exchange information and use the information that has been exchanged

Table 11: Sub-Factors of Usability & Definitions

Usability	
Appropriateness Recognizability	The degree to which users can recognize whether a product or system is appropriate for their needs
Learnability	The degree to which specified users can use a product or system to achieve specified goals of learning to use the product or the system with effectiveness, efficiency, freedom from risk, and satisfaction in a specified context of the use
Operability	The degree to which a product or system has attributes that make it easy to operate and control
User Error Protection	The degree to which a system protects users against making errors
User Interface Aesthetics	The degree to which a user interface enables pleasing and satisfying interaction for the user
Accessibility	The degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use

Table 12: Sub-Factors of Reliability & Definitions

Reliability	
Maturity	The degree to which a system, product, or component meets the needs for reliability under normal operation
Availability	The degree to which a system, product, or component is operational and accessible when required for use
Fault Tolerance	The degree to which a system, product, or component operates as intended despite the presence of hardware or software faults
Recoverability	The degree to which, in the event of an interruption or a failure, a product or system can recover the Data directly affected and re-establish the desired state of the system

Table 13: Sub-Factors of Security & Definitions

Security	
Confidentiality	The degree to which a product or system ensures that data are accessible only to those authorized to have access
Integrity	The degree to which a system, product, or component prevents unauthorized access to, or modification of, computer programs or Data
Non-repudiation	The degree to which actions or events can be proven to have taken place so that the events or actions cannot be repudiated later
Accountability	The degree to which the actions of an entity can be traced uniquely to the entity
Authenticity	The degree to which the identity of a subject or resource can be proved to be the one claimed

Table 14: Sub-Factors of Maintainability & Definitions

Maintainability	
Modularity	The degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components
Reusability	The degree to which a system, product, or component prevents unauthorized access to, or modification of, computer programs or Data
Analyzability	The degree to which actions or events can be proven to have taken place so that the events or actions cannot be repudiated later
Modifiability	The degree to which the actions of an entity can be traced uniquely to the entity
Testability	The degree to which the identity of a subject or resource can be proved to be the one claimed

Table 15: Sub-Factors of Portability & Definitions

Portability	
Adaptability	The degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software, or other operational or usage environments
Installability	The degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment
Replaceability	The degree to which a product can replace another specified software product for the same purpose in the same environment

Non-functional requirements listed above can help compare and select DS technologies. In addition, not all the NFRs listed above may be used for each step of the DSW. Each step of the DSW may require different technologies; therefore, each step of the DSW may have different functional and non-functional requirements. Also, input from all the dimensions of the Technology Vision Statement would help shape the requirements for each step of the DSW. Therefore, each workflow step may utilize different NFR sets from Table 7 to Table 15. Also, an SLR on software quality for AI-based software [78] presents the most critical NFR as Reliability, Maintainability, Security, Functional Suitability, Usability, and Performance Efficiency. Horkoff [79] has also determined the non-functional requirements for machine learning. These NFRs for ML include Accuracy & Performance, Fairness, Transparency, Security & Privacy, Testability, and Reliability. With further research, Habibullah and Horkoff [80] have surveyed analytics professionals and determined additional NFRs, including, Explainability, Safety, Fairness, Transparency, Integrity, Justifiability, Completeness, Fault Tolerance, Interpretability, Consistency, and Complexity. Also, Köhl et al. [81] support [80] by indicating the need for Explainability to be considered an NFR for analytics software.

4.3 Data Science Technology Selection Method

1. Form facilitator and evaluator teams. Steps 2-6 are carried out by both groups together.

2. Both teams construct a Technology Vision Statement according to the ODISEE interview questions.
3. Both teams plan the Data Science Workflow according to the business problem, Technology Vision Statement, CRISP-DM, or Team Data Science Process.
4. Evaluators order the Data Science Workflow steps in terms of their criticality to determine the bottlenecks and streamline the cascading requirements.
5. Both teams specify functional requirements for each step of the DSW using the Technology Vision Statement.
6. Facilitators alone map the functional requirements to the DSW steps considering their criticality order and list the most relevant and popular alternatives for each category. This mapping can be either in the shape of a grid or a checklist.
7. Facilitators filter technology alternatives that do not meet the requirements using a decision tree or a checklist. Inadequate and insufficient alternatives are eliminated.
8. Facilitators identify the differentiating factors (i.e., functional and non-functional requirements) for the technology alternatives specified in each step of the DSW using a customized multivocal research approach. In addition, cascading functional requirements are identified and added to the preceding or the following steps.
9. If the differentiating factors include additional functional requirements, both teams filter and select the relevant additional functional requirements and repeat step 7. Else, proceed with step 10 with differentiating non-functional requirements.
10. Facilitators prepare the AHP document by considering differentiating non-functional requirements for each DSW step.
11. Facilitators help evaluators rank the NFRs for each DSW step by using AHP.
12. Evaluators rank the criticality of each DSW step by using AHP.
13. Facilitators score the technology alternatives in each DSW step according to NFRs and calculate the weighted scores.
14. Facilitators construct the most suitable technology stack or stacks considering both weighted scores, the criticality of DSW steps, and interoperability.

15. If there are multiple alternatives, evaluators decide on the best data science technology stack using Delphi Method, and if there is only one available stack, the selection is finalized.

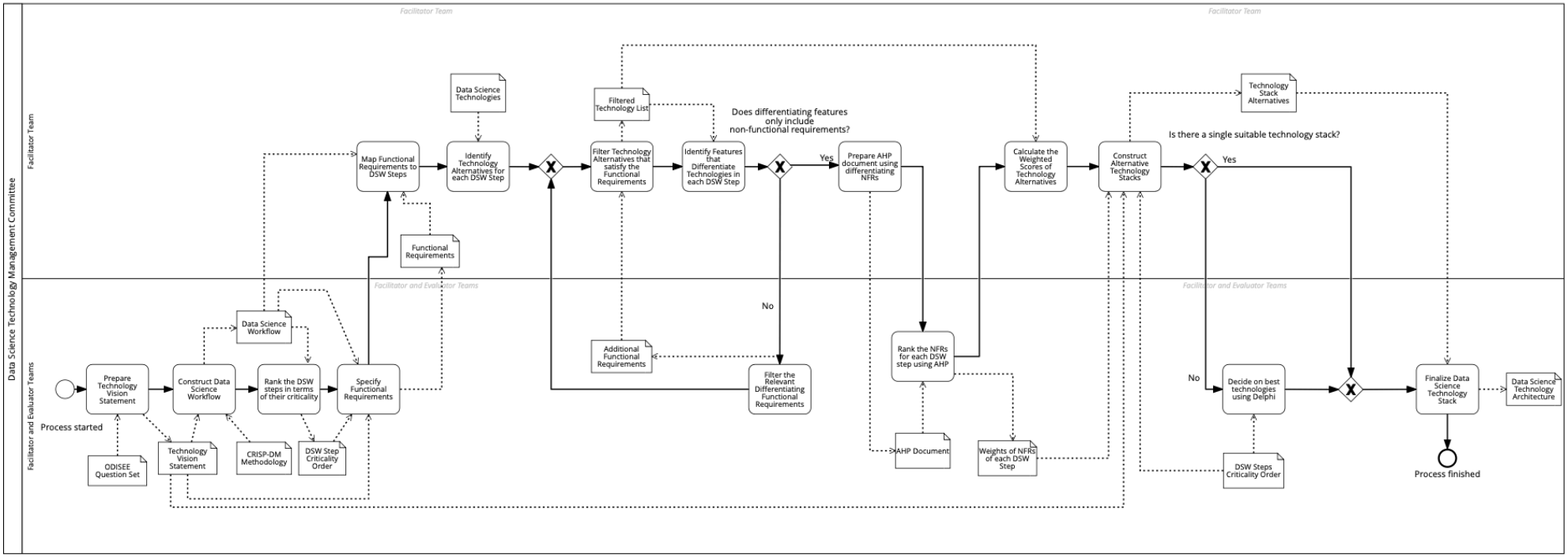


Figure 9: Data Science Technology Selection Method

4.3.1 Determining Differentiating Features for Technology Alternatives

The eighth step of the Data Science Technology Selection Method includes a customized multivocal research approach to identify the differentiating features for the technology alternatives specified in each step of the DSW. The following part proposes a systematic way to identify differentiating features for technology alternatives.

IT professionals use open-source, commercial, in-house tools and technologies for daily business routines. Currently, the industry leads cutting-edge research in software development and conversations about software development instead of the academia. Therefore, news and updates about new software technologies regarding Data Science are not mostly shared on academic platforms but instead on social platforms and blogs dedicated to software developers, including GitHub, StackOverflow, KDNuggets, HackerNews, YouTube, Medium.com, TowardsDataScience.com, and documentation web pages of tools and technologies. These pages include content from the developers and the practitioners of these technologies. Conducting Systematic Literature Reviews limits the search to academic search engines like Scopus, Web of Science, and Google Scholar and misses out on these blogs and web pages. Garousi et al. [82] argue that essential types of knowledge can be missed when systematic mapping only includes academic literature, such as test tools and cutting-edge automation approaches which are available in the industry. Therefore, conducting SLRs to attain information about these technologies may not provide enough information for technology assessment, forecasting, and selection. On the other hand, Multivocal literature reviews cover both the academic and the gray literature, including presentations, lecture notes, reports, blogs, and websites.

Çaldağ and Gökalp [83] implement an Mutivocal Literature Review (MLR) approach on software engineering domain by investigating Open Government Data. The following methodology is developed to identify differentiating features among technology alternatives by analyzing and customizing the approach designed by Çaldağ and Gökalp. The following diagram depicts the blueprint for conducting MLR to identify differentiating factors among technology alternatives.

.

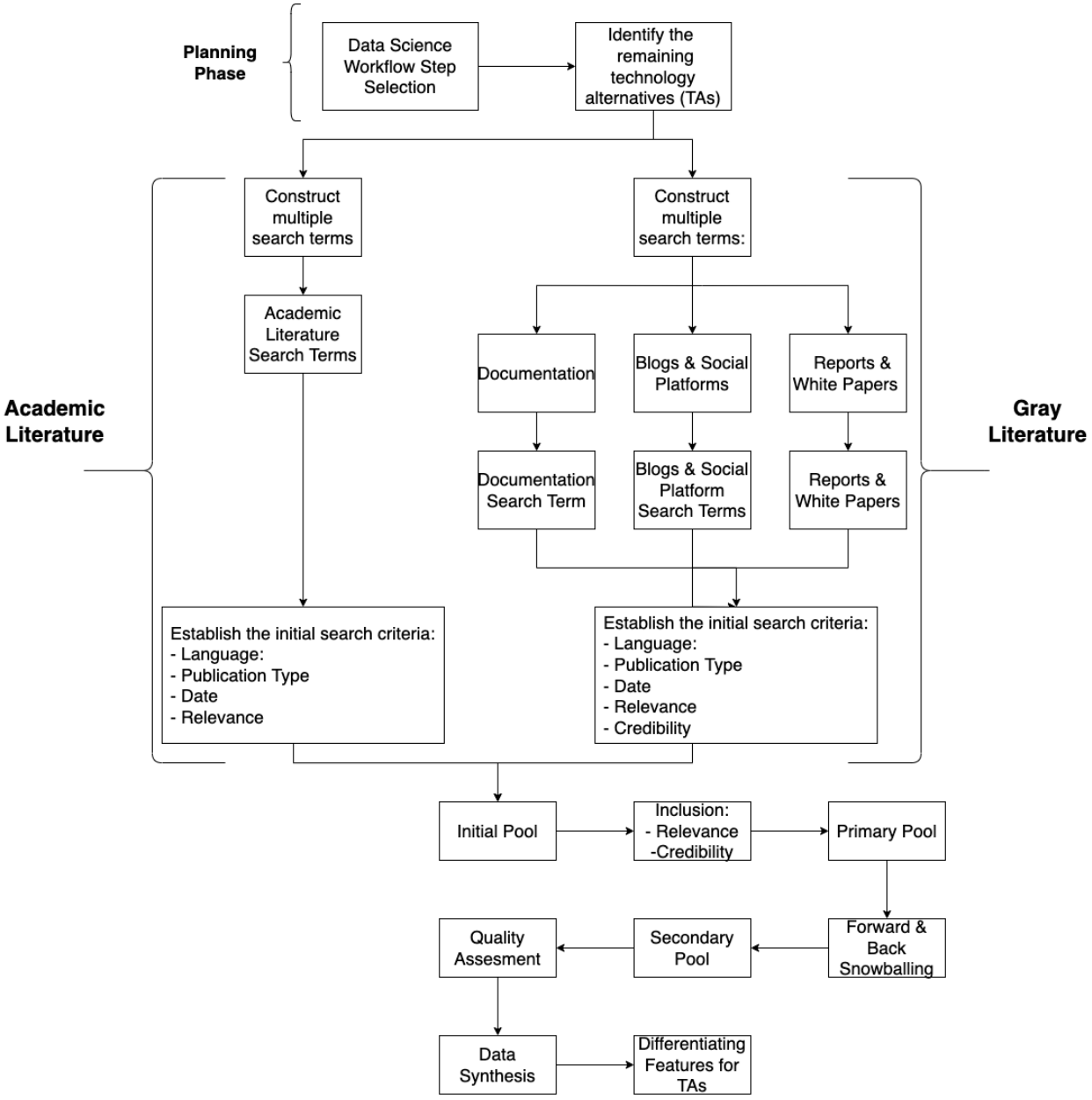


Figure 10: Customized MLR Approach to Identify Differentiating Features based on Çaldağ and Gökalp [83]

The customized MLR approach should be applied for each step of the DSW:

1. Select a Data Science Workflow step.
2. Identify the technology alternatives (TAs).
3. Construct multiple search strings for academic literature:
    a. Separate TAs by comma and construct a search string that includes all of them
    b. Group TAs in pairs and triplets and construct multiple search strings for possible combinations
4. Establish Initial Search Criteria for academic literature:
    a. Language: English
    b. Publication Type: Conference Paper, Journal Paper, Book Chapter
    c. Date: 2017-2022
    d. Relevance: Shall include description, comparison, and benchmarks of technology alternatives
5. Construct multiple search strings for gray literature:
    a. Documentation Pages:
        i. TA name + "Documentation"
    b. Blogs and Social Platforms:
        i. "Compare" + TA name 1+ "," + TA name 2 +...
        ii. "Comparison of" + TA name 1+ "," + + TA name 2 +...
        iii. TA name 1 + "vs" +TA name 2 + "vs" +…
        iv. TA Category + "comparison."
    c. Reports & White Papers:
        i. TA Category + "Report" OR "White Paper."
        ii. TA name 1 + "," + TA name 2

6. Establish Initial Search Criteria for Gray Literature:
    a. Language: English
    b. Search Engine: Google
    c. Publication Type: Blog Post, White Paper, Report, Discussion, Documentation & Guidelines
    d. Date: 2017-2022
    e. Relevance: Shall include description, comparison, and benchmarks of technology alternatives
    f. Credibility: Fact and frequency checks from multiple resources
    g. Limit: First 5 Pages of Google Search Results
7. The primary pool of resources is gathered. Filter these resources according to relevance and credibility.
8. Expand the resource pool using forward and backward snowballing.
9. Assess the quality and extract differentiating features from the documents.

CHAPTER 5

CASE STUDY & RESULTS

This section consists of two parts. The first part follows the steps of the Data Science Technology Selection Method by testing it on a Case Study and the second part provides a statistical analysis of the results of the Case Study.

5.1 Case Study

This section of this thesis includes the case study with Middle East Technical University Informatics Institute's Computer Vision Laboratory members. Their project lies in the intersection of Deep Learning and Medical Imaging. The case study will follow the steps explained in section 4.3.

5.1.1 Team Forming

Construct the Facilitator and the Evaluator teams. Both teams together will run steps between 5.1.1 and 5.1.6. The facilitator team will consist of Middle East Technical University Informatics Institute's Digital Transformation Research Lab, and the evaluators are the Computer Vision Laboratory's members.

5.1.2 Technology Vision Statement Creation

Both teams construct a Technology Vision Statement according to the ODISEE interview questions. The technology vision statement, the artifact that describes the given conditions, prerequisites, and priorities under Organization, Data, Infrastructure, Software, Experience, Environment (ODISEE), is constructed using the interview questions are explained in the following part:

<u>Organization</u>

- Researchers at METU Informatics Institute Computer Vision Laboratory conduct the research.
- METU Informatics Institute is an interdisciplinary research institute covering Information Systems, Data Informatics, Medical Informatics, and Cognitive Science.

- The METU Central IT Team advises the Informatic Institute to manage the Informatics Institute's on-premises IT resources.
- Researchers from Turkish Universities can collaborate with The Scientific and Technological Research Council of Turkey's (TUBITAK) High-Performance Computing Center (TRUBA).
- Therefore, the project is planned with an academic **goal and a limited budget.**

Data:

- The dataset includes three channeled RGB image data with 352x288 dimensions.
- Data includes sensitive patient information **that requires local and safe storage**.
- Data is labeled independently by three expert gastroenterologists.
- Data is gathered **once in a batch**, which would not require **additional data collection.**

Infrastructure:

- Group prefers to work with multiple infrastructures at a time.
- The infrastructure at the CV Lab is a workstation **with two dedicated NVIDIA GPUs with a total of 12 GB of memory.**
- The workstation has 8-core Intel processors, 2.5 GHz, and 32 GB DDR4 RAM.
- METU Informatics Institute has its **on-premises hardware resources** managed by the research assistants at the Informatics Institute.
- The on-premises server provided by the institute **has 128 GB of ram with 32 CPUs and a dedicated NVIDIA GPU with 16 GB of memory.**
- The infrastructure at the Informatics Institute provides a 100 Mbps ethernet connection.
- TRUBA's infrastructure includes a CUDA dedicated cluster named Barbun-CUDA consisting of **24 servers, each having 40 CPU cores & 2 NVIDIA P100 GPUs.**
    - Akya-Cuda: 24 nodes; each node has 40 CPU cores and 4 V100 GPUs.
    - Palamut-Cuda: 9 nodes, each node with 128 cores and 8 A100 80 GB GPUs.

Software:

- Due to their infrastructure preference, the group requires a software stack that can work with multiple infrastructures with different characteristics.
- The research group prefers to use **open-source or free software technologies.**
- The resources at the CV Lab use Ubuntu OS.
- The on-premises resources at the Informatics Institute are virtualized using Proxmox VE and can provide Windows 10 OS and Linux OS-based virtual machines.

- TRUBA's resources are provided over a CentOS/RHEL8 OS. To use these resources, researchers need to use Slurm to submit jobs.

Experience:

- The research team consists of experienced computer vision researchers.
- The research team is experienced in using Python programming language and PyTorch library.
- The research team would prefer technologies that they have prior experience using.
- The research team has experience in using TUBITAK's TRUBA HPC platform.
- The Informatics Institute IT team has experience virtualizing and providing the necessary virtual machines with dedicated hardware resources.

5.1.3 Data Science Workflow Design

The Data Science Workflow is designed by both teams according to Technology Vision Statement, and CRISP-DM can be found in the figure below.

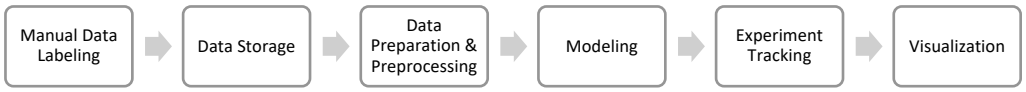


Figure 11: The Data Science Workflow

5.1.4 Criticality Ranking of Data Science Workflow Steps

Evaluators order the Data Science Workflow steps in terms of their criticality to determine the bottlenecks and streamline the cascading requirements.

1. Modeling
2. Experiment Tracking
3. Data Preparation & Preprocessing
4. Visualization
5. Data Storage

5.1.5 Specifying Functional Requirements

Both teams specify the functional requirements for each step of the DSW according to DSW and Technology Vision Statement. The functional requirements for each step of the Data Science Workflow are listed in the following part.

Manual Data Labeling:

- Experts from the gastroenterology domain should manually label image data.

Data Storage:

- Should store three-channel RGB Images with 352x88 dimensions and the labels of these images.
- Should support 10GB of Image Data.
- Should be stored locally in the Informatics Institute.
- Should be either open-source or provided by the Operating System.
- Should support either PNG or JPEG.

Data Preparation & Data Preprocessing:

- Should be based on Python Programming Language.
- Should be interoperable with the chosen data storage technology.
- Should be able to compare label lists from three experts and discard the images and labels that do not match.
- Should be able split image data into train, test, and validation sets.
- Should be able to resize and flatten and modify image data.

Model Development

- Should be based on Python Programming Language.
- Should support the following deep learning models and algorithms: Resnet18, Resnet50, DensetNet121, and Inception -v3.
- Should support GPU-based processing.
- Models should be able to support the following data formats: ndarray, NumPy array, and pillow.
- Should be interoperable with the chosen data preparation & data preprocessing technology.
- The developed models should be containerized to work on either infrastructure choice.

Experiment Tracking:

- Should support Python Programming Language.
- Should provide confusion matrix metrics calculations: precision, recall, f1-score, losses accuracy over multiple experiments.
- Should provide Computer Vision Metrics calculations: Intersection Over Union, Mean Average Precision, PSNR, SSIM over multiple experiments.
- The developed models should be reproducible when tested by other researchers.

Visualization:

- Should provide bar, line, histogram, boxplot.
- Should support the data structures chosen at the Model Development step.
- Should be able to work with Model Evaluation packages to visualize results with dashboards.

5.1.6 Mapping & Filtering of Technology Alternatives

Facilitators alone map the functional requirements to the DSW steps considering their criticality order and list the most relevant and popular alternatives for each category. Then, facilitators filter technology alternatives that do not meet the requirements using a decision tree or a checklist. Inadequate and insufficient alternatives are eliminated.

Table 16:Technology Mapping for Modeling Technologies

Infrastructure For Model Development & Evaluation:	TRUBA HPC Clusters OR METU CV Lab Workstation OR II On-Premise Servers										
		Alternatives									
Requirement Group	Requirement	PyTorch w/ Timm	PyTorch Lightning	Keras on Tensorflow	Tensorflow	Caffe	Sklearn	Apache Mxnet w/ GluonCV	Fast.ai on PyTorch	MATLAB w/ Mathworks Deep Learning Toolbox	Theano
Organization	Shall be an Open-Source Technology	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	Shall follow and adapt state-of-the-art Deep Learning methods and techniques	✓	✓	✓	✓	✓	×	✓	✓	×	×
	Shall have active development and support	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
Data	Shall process RGB image data with 352x288 dimensions	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Shall support the following Data Structures: ndarray, NumPy array, pillow	✓	✓	✓	✓	✓	✓	×	✓	×	✓
	Shall process 100 GB of Image Data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Infrastructure	Shall run on TRUBA HPC & Support CentOS	✓	✓	✓	✓	✓	×	✓	✓	✓	✓
	Shall support CUDA on Truba	✓	✓	✓	✓	×	×	×	✓	×	✓
Software	Shall support the following DL Models: Resnet18, Resnet50, DensetNet121, Inception -v3	✓	✓	✓	✓	✓	×	✓	✓	✓	×
	Shall provide Confusion Matrix Statistics: Precision, Recall, F1-Score, QWK, MAE, Accuracy	✓	✓	✓	✓	×	✓	✓	✓	✓	×
	Shall include state-of-the-art training strategies and techniques	✓	✓	✓	✓	×	✓	✓	✓	✓	×
	Shall be able to save the model architecture and its parameters	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Experience	Shall support Python Programming Language	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
Result		✓	✓	✓	✓	×	×	×	✓	×	×

Table 17: Technology Mapping for Experiment Tracking Technologies

Infrastructure For This Step:	METU CV Lab Workstation						
		Alternatives					
Criteria Group	Details	TensorBoard	Sacred + Omniboard	Guild AI	Weights & Biases	Mlflow	ClearML
Infrastructure	Shall support Ubuntu OS	✓	✓	✓	✓	✓	✓
Software	Shall support Hyperparameter Tracking	✓	×	✓	✓	✓	✓
	Shall support Model Metrics Tracking	✓	✓	✓	✓	✓	✓
	Shall have Web-based UI	✓	✓	✓	✓	✓	✓
	Shall support PyTorch	✓	✓	✓	✓	✓	✓
	Shall support TensorFlow & Keras	✓	✓	✓	✓	✓	✓
	Shall support fast.ai	✓	×	×	✓	✓	×
	Shall support 100+ experiements	✓	✓	✓	✓	✓	✓
Experience	Shall Support Python Programming Language	✓	✓	✓	✓	✓	✓
Result		✓	✓	✓	✓	✓	✓

Table 18: Technology Mapping for Data Preprocessing Technologies

Infrastructure For This Step:	METU CV Lab Workstation						
		Alternatives					
Criteria Group	Details	Pillow	Scikit-image	OpenCV	Tf.keras. preprocessing.image	PyTorch + TorchVision	MATLAB w/ Mathworks Deep Learning Toolbox
Organization	Shall be an Open-Source Technology	✓	✓	✓	✓	✓	✕
Data	Shall process RGB image data with 352x288 dimensions	✓	✓	✓	✓	✓	✓
	Shall process PNG & JPEG, Bitmap	✓	✓	✓	✓	✓	✓
	Shall process 10 GB Storage Space	✓	✓	✓	✓	✓	✓
Infrastructure	Shall support Ubuntu OS	✓	✓	✓	✓	✓	✓
Software	Shall resize, Flatten, Transform, Modify, Normalize, Augment, Censoring, Blacking Out and other image transformation techniques	✓	✓	✓	✓	✓	✓
	Shall split Image Data into Train, Validation, Test Sets	✓	✓	✓	✓	✓	✓
	Shall support PyTorch	✓	✓	✓	✕	✓	✕
	Shall support TensorFlow & Keras	✓	✓	✓	✓	✕	✕
	Shall support fast.ai	✓	✓	✓	✓	✓	✕
Experience	Shall support Python Programming Language	✓	✓	✓	✓	✓	✕
Result		✓	✓	✓	✓	✓	✕

Table 19: Technology Mapping for Data Storage Technologies

Infrastructure For This Step:	METU CV Lab Workstation			
		Alternatives		
Criteria Group	Details	Linux File System	MySQL (RDBMS)	MongoDB (NoSQL)
Organization	Shall be an Open-Source Technology	✓	✓	✓
	Shall support on Premise Storage	✓	✓	✓
Data	Shall be able to store RGB image data with 352x288 dimensions	✓	✓	✓
	Shall be able to store PNG, JPEG, Bitmap	✓	✓	✓
	Shall provide 100 GB Storage Space	✓	✓	✓
Infrastructure	Shall support Ubuntu OS	✓	✓	✓
Software	Shall support Data Transfer to CentOS Machines	✓	✓	✓
	Shall support single code data loading using Python	✓	✓	✓
Result		✓	✓	✓

Table 20: Technology Mapping for Visualization & Dashboarding Technologies

Infrastructure For This Step:	METU CV Lab Workstation					
Criteria Group	Details	Seaborn + Streamlit	Dash + Plotly	ggplot	Streamlit + Matplotlib	Gradio + MatPlotlib + Plotly
Organization	Shall be an Open-Source Technology	✓	✓	✓	✓	✓
Data	ShallSupport the following Data Structures: ndarray, NumPy array, pillow, Pandas	✓	✓	✓	✓	✓
Infrastructure	Shall supports Ubuntu OS	✓	✓	✓	✓	✓
Software	Shall support Line Charts, Histograms, Bar plots, Scatter plots, Box plots	✓	✓	✓	✓	✓
	Shall support Heatmaps	✓	✓	✓	✓	✓
	Shall support customizable Web based Dasboard Apps	✓	✓	✗	✓	✓
Experience	Shall Support Python Programming Language	✓	✓	✗	✓	✓
Result		✓	✓	✗	✓	✓

Table 16 depicts the technology mapping for the Model Development step of the Data Science Workflow. According to evaluators, the modeling table is first constructed since it is the most critical workflow step. Alternatives are eliminated since they do not suffice the functional requirements. Alternatives colored in green are available for both infrastructure options, whereas the ones colored in yellow can only run on local workstations. Tables 16-20 are prepared, placing the Model Development step as the center, considering the available options in Table 16 as cascading functional requirements for the other steps in the Data Science Workflow.

5.1.7 Identifying the Differentiating Features

The following part includes an exemplary customized MLR for determining the differentiating features among technology alternatives, as explained in 4.3.1. The customized MLR would be run for the remaining technology alternatives in the Modeling step of the DSW.

1. Differentiating features for the Modeling step would be selected.
2. The alternatives to be compared: PyTorch, Keras w/ Tensorflow and Fast.ai
3. Construct multiple search strings for academic literature on Scopus and Web of Science:
    a. Separate TAs by comma and construct a search string that includes all of them:
    b. TA Category
        i. “Deep Learning Libraries/Frameworks”
    c. Group TAs in pairs and triplets and construct multiple search strings for possible combinations
        i. “PyTorch” and (“Tensorflow” OR “Keras”) and “Fastai”
        ii. “PyTorch“ and (“Tensorflow” OR “Keras”)
        iii. “PyTorch” and “Fastai”
        iv. (“Tensorflow” OR “Keras”) and “Fastai”
4. Establish Initial Search Criteria for academic literature:
    a. Language: English
    b. Publication Type: Conference Paper, Journal Paper, Book Chapter

c. Date: 2017-2022

d. Relevance: Shall include description, comparison, and benchmarks of relevant technology alternatives

5. Construct multiple search strings for gray literature:

    a. Documentation Pages:

        i. TA name + “Documentation”

            1. “PyTorch Documentation”

            2. “Tensorflow Documentation”

            3. “Keras Documentation”

            4. “Fast.ai Documentation”

    b. Blogs and Social Platforms:

        i. “Compare” + TA name 1+ “,” + TA name 2 +...

            1. “Compare PyTorch, Keras, and Fast.ai”

            2. “Compare PyTorch, Tensorflow, and Fast.ai”

        ii. “Comparison of” + TA name 1+ “,” + + TA name 2 +...

            1. “Comparison of PyTorch, Keras, Fast.ai”

            2. “Comparison of PyTorch, TensorFlow, Fast.ai”

        iii. TA name 1 + “vs” +TA name 2 + “vs” +…

            1. “PyTorch vs. Keras vs Fast.ai”

            2. “PyTorch vs. Tensorflow vs. Fast.ai”

        iv. TA Category + “comparison”

            1. “Deep Learning Libraries Comparison”

            2. “Deep Learning Frameworks Comparison”

    c. Reports & White Papers:

        i. TA Category + “Report” OR “White Paper”

1. “Deep Learning Libraries” AND (“Report” OR “White Paper”)

6. Establish Initial Search Criteria for Gray Literature:

    a. Language: English

    b. Search Engine: Google

    c. Publication Type: Blog Post, White Paper, Report, Discussion, Documentation & Guidelines

    d. Date: 2017-2022

    e. Relevance: Shall include description, comparison, and benchmarks of technology alternatives.

    f. Credibility: User votes, credible websites and fact checks from multiple sources, and documentation

    g. Limit: First 5 Pages of Google Search Results

7. The primary pool of resources is gathered. Filter these resources according to relevance and credibility.

    a. Academic Resources

Table 21: Academic Literature Document Statistics

	Initial Search	First Pass	Second Pass
Scopus	1958	50	40
Web of Science	12568	34	27
Total Number of Unique Documents			46

b. Gray Literature (Google Search 14.09.2022)

Table 22: Gray Literature Document Statistics

Document Type	Count
Blog Posts	22
Master Thesis/ Report/White Paper	1
Official Documentation	4

8. Assess the quality and extract differentiating features from the documents.

The academic literature covers studies that compare technologies by making performance tests and includes comparisons of these technologies on different dimensions. The following table depicts the differentiating factors found in the academic literature.

Table 23: Differentiating Factors according to Academic Literature

Differentiating Factor	Number of Occurrences	References
Performance & Speed	21	[84]–[94] [95]–[104]
API Language	9	[86], [88], [91], [97]–[99], [105]–[107]
Multi-GPU Support	7	[86], [91], [96], [98], [100], [103], [106]
Platform	6	[86], [88], [98] [99] [106] [108]
Parallelizing Technique	6	[86], [91], [97], [98], [100], [106]
Community &Documentation	6	[88], [96], [97], [103], [104], [109]
Core Language	5	[86], [91], [98], [100], [106]
Pretrained Models / Model Availability	4	[86], [96], [98], [106]
Popularity	4	[88], [97], [100], [103]
User Friendliness	4	[88], [97], [103], [104]
Computation Graph Type	3	[88], [97], [103]
License	2	[88], [98]
Domain (Academic/Industrial)	1	[88]
Activation Function Diversity	1	[110]
Loss Function Diversity	1	[110]
Optimizer Diversity	1	[110]

Table 24: Differentiating Factors according to Gray Literature

Differentiating Factor	Number of Occurrences	References
Debugging	9	[111]–[119]
Performance & Speed	8	[112]–[118], [120]
User Friendliness	8	[112] [113]–[119]
Level of API	7	[112]–[118]
Popularity	6	[113]–[115], [117]–[119]
Visualization	4	[111], [112], [118], [119]
Dataset Volume	4	[112], [114], [115], [117]
Community& Documentation	4	[112] [116], [118], [119]
Domain (Academic/Industrial)	4	[115] [111], [119], [121]
Computational Graph	4	[111], [112], [116], [119]
Core Language	3	[112], [114], [115]
Deployment	3	[111], [119], [121]
Pretrained Models / Model Availability	2	[114], [121]
API Language	1	[118]
License	1	[118]
Data Parallelism	1	[111]
Ecosystem	1	[121]

Tables 23 and 24 include the differentiating factors from the Academic and the Gray Literature. These factors consist of both functional and non-functional requirements. Some of these functional requirements match the requirements identified in the technology vision statement. These are:

- API Language: Python
- Multi-GPU Support & Parallelizing Technique
- Pretrained Model Availability
- Platform: Ubuntu OS

Also, some differentiating factors can map to new functional requirements:

- Computational Graph Type
- Core Language
- Deployment
- License
- Data Parallelism
- Visualization
- Activation Function, Loss Function, Optimizer Diversity
- Dataset Volume

After these factors are presented to the evaluator team, they have stated that these factors are not crucial for their selection or considered in other DSW steps. Moreover, these factors are not added to the list of functional requirements for this step of the DSW. There are some differentiating factors in terms of how these technologies operate. These are the non-functional requirements.

- Community & Documentation
- Performance
- Popularity
- User Friendliness
- Domain (Academic/Industrial)

- Debugging
- Level of API
- Ecosystem

Some of these NFRs are correlated and can be categorized into groups and mapped to ISO 25010 standards to finalize the criteria for quantitative evaluation:

- Popularity & Ecosystem ➔ Maturity
- Level of API & Debugging & User Friendliness ➔ Operability
- Performance & Speed ➔ Performance Efficiency
- Community & Documentation will remain the same.
- Domain (Academic/Industrial)➔ Prior User Experience

Also, according to the Experience dimension of the technology vision statement, Prior User Experience is a critical criterion for the evaluators. Therefore, Prior User Experience will also be added to the criteria list. The steps mentioned above are implemented for each step of the Data Science Workflow, and Table 26 is constructed.

Table 25: Differentiating Features Chosen for the Data Science Workflow

Data Storage	Data Preprocessing	Modeling	Experiment Tracking	Visualization
• Scalability • Community & Documentation • Operability • Fault Tolerance • Recoverability • Prior User Experience	• Performance Efficiency • Community & Documentation • Operability • Prior User Experience	• Performance Efficiency • Learnability • Operability • Maturity • Prior User Experience	• Community & Documentation • Installability • Dashboard Customization • User Interface Aesthetics • Operability • Maturity • Prior User Experience	• User Interface Aesthetics • Community & Documentation • Operability • Plot Variety • Prior User Experience

5.1.8 AHP Document Construction

Facilitators prepare the AHP document by considering the differentiating non-functional requirements for each DSW step The AHP document prepared for the Modeling step can be seen below.

Table 26: AHP Template Customized for the Modeling Step

Analytic Hierarchy Template: n=	5	Criteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Modeling Criteria

Pairwise Comparison Matrix

	Community & Docun	Maturity	Prior User Experien	Customizability	Performance Efficie
Community & Documentation	1	1	1	1	1
Maturity	1	1	1	1	1
Prior User Experience	1	1	1	1	1
Customizability	1	1	1	1	1
Performance Efficiency	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1

5.1.9 Ranking Non-Functional Requirements.

Facilitators help evaluators rank the NFRs for each DSW step by using AHP. The tables below depict the calculated weights of NFRs using AHP according to each step of the Data Science Workflow.

Table 27: Weights of Data Storage NFRs

Data Storage						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Performance Efficiency	0.126	0.488	0.257	0.172	0.261	0.261
Operability	0.502	0.206	0.150	0.320	0.295	0.295
Fault Tolerance	0.053	0.122	0.045	0.030	0.063	0.063
Community & Documentation	0.081	0.097	0.290	0.329	0.199	0.200
Scalability	0.042	0.050	0.069	0.100	0.065	0.065
Prior User Experience	0.195	0.030	0.188	0.050	0.116	0.116

Table 28: Weights of Data Preprocessing NFRs

Data Preprocessing						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Performance Efficiency	0.143	0.454	0.143	0.055	0.199	0.198
Community & Documentation	0.293	0.236	0.462	0.234	0.306	0.306
Operability	0.506	0.239	0.251	0.603	0.400	0.399
Prior User Experience	0.058	0.079	0.143	0.108	0.097	0.097

Table 29: Weights of Modeling NFRs

Modeling						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Maturity	0.147	0.462	0.081	0.382	0.268	0.268
Community & Documentation	0.370	0.210	0.598	0.378	0.389	0.389
Customizability	0.330	0.178	0.131	0.097	0.184	0.184
Performance Efficiency	0.053	0.103	0.060	0.043	0.065	0.065
Prior User Experience	0.096	0.046	0.131	0.100	0.093	0.093

Table 30: Weights of the Experiment Tracking NFRs

Experiment Tracking						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Dashboard Customization	0.100	0.408	0.045	0.043	0.149	0.149
Maturity	0.390	0.157	0.222	0.407	0.294	0.294
Community & Doc.	0.145	0.152	0.222	0.084	0.151	0.151
Operability	0.075	0.093	0.119	0.175	0.116	0.116
Installability	0.043	0.093	0.117	0.204	0.114	0.114
User Interface Aesthetics	0.034	0.056	0.051	0.033	0.044	0.044
Prior User Experience	0.213	0.040	0.224	0.054	0.133	0.133

Table 31: Weights of Visualization & Dashboarding NFRs

Visualization & Dashboarding						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Community & Documentation	0.273	0.364	0.448	0.122	0.302	0.302
Operability	0.510	0.325	0.142	0.497	0.369	0.369
User Interface Aesthetics	0.063	0.149	0.071	0.042	0.081	0.081
Customizability	0.044	0.100	0.083	0.065	0.073	0.073
Prior User Experience	0.109	0.063	0.256	0.274	0.176	0.176

5.1.10 Ranking Criticality of Data Science Workflow Steps

Table 32: Criticality Weights of the Data Science Workflow Steps

Criticality of the Data Science Workflow Steps						
Criteria	Domain Experts				Average Weights	Normalized Weights
	Expert 1	Expert 2	Expert 3	Expert 4		
Data Storage	0.052	0.051	0.056	0.055	0.054	0.054
Data Preprocessing	0.136	0.134	0.106	0.123	0.125	0.125
Modeling	0.601	0.600	0.502	0.555	0.565	0.565
Experiment Tracking	0.136	0.141	0.266	0.190	0.183	0.183
Visualization & Dashboarding	0.076	0.074	0.070	0.077	0.074	0.074

According to Table 32, Data Science Workflow steps are ranked according to their criticality for the project as the following:

1. Modeling
2. Experiment Tracking
3. Data Preprocessing
4. Visualization & Dashboarding
5. Data Storage

5.1.11 Scoring Technology Alternatives

Facilitators score the technology alternatives in each DSW step according to the NFRs and calculate the weighted scores. The Prior User Experience score is calculated according to inputs from the experts. The following part consists of Prior User Experience Scores for each step of the Data Science Workflow, considering multiple alternatives.

Table 33: Prior User Experience Scores for Data Storage

Prior User Experience in Data Preprocessing Technologies					
	Expert 1	Expert 2	Expert 3	Expert 4	Average Score
Linux File System	5	10	5	10	7.50
MySQL (RDBMS)	2	5	1	3	2.75
MongoDB (NoSQL)	1	2	3	0	1.50

Table 34: Prior User Experience Scores for Data Preprocessing

Prior User Experience in Data Preprocessing Technologies					
	Expert 1	Expert 2	Expert 3	Expert 4	Average Score
Pillow	6	7	8	8	7.25
Scikit-image	6	6	7	0	4.75
OpenCV	7	9	9	0	6.25
Tf.keras.preprocessing.image	1	2	4	0	1.75
TorchVision	7	7	10	6	7.50

Table 35: Prior User Experience Scores for Modeling

Prior User Experience in Modeling Technologies					
	Expert 1	Expert 2	Expert 3	Expert 4	Average Score
PyTorch	8	9	10	8	8.75
PyTorch Lightning	0	1	10	0	2.75
Keras on Tensorflow	4	4	7	5	5.00
Fast.ai on PyTorch	0	1	2	3	1.50
Tensorflow	2	3	7	5	4.25

Table 36: Prior User Experience Scores for Experiment Tracking

Prior User Experience in Experiment Tracking Technologies					
	Expert 1	Expert 2	Expert 3	Expert 4	Average Score
Tensorboard	8	6	8	3	6.25
Guild.ai	0	0	1	0	0.25
Weights & Biases	8	9	10	9	9.00
MLFlow	0	1	2	0	0.75
ClearML	0	1	0	0	0.25

Table 37: Prior User Experience Scores for Dashboarding & Visualization

Prior User Experience in Dashboarding & Visualization Technologies					
	Expert 1	Expert 2	Expert 3	Expert 4	Average Score
Seaborn	4	7	3	6	5.00
Plotly	6	6	0	0	3.00
Matplotlib	8	9	7	6	7.50
Streamlit	9	1	2	0	3.00
Gradio	0	0	3	0	0.75
Dash	0	0	0	1	0.25

The following tables show the weighted scores for each step of the Data Science Workflow. According to experts, Prior User Experience Scores are determined from Table 33-37. Other NFRs are scored by the facilitators objectively. These tables are ordered according to the criticality ranking from Step 11. The cells colored in green in the weighted score highlight the technology alternative with the highest score.

Table 38: Weighted Scores for Modeling

Alternatives	Differentiating Factors					Weighted Score
	Community & Documentation	Maturity	Customizability	Prior User Experience	Performance Efficiency	
PyTorch	10.00	9.00	10.00	8.75	10.00	9.61
PyTorch Lightning	7.00	7.00	7.00	2.75	7.00	6.60
Keras on Tensorflow	10.00	9.00	6.00	5.00	7.00	8.33
Fast.ai on PyTorch	7.00	6.00	7.00	1.5	7.00	6.22
Tensorflow	10.00	9.00	10.00	4.25	9.00	9.13

Table 39: Weighted Scores for Experiment Tracking

Alternatives	Differentiating Factors							Weighted Score
	Maturity	Community & Documentation	Dashboard Customization	Prior User Experience	Operability	Installability	User Interface Aesthetics	
Tensorboard	10.00	5.00	6.00	6.25	7.00	8.00	5.00	7.36
Guild.ai	6.00	5.00	6.00	0.25	6.00	7.00	6.00	5.20
Weights & Biases	8.00	10.00	10.00	9.00	7.00	7.00	10.00	8.59
MLFlow	6.00	8.00	4.00	0.75	7.00	8.00	6.00	5.65
ClearML	6.00	7.00	8.00	0.25	6.00	9.00	10.00	6.20

:

Table 40: Weighted Scores for Data Preprocessing

Alternatives	Differentiating Factors				Weighted Score
	Operability	Community & Documentation	Performance Efficiency	Prior User Experience	
Pillow	9.00	9.00	7.00	7.25	8.43
Scikit-image	9.00	7.00	7.00	4.75	7.58
OpenCV	8.00	10.00	10.00	6.25	8.84
Tf.keras preprocessing	8.00	7.00	5.00	1.75	6.49
TorchVision	8.00	8.00	7.00	7.5	7.75

Table 41: Weighted Scores for Data Visualization

Alternatives	Differentiating Factors					Weighted Score
	Community & Documentation	Operability	Prior User Experience	User Interface Aesthetics	Customizability	
Seaborn	10.00	8.00	5.00	9.00	6.00	8.01
Plotly	10.00	7.00	3.00	9.00	9.00	7.51
Matplotlib	7.00	5.00	7.50	6.00	9.00	6.41

Table 42: Weighted Scores for Dashboarding

Alternatives	Differentiating Factors					Weighted Score
	Community & Documentation	Operability	Prior User Experience	User Interface Aesthetics	Customizability	
Streamlit	8.00	9.00	3.00	7.00	7.00	7.33
Gradio	6.00	10.00	0.75	7.00	7.00	6.70
Dash	10.00	7.00	0.25	9.00	10.00	7.10

Table 43: Weighted Scores for Data Storage

Alternatives	Differentiating Factors						Weighted Score
	Performance Efficiency	Operability	Community & Documentation	Prior User Experience	Fault Tolerance	Scalability	
Linux File System	5.00	10.00	10.00	7.50	5.00	5.00	7.76
MySQL (RDBMS)	6.00	6.00	8.00	2.75	8.00	6.00	6.15
MongoDB (NoSQL)	8.00	5.00	6.00	1.50	6.00	8.00	5.84

5.1.12 Construction of Possible Technology Stacks

In this step, facilitators construct the most suitable technology stack or stacks considering both weighted scores, criticality of DSW steps, and interoperability. The highest-scored Data Science Technology Stack consists of Linux File System, PyTorch[122], OpenCV[123], Weights & Biases[124], Seaborn[125], and Streamlit[126]. These tools are interoperable and can be installed on Linux-based systems according to their documentation pages and Step 6 & 7 selection methodology.

5.1.13 Finalizing the Technology Stack

If there are multiple alternatives, evaluators decide on the best Data Science technology stack using Delphi Method. If there is only one available stack, the selection is finalized. In this case there is only a single stack considered.

Table 44: Finalized Data Science Technology Stack

Data Science Tech Category	Data Storage	Data Preprocessing	Modeling	Experiment Tracking	Visualization	Dashboarding
Technology Alternative	Linux File System	OpenCV	PyTorch	Weights & Biases	Seaborn	Streamlit

5.2 Results

The following section summarizes the findings using three tables. Table 45 exhibits the descriptive statistics of the Domain Experts' inputs. Next, Table 46 depicts the descriptive statistics of Prior User Experience in technology alternatives. Finally, Table 47 represents the local and global weights of each NFR and ranks them altogether.

Table 45: Descriptive Statistics for each step of the DSW

Data Science Workflow Step	Non-Functional Requirements	Statistics								
		Mean	Median	Mode	Standard Deviation	Kurtosis	Skewness	Correlation of Variation (CV) %	Range	Count
Data Storage	Performance Efficiency	0.261	0.215	None	0.161	1.703	1.374	61.716	0.362	4
	Operability	0.295	0.263	None	0.155	-0.048	0.931	52.757	0.352	4
	Fault Tolerance	0.063	0.049	None	0.041	3.047	1.670	65.274	0.092	4
	Community & Documentation	0.199	0.194	None	0.128	-5.470	0.066	64.474	0.248	4
	Scalability	0.065	0.060	None	0.026	0.107	1.007	39.519	0.058	4
	Prior User Experience	0.116	0.119	None	0.088	-5.712	-0.039	75.935	0.165	4
Data Preprocessing	Performance Efficiency	0.199	0.143	0.143	0.175	3.116	1.652	88.126	0.399	4
	Community & Documentation	0.306	0.265	None	0.107	2.551	1.644	35.062	0.228	4
	Operability	0.400	0.379	None	0.183	-4.627	0.234	45.801	0.364	4
	Prior User Experience	0.097	0.094	None	0.037	-1.058	0.445	38.028	0.085	4
Modeling	Maturity	0.268	0.265	None	0.183	-4.476	0.052	68.207	0.381	4
	Community & Documentation	0.389	0.374	None	0.159	1.662	0.558	40.971	0.388	4
	Customizability	0.184	0.155	None	0.103	1.981	1.412	55.893	0.233	4
	Performance Efficiency	0.065	0.057	None	0.026	2.767	1.586	40.829	0.060	4
	Prior User Experience	0.093	0.098	None	0.035	1.789	-0.787	37.716	0.085	4
Experiment Tracking	Dashboard Customization	0.149	0.073	None	0.175	3.504	1.869	117.232	0.365	4
	Maturity	0.294	0.306	None	0.124	-4.585	-0.217	42.090	0.250	4
	Community & Documentation	0.151	0.149	None	0.056	1.455	0.237	37.460	0.138	4
	Operability	0.116	0.106	None	0.044	0.799	1.070	37.736	0.100	4
	Installability	0.114	0.105	None	0.067	1.257	0.769	58.912	0.161	4
	User Interface Aesthetics	0.044	0.043	None	0.012	-5.079	0.149	26.973	0.023	4
	Prior User Experience	0.133	0.134	None	0.099	-5.840	-0.007	74.789	0.184	4
Dashboarding & Visualization	Community & Documentation	0.302	0.319	None	0.140	-0.155	-0.604	46.238	0.326	4
	Operability	0.369	0.411	None	0.173	-1.170	-0.870	46.932	0.368	4
	User Interface Aesthetics	0.081	0.067	None	0.047	2.883	1.583	57.591	0.107	4
	Plot Variety	0.073	0.074	None	0.024	-1.000	-0.202	32.934	0.056	4
	Prior User Experience	0.176	0.183	None	0.105	-4.935	-0.137	59.997	0.211	4

Table 46: Descriptive Statistics of Prior User Experience for each step of the DSW

Technology	Alternative	Score	Mean	Median	Mode	Standard Deviation	Skewness	Range	Count	Selection
Data Storage	Linux File System	7.5	3.92	2.75	None	3.17	1.43	6.00	3.00	Linux File System
	MySQL	2.75								
	NoSQL	1.5								
Data Preprocessing	Pillow	7.25	5.50	6.25	None	2.36	-1.25	5.75	5.00	OpenCV
	Scikit-image	4.75								
	OpenCV	6.25								
	Tf.keras. preprocessing. image	1.75								
	TorchVision	7.5								
Modeling	PyTorch	8.75	4.45	4.25	None	2.76	0.97	7.25	5.00	PyTorch
	PyTorch Lightning	2.75								
	Keras on Tensorflow	5								
	Fast.ai on PyTorch	1.5								
	Tensorflow	4.25								
Experiment Tracking	Tensorboard	6.25	3.30	0.75	None	4.07	0.85	8.75	5.00	Weights and Biases
	Guild.ai	0.25								
	Weights & Biases	9								
	MLFlow	0.75								
	ClearML	0.25								
Visualization	Seaborn	5	5.17	5.00	None	2.25	0.33	4.50	3.00	Seaborn
	Plotly	3								
	Matplotlib	7.5								
Dashboarding	Streamlit	3	1.33	0.75	None	1.46	1.51	2.75	3.00	Streamlit
	Gradio	0.75								
	Dash	0.25								

Table 47: Weights of the NFRs

Data Science Workflow Step	Weight	Priority	Non-Functional Requirements	Local Weights	Local Priority	Global Weights	Normalized (%)	Global Priority
Data Storage	0.054	5	Performance Efficiency	0.261	2	0.0140	1%	8
			Operability	0.295	1	0.0158	2%	7
			Community & Documentation	0.200	3	0.0107	1%	8
			Prior User Experience	0.116	4	0.0062	1%	8
			Fault Tolerance	0.063	6	0.0034	0%	9
			Scalability	0.065	5	0.0035	0%	9
Data Preprocessing	0.125	3	Operability	0.399	1	0.0498	5%	4
			Community & Documentation	0.306	2	0.0381	4%	5
			Performance Efficiency	0.198	3	0.0248	3%	6
			Prior User Experience	0.097	4	0.0121	1%	8
Modeling	0.565	1	Community & Documentation	0.389	1	0.2196	22%	1
			Maturity	0.268	2	0.1514	15%	2
			Customizability	0.184	3	0.1040	10%	3
			Prior User Experience	0.093	4	0.0527	5%	4
			Performance Efficiency	0.065	5	0.0366	4%	5
Experiment Tracking	0.183	2	Maturity	0.294	1	0.0276	3%	6
			Community & Documentation	0.151	2	0.0276	3%	6
			Dashboard Customization	0.149	3	0.0273	3%	6
			Prior User Experience	0.133	4	0.0243	2%	7
			Operability	0.116	5	0.0212	2%	7
			Installability	0.114	6	0.0209	2%	7
			User Interface Aesthetics	0.044	7	0.0080	1%	8
Dashboarding. & Visualization	0.077	4	Community & Documentation	0.302	2	0.0232	2%	7
			Operability	0.369	1	0.0284	3%	6
			Prior User Experience	0.176	3	0.0135	1%	8
			User Interface Aesthetics	0.081	4	0.0063	1%	8
			Customizability	0.073	5	0.0056	1%	8

5.3 Analysis on Findings

Table 45 depicts the descriptive statistics for the AHP results:

- There is a total of 27 Differentiating Factors under different technology categories. 11 of them have a skewness between -0.5 and 0.5, which makes them symmetrical, 13 of these are positively skewed with a skewness value higher than 0.5, and 3 of them are negatively skewed, with a skewness value lower than -0.5. This result shows that approximately %60 of NFRs are skewed, indicating that evaluators have different preferences and tendencies. Therefore, using the mean value as the weight helps find common ground.

- The coefficients of variation for DFs are between %26.97 and %117.23. The DFs with the highest CVs are Dashboard Customization under Experiment Tracking at %117.23, Performance Efficiency under Data Preprocessing at %88.13, Prior User Experience under Data Storage at %75.94, Prior User Experience under Experiment Tracking at %74.79 and Maturity from Modeling with %68.20. These results show that the importance of these DFs fluctuates highly amongst evaluators.

- When the distribution of DFs' CV scores is considered, values below %40 can be considered low. The factors with low CV scores include User Interface Aesthetics under Experiment Tracking at %26.97, Plot Variety under Dashboarding & Visualization at %32.93, Community & Documentation under Data Preprocessing at %35.06, Community & Documentation and Operability under Experiment Tracking at %37.46 and 37.73. Also, others include Prior User Experience under Modeling and Data Preprocessing, respectively, at %37.72 and %38.02, and Scalability Under Data Storage at %39.52. Lower CV values demonstrate that there is a consensus among Evaluators.

- In addition, the DFs with the lowest CV are the ones with the lowest weights, such as User Interface Aesthetics and Plot Variety. Also, Community & Documentation under Modeling, the factor with the highest weight score under the most critical step of the DSW, has a CV of %40.91. It is not as low as the DFs mentioned before; nevertheless, it shows a consensus in deciding the most important factor. These scores demonstrate that domain experts have a shared understanding of the DFs that are at the extremes for their selection. However, it is imperative to note that the remaining DFs with higher weights do not conform to any trend or rule.

- Furthermore, the limited dataset size should also be considered when analyzing the results mentioned above. Due to its size, any data point can substantially affect the variation and values of weights.

Among the differentiating factors, the only factor that depends explicitly on the evaluators is the Prior User Experience (PUE). Since Tables 33-37 demonstrate an imbalanced experience score, it is imperative to investigate the PUE distributions and effects. Table 46 exhibits the descriptive statistics for PUE in different technology categories: Except for the Visualization technologies, all the technology cate PUE scores are either positively or negatively skewed.

- Among the skewed ones, except Data Preprocessing, the technology alternative with the highest PUE Score has the highest weighted score. This case shows that whatever the weight of the PUE score is, if the experience score is skewed, PUE e may be the factor that dominates the others and thus may be the factor that determines the most suitable technology.

- In the case of Visualization technologies, the scores are not skewed, and the alternative with the second highest score, Seaborn, is selected. So, this shows that when the PUE score is not skewed, it does not dominate and contributes to the weighted sum score in its respective weight.

- Even though Data Preprocessing is skewed in terms of PUE score, the fact that the PUE is the differentiating factor with the lowest weight within Data Preprocessing factors causes PUE not to dominate and fall behind.

- The results show that this population's experience with specific tools and not others is causing this PUE score bias in the weighted sum calculation.

- In addition, the PUE score distribution should be unique to all groups. Since this team is working in an academic environment, and their main priorities are the research itself and the academic outputs, they may be expected to be highly experienced in a specific set of tools and not be versatile in a larger technology pool. This case may be the opposite for a software team working in the industry.

- The variation in experience can also be due to the maturity of that technology category's market. For example, Experiment Tracking technologies are relatively newer when compared to Data Preprocessing and Data Storage. This situation may have led to Weights and Biases dominating the Experiment Tracking scores. In contrast, in the case of Data Preprocessing's PUE, it may have contributed in line with its weight.

- It is also important to note that PUE may not be as important in other workflows from the industry. Since productivity and efficiency are prioritized in the industry, PUE may not even be considered a significant factor for the weighted sum since the industry demands the developers to be versatile, adapt, and work with the technologies provided.

Table 47 demonstrates the differentiating factors' local and global weights and rankings. According to these results, the following outcomes are determined:

- The order of criticality of the DSW steps are as follows: Modeling, Experiment Tracking, Data Preprocessing, Dashboarding & Visualization, and Data Storage. Since the primary goal of this research is to create state-of-the-art modeling artifacts, it is a must to prioritize the modeling technologies.

- The most critical factors are Community & Documentation at %22, Maturity at %15, and Customizability at %10, all under Modeling. So, these three factors directly influence the %47 of the technology selection process. The remaining factors are spread out between %0 and %5. These outcomes can be considered reasonable since the primary goal of this research group is to develop cutting-edge models and techniques. To do so, they need reliable, customizable, and well-documented technologies to utilize these technologies up to their limits and make the best of their resources.

- The least important factors are Scalability and Fault Tolerance under Data Storage, with a %0.1 criticality score rounded to %0. Since this case study's dataset is stable and static, it is not regularly updated, making the problem of Scalability less critical. Moreover, the project is not a live deployed one, and it is backed up in multiple data stores, thus making the Fault Tolerance a less critical concern.

- Some DFs are common in different DSW steps. However, their importance can be quite different. For example, Modeling's Customizability has a global weight of %10, whereas Dashboarding & Visualization's Customizability has a global weight of %1. These differences can also be seen in other DFs, including Prior User Experience, Community & Documentation, and Operability. These differences in weights show that it is not practical to generalize the weights and rankings of DFs.

- Furthermore, these weights can change from workflow to workflow as well. The modeling step may be the most important for a Deep Learning based one, whereas Data Storage can be a critical step for a Big Data related system.

- Table 4 allows fragmenting decision-making process. Therefore, it enables decision-makers to see the effects of each component and its respective factors. This table also allows demonstrating what-if analyses in the absence or the change in the value of a factor.

5.3.1 What-if (Sensitivity) Analyses

What-if analysis is used to determine how the calculated result is affected by changes in the conditions. In this case, the conditions are the decision weights of the weighted sum. What-if analyses enable testing out different fictitious cases where the decision weights are altered and seeing the changes in the results. The following section investigates two fictitious cases to see the relationships between the decision weights and the selected alternatives.

Case 1:

Instead of a team with skewed technology experience, let the team be a highly-experienced team from a software company, in which software developers can work with any technology provided and do not have any specific technology preferences. So, this demonstrates a scenario where Prior User Experience is not considered. In this case, what would be the most suitable technology stack? Prior User Experience weight is set to 0 to answer this question, and once again, the weights are calculated after weights are normalized. Table 48 exhibits the changes in the selected alternatives for the conditions of the first case.

Table 48: What-if Analysis for Case 1

	Academic Case Study Weighted Score	Academic Case Study Selection	Case 1 Weighted Score (After Weights are Normalized)	Case 1 Selection
Data Storage	9.61	Linux File System	7,80	Linux File System
Data Preprocessing	8.59	OpenCV	9.12	OpenCV
Modeling	8.84	PyTorch	9.3	PyTorch
Experiment Tracking	8.01	Weights & Biases	8.43	Weights & Biases
Visualization	7.33	Seaborn	8.47	Plotly
Dashboarding	7.76	Streamlit	8.56	Dash

So, in this case where PUE is removed from consideration, there is a change in technologies regarding Visualization and Dashboarding. However, in the other cases there is no change the selected technologies.

Case 2:

Let the second case study be for a profit-oriented organization that wants to productize the given DSW. Next, they must deploy their project in a virtual environment. This deployment will bring additional costs, and their goal is to make the best of their resources and maximize profit while they may accept some loss in model performance. So, in this case, Performance Efficiency (Throughput) will be prioritized where available. For the sake of this case, let the decision weight of Performance Efficiency be 0.4, Prior User Experience be 0 and the remaining weights are normalized accordingly. Table 49 exhibits the changes in the selected alternatives for the conditions of the second case.

Table 49: What-if Analysis for Case 2

	Academic Case Study Weighted Score	Academic Case Study Selection	Case 2 Weighted Score (After Weights are Normalized)	Case 2 Selection
Data Storage	9.61	Linux File System	6,75	NoSQL (MongoDB)
Data Preprocessing	8.59	OpenCV	8.93	OpenCV
Modeling	8.84	PyTorch	9,25	PyTorch
Experiment Tracking	8.01	Weights & Biases	8.01	Weights & Biases
Visualization	7.33	Seaborn	8.47	Plotly
Dashboarding	7.76	Streamlit	8.56	Dash

So, in this case where Performance Efficiency is prioritized and PUE is removed, Linux File System is replaced with MongoDB and similar with the first case Visualization and Dashboarding tools are replaced accordingly.

5.4 Discussion

The following section discusses the results and findings of the proposed method and its respective case study.

- The modular architecture of the weighted sum enables breaking down the effect of each differentiating factor on the total score. Due to this modularity, the dominance of PUE in the final decision can be observed in Table 46. Also, in another Data Science Project, a different factor may affect or directly dictate the selection decision when this method is used. In such situations, the proposed method enables what-if (sensitivity) analyses and helps make deep dives.

- It needs to be emphasized that the evaluators with an academic mindset evaluated and scored the DFs. For example, this case study aims to develop a model with better confusion matrix statistics than academic baselines. However, if the evaluators scored from a profit-oriented perspective, they may have prioritized performance efficiency and throughput for better deployment and customer experience by compromising on model accuracy.

- Tables 45-47 show that Community & Documentation and Maturity are critical for all the steps. However, it is critical to note that the results attained here are unique to this research team. Even if it were a different team from the same domain, there could also be differing results.

- In addition, if a development team from the industry had scored the differentiating factors, it would not be surprising to see different results. If a development team from the industry presented the project, the expected requirements of the project would also be different. Even the DSW could have been different, including different steps. An additional deployment step could have been considered since it would be planned as a complete product.

- A team can have multiple projects they develop. Consequently, an analytics department can consist of multiple teams as well. So, scaling the technology selection approach to the organization level may be challenging and require tailoring the proposed method and integrations with other technology management tools. Thanks to its modular structure, the proposed method can be integrated with Data Science Roadmapping [9] and Data-Integrated Roadmaps [127] to align organizational strategy in AI & DS contexts, resources, and limitations with the requirements of each DSW.

- Finally, tables 45-47 show that each development team member can have different preferences and priorities, and using a multi-round AHP approach democratizes the decision-making process.

CHAPTER 6

CONCLUSION

6.1 Summary

This study aims to propose a solution to the Data Science technology selection problem. First, this study presents a literature review of the existing technology selection studies and evaluates these approaches in terms of their versatility and extensiveness under Table 4. The literature review shows that no approach comprehensively covers the end-to-end Data Science Workflow and the technologies.

Next, the new technology selection method that utilizes multi-criteria-decision-making and group decision-making has been developed iteratively in group meetings considering the strengths and shortcomings of the current methods. The method starts with constructing two teams: facilitators who run the process and evaluators who provide information. Then, the analytics project is mapped into a data science workflow using CRISP-DM or TDSP. Afterward, evaluators rank the steps of the DSW according to their criticality for the project. Subsequently, the functional requirements for each step are gathered and mapped considering Organization, Data, Infrastructure, Software, and Environment and Experience dimensions. Then, the facilitators filter the technology alternatives at each step according to these functional requirements. At this point, all the filtered technologies can provide the necessary functionality, but they differ in how they provide these functionalities. After this preliminary elimination, the facilitators work out the differentiating features for each step of the DSW using a customized MLR approach, as explained in section 4.3.1. Then, evaluators rank and score these features in terms of their criticality for their respective workflow steps using the AHP. Lastly, the facilitators score the technology alternatives in each DSW according to differentiating features and construct the most suitable technology stack or, if possible, stacks. If there are multiple stack alternatives, evaluators reach a consensus on a stack using Delphi Method.

Finally, we tested the proposed method in an academic computer vision workflow, and we selected the most suitable Data Science Technology Stack thanks to the inputs of computer vision researchers.

6.2 Contributions

This study presents multiple contributions to the literature by bridging some research gaps:

1. This study investigates the analytics technology selection literature and baselines them on key principles for developing strategic technology management toolkits [14]. As stated in the literature review, this study identifies the strengths and weaknesses of the methods presented in the literature. Then according to these insights, a systematic and comprehensive technology selection method that complies with all the key principles for developing strategic technology management toolkits is developed and tested with a Case Study. Compared with the alternatives from the literature, this selection method covers the end-to-end DSW steps and the entire analytics landscape. In contrast, the studies in the literature only focus on a single technology category like BI tools or present guidelines for a specific domain like Big Data with limited alternatives. Furthermore, as seen in Table 47, the proposed method using the local and global weights helps streamline and fragment the selection process while making the decision-making process transparent.

2. This study uses a Data Science Project from the Medical Computer Vision domain to validate the proposed method. Thus, the weights of the decision factors and the rankings for each step of the DSW are calculated. So, these factors, with their respective weights, provide the decision-making template for the projects of this domain while providing future practitioners the opportunity to play and customize the values of weights.

3. It is common for teams to make decisions based on the highest-paid person's opinion, seniority, or intuition [128]. These non-systematic decision-making approaches can lead to unexpected technical debts. With this proposed selection approach, not only everyone's opinion is heard, but also these opinions are directly integrated into the selection decision while democratizing the decision-making process.

4. During the facilitation of the proposed method, facilitators create multiple artifacts. Since this method is modular, each artifact can be utilized separately with minor modifications if needed.

    - **ODISEE Survey & Technology Vision Statement:** Technology Vision Statement is created using the responses to ODISEE questions. These both contribute to determining the functional requirements and limitations of a Data Science Project in a structured manner.

    - **Data Science Workflow Criticality Ranking:** This artifact provides insight into DSW steps where teams may encounter technology bottlenecks. Also, it enables teams to select the DSW steps to prioritize and helps them choose the DSW step where the functional requirements should originate and the steps these requirements cascade from the origin step.

    - **Functional Requirements Mapping:** This artifact demonstrates the mapping of each ODISEE dimension with the technology alternatives and the cascading requirements from the most critical DSW to the least critical one.

    - **Local & Global Weights Table:** This artifact exhibits the importance of each differentiating factor for each DSW step as well as for the entire workflow. It also enables practitioners to make what-if analyses and understand cause-and-effect relationships between factors and the selections.

    - **Customized MLR Approach for Determining Differentiating Features:** The tailored MLR approach enables finding the differences among technology alternatives. It also approaches the problem with a holistic point of view, and this way, it aligns academic expectations with industrial needs. Moreover, this method is not limited only to analytics-related technologies, but technologies from the entire software landscape can be compared with the help of it.

5. We have tested the customized MLR Approach for Determining Differentiating Features on the modeling technologies. The most prominent Deep Learning libraries and Frameworks, PyTorch, Tensorflow, Keras, and Fastai, are investigated. Thus, differentiating factors are determined by considering academic and industrial concerns.

6.3 Limitations

- Like any expert-based system, the proposed method is subject to expert bias [129].

- Not all differentiating methods, like Operability or Maturity, can be measured systematically.

- The proposed method may not be considered practical. Especially running the MLR process to determine the differentiating factors is laborious and demanding. A solution would be to dive deeply into the most critical technology category and do more superficial research for the remaining technology categories. Another approach would be to limit the research to a set of resources, including blogs and documentation pages. In addition, as facilitators conduct these MLRs for technologies, they will have the chance to build their know-how, and when documented, these can be reused for other projects. So, in the beginning, it may require hard work. However, after running it for multiple technology categories, facilitators would be expanding their know-how, and thus the amount of laborious work required would decrease in time.

- Getting the correct information about emerging technologies may be challenging due to a lack of resources or IP-related issues. This case is also valid for proprietary technologies since their documentation may include biased information, or they may collaborate with Data Science Blogs for sponsored content. Nevertheless, since open-source technologies have become the de facto standard in the analytics domain and the open-source community's transparency allows practitioners to know what to expect when considering alternatives that are not rich in information or may include biased specifications. [130]

- Since the method was tested on a single case study, it is not possible to make definite comments about the end-to-end method's generalizability. Nevertheless, the artifacts developed can be generalized and used in other projects.

6.4 Future Work

The following research opportunities are identified following the results of this study:

- Additional case studies may be conducted to assess the generalizability of the proposed method.

- In addition, case studies regarding project, department, and organizational levels can be conducted to test out the proposed method's scalability. Since the proposed method is modular, it can be integrated with Data Science Roadmapping [9] to align the organizational goals, resources, and limitations with the requirements to scale the proposed method.

- Furthermore, this study's scope may not be limited to the analytics field but also the software development landscape. The analytics industry requires the extensive use of several technologies at once. However, technology stack selection is not only a problem for analytics but also an issue for different practices of software engineering as well. Thus, software engineering leaders can modify the method according to their use to streamline their decision-making process and mitigate technical debt.

REFERENCES

[1] F. Provost and T. Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making," *Big Data*, vol. 1, no. 1, pp. 51–59, Mar. 2013, doi: 10.1089/big.2013.1508.

[2] S. Balkhi, "How Companies Are Using Big Data to Boost Sales, and How You Can Do the Same," *Entrepreneur*. https://www.entrepreneur.com/article/325923 (accessed Apr. 04, 2022).

[3] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for Big Data professions: A systematic classification of job roles and required skill sets," *Information Processing & Management*, vol. 54, no. 5, pp. 807–817, Sep. 2018, doi: 10.1016/j.ipm.2017.05.004.

[4] M. O. Gökalp, K. Kayabay, M. Zaki, A. Koçyiğit, P. E. Eren, and A. Neely, "Open-Source Big Data Analytics Architecture for Businesses," in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, Nov. 2019, pp. 1–6. doi: 10.1109/UBMYK48245.2019.8965572.

[5] "23 Must See Facts about State of Data Science and its Challenges in 2020 - 2021 - MLK - Machine Learning Knowledge." https://machinelearningknowledge.ai/must-see-facts-about-state-of-data-science-and-its-challenges/ (accessed Apr. 04, 2022).

[6] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AIMag*, vol. 17, no. 3, p. 37, Mar. 1996, doi: 10.1609/aimag.v17i3.1230.

[7] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," presented at the Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 2000, pp. 29–39.

[8] D. Sculley *et al.*, "Machine Learning: The High-Interest Credit Card of Technical Debt," p. 9, 2014.

[9] K. Kayabay, M. O. Gökalp, E. Gökalp, P. Erhan Eren, and A. Koçyiğit, "Data science roadmapping: An architectural framework for facilitating transformation towards a data-driven organization," *Technological Forecasting and Social Change*, vol. 174, p. 121264, Jan. 2022, doi: 10.1016/j.techfore.2021.121264.

[10] J. Kachaoui and A. Belangour, "A Multi-criteria Group Decision Making Method for Big Data Storage Selection," Sep. 2019, pp. 381–386. doi: 10.1007/978-3-030-31277-0_25.

[11] A. A. B. Abdelnabi, "An Analytical Hierarchical Process Model to Select

Programming Language for Novice Programmers for Data Analytics Applications," in *2019 International Arab Conference on Information Technology (ACIT)*, Dec. 2019, pp. 128–132. doi: 10.1109/ACIT47987.2019.8990995.

[12] M. Lněnička, "AHP Model for the Big Data Analytics Platform Selection," *AIP*, vol. 4, no. 2, pp. 108–121, Dec. 2015, doi: 10.18267/j.aip.64.

[13] S. Grandhi and S. Wibowo, "A Multi-Criteria Group Decision Making Method for Selecting Big Data Visualization Tools," vol. 10, no. 1, p. 6.

[14] C. Kerr, C. Farrukh, R. Phaal, and D. Probert, "Key principles for developing industrially relevant strategic technology management toolkits," *Technological Forecasting and Social Change*, vol. 80, no. 6, pp. 1050–1070, Jul. 2013, doi: 10.1016/j.techfore.2012.09.006.

[15] "Data science - Data Cymru." https://www.data.cymru/data-science (accessed Apr. 20, 2022).

[16] "NewVantage Partners Releases Annual Big Data Executive Survey for 2018," Jan. 08, 2018. https://www.businesswire.com/news/home/20180108005172/en/NewVantage-Partners-Releases-Annual-Big-Data-Executive-Survey-for-2018 (accessed Apr. 08, 2022).

[17] "Discuss | GeeksforGeeks | A computer science portal for geeks." https://discuss.geeksforgeeks.org/discuss/gfg_post_370924/11+Industries+That+Benefits+the+Most+From+Data+Science/gfg/?rel=https://www.geeksforgeeks.org/11-industries-that-benefits-the-most-from-data-science/&mode=light (accessed Apr. 08, 2022).

[18] "Importance of Data Science: A Simple Guide in 10 Steps." https://www.jigsawacademy.com/blogs/data-science/importance-of-data-science/, https://www.jigsawacademy.com/blogs/data-science/importance-of-data-science/ (accessed May 24, 2022).

[19] M. Frické, "The knowledge pyramid: a critique of the DIKW hierarchy," *Journal of Information Science*, vol. 35, no. 2, pp. 131–142, Apr. 2009, doi: 10.1177/0165551508094050.

[20] M Tabladillo, "What is the Team Data Science Process? - Azure Architecture Center." https://docs.microsoft.com/en-us/azure/architecture/data-science-process/overview (accessed Apr. 08, 2022).

[21] D. M. Vollmer, "How to make it simple to explain AI, ML, DL and Data Science?," *Medium*, Aug. 21, 2018. https://medium.com/@marcellvollmer/how-to-make-it-simple-to-explain-ai-ml-dl-and-data-science-a49e54d54a12 (accessed Apr. 25, 2022).

[22] "How are AI, Machine Learning, Deep Learning & Data Science Related?," *Corpnce*, Aug. 27, 2019. https://www.corpnce.com/relationship-ai-ml-dl-ds/ (accessed Apr. 25, 2022).

[23] "Data Science vs Machine Learning vs AI vs Deep Learning vs Data Mining: Know the Differences," *AltexSoft*. https://www.altexsoft.com/blog/data-science-artificial-intelligence-machine-learning-deep-learning-data-mining/ (accessed Apr. 25, 2022).

[24] "Apache Storm." https://storm.apache.org/ (accessed Apr. 25, 2022).

[25] "Apache Heron · A realtime, distributed, fault-tolerant stream processing engine." https://heron.apache.org/ (accessed Apr. 25, 2022).

[26] "Apache Kafka," *Apache Kafka*. https://kafka.apache.org/ (accessed Apr. 25, 2022).

[27] "Prophet," *Prophet*. http://facebook.github.io/prophet/ (accessed Apr. 25, 2022).

[28] "PyTorch." https://www.pytorch.org (accessed May 27, 2022).

[29] "TensorFlow," *TensorFlow*. https://www.tensorflow.org/?hl=tr (accessed Apr. 25, 2022).

[30] "Apache Hadoop." https://hadoop.apache.org/ (accessed May 27, 2022).

[31] "Apache Beam." https://beam.apache.org/ (accessed May 27, 2022).

[32] "Kubernetes." https://kubernetes.io/ (accessed May 27, 2022).

[33] "About," *OpenCV*. https://opencv.org/about/ (accessed May 27, 2022).

[34] "Bulut Bilişim Hizmetleri - Amazon Web Services (AWS)." https://aws.amazon.com/tr/ (accessed Apr. 29, 2022).

[35] "Bulut Bilişim Hizmetleri | Microsoft Azure." https://azure.microsoft.com/tr-tr/ (accessed Apr. 29, 2022).

[36] "Cloud Computing Services," *Google Cloud*. https://cloud.google.com/ (accessed Apr. 29, 2022).

[37] "GPU for Deep Learning." https://www.run.ai/guides/gpu-deep-learning (accessed Apr. 25, 2022).

[38] "NVIDIA BlueField Data Processing Units," *NVIDIA*. https://www.nvidia.com/en-us/networking/products/data-processing-unit/ (accessed Apr. 25, 2022).

[39] "Cloud Tensor Processing Units (TPUs)," *Google Cloud*. https://cloud.google.com/tpu/docs/tpus (accessed Apr. 25, 2022).

[40] S. Sagiroglu and D. Sinanc, "Big data: A review," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*, May 2013, pp. 42–47. doi: 10.1109/CTS.2013.6567202.

[41] D. Cetindamar, O. Can, and O. Pala, "Technology Management Activities and Tools: The Practice in Turkey," in *2006 Technology Management for the Global Future - PICMET 2006 Conference*, Istanbul, Turkey, Jul. 2006, pp. 92–98. doi: 10.1109/PICMET.2006.296557.

[42] D. Çetindamar, R. Phaal, and D. Probert, *Technology management: activities and tools*, Second edition. New York: Palgrave Macmillan, 2016.

[43] F. T. S. Chan, M. H. Chan, and N. K. H. Tang, "Evaluation methodologies for technology selection," *Journal of Materials Processing Technology*, vol. 107, no. 1–3, pp. 330–337, Nov. 2000, doi: 10.1016/S0924-0136(00)00679-8.

[44] "Strategy First in Technology Selection," *Contact Center Pipeline*. https://www.contactcenterpipeline.com/Article/strategy-first-in-technology-selection (accessed Apr. 11, 2022).

[45] A. V. Sivarama Prasad and N. Somasekhara, "The analytic hierarchy process for choice of technologies," *Technological Forecasting and Social Change*, vol. 38, no. 2, pp. 151–158, Sep. 1990, doi: 10.1016/0040-1625(90)90036-U.

[46] M. Feroze Ahmed and T. Ahmed, "1.8 - Status of Remediation of Arsenic Contamination of Groundwater in Bangladesh," in *Comprehensive Water Quality and Purification*, S. Ahuja, Ed. Waltham: Elsevier, 2014, pp. 104–121. doi: 10.1016/B978-0-12-382182-9.00002-5.

[47] Y. Yah Jusoh, "Open Source Software Selection Using an Analytical Hierarchy Process (AHP)," *AJSEA*, vol. 3, no. 6, p. 83, 2014, doi: 10.11648/j.ajsea.20140306.13.

[48] U. Şener, E. Gökalp, and P. E. Eren, "ClouDSS: A Decision Support System for Cloud Service Selection," in *Economics of Grids, Clouds, Systems, and Services*, Cham, 2017, pp. 249–261. doi: 10.1007/978-3-319-68066-8_19.

[49] T. L. Saaty, "What is the Analytic Hierarchy Process?," in *Mathematical Models for Decision Support*, G. Mitra, H. J. Greenberg, F. A. Lootsma, M. J. Rijkaert, and H. J. Zimmermann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1988, pp. 109–121. doi: 10.1007/978-3-642-83555-1_5.

[50] Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, "TOPSIS for MODM," *European Journal of Operational Research*, vol. 76, no. 3, pp. 486–500, Aug. 1994, doi: 10.1016/0377-

2217(94)90282-8.

[51] D. H. Hong and C.-H. Choi, "Multicriteria fuzzy decision-making problems based on vague set theory," *Fuzzy Sets and Systems*, p. 11, 2000.

[52] N. Gerdsri and D. F. Kocaoglu, "Applying the Analytic Hierarchy Process (AHP) to build a strategic framework for technology roadmapping," *Mathematical and Computer Modelling*, vol. 46, no. 7, pp. 1071–1080, Oct. 2007, doi: 10.1016/j.mcm.2007.03.015.

[53] C. Reddy, "Group Decision Making Advantages and Disadvantages," *WiseStep*, Jul. 03, 2016. https://content.wisestep.com/group-decision-making-top-advantages-disadvantages/ (accessed Apr. 12, 2022).

[54] "Group Decision Making - advantages, model, type, disadvantages, Group decision making methods." https://www.referenceforbusiness.com/management/Gr-Int/Group-Decision-Making.html (accessed Apr. 12, 2022).

[55] V. L. Putman and P. B. Paulus, "Brainstorming, Brainstorming Rules and Decision Making," *The Journal of Creative Behavior*, vol. 43, no. 1, pp. 29–40, 2009, doi: 10.1002/j.2162-6057.2009.tb01304.x.

[56] D. M. Schweiger, W. R. Sandberg, and J. W. Ragan, "Group Approaches for Improving Strategic Decision Making: A Comparative Analysis of Dialectical Inquiry, Devil's Advocacy, and Consensus," *AMJ*, vol. 29, no. 1, pp. 51–71, Mar. 1986, doi: 10.5465/255859.

[57] N. Harvey and C. A. Holmes, "Nominal group technique: An effective method for obtaining group consensus," *International Journal of Nursing Practice*, vol. 18, no. 2, pp. 188–194, 2012, doi: 10.1111/j.1440-172X.2012.02017.x.

[58] C.-C. Hsu and B. A. Sandford, "The Delphi technique: making sense of consensus," *Practical assessment, research, and evaluation*, vol. 12, no. 1, p. 10, 2007.

[59] K. Collier, B. Carey, D. Sautter, and C. Marjaniemi, "A methodology for evaluating and selecting data mining software," in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, Jan. 1999, vol. Track6, p. 11 pp.-. doi: 10.1109/HICSS.1999.772607.

[60] D. Dakic, S. Darko, S. Sladojevic, M. Arsenovic, and T. Lolić, *A Comparison of Contemporary Data Mining Tools*. 2017.

[61] A. H. Altalhi, J. M. Luna, M. A. Vallejo, and S. Ventura, "Evaluation and comparison of open source software suites for data mining and knowledge discovery," *WIREs Data Mining and Knowledge Discovery*, vol. 7, no. 3, p. e1204, 2017, doi: https://doi.org/10.1002/widm.1204.

[62] J. Klein and I. Gorton, “Design Assistant for NoSQL Technology Selection,” in *Proceedings of the 1st International Workshop on Future of Software Architecture Design Assistants - FoSADA ’15*, Montréal, QC, Canada, 2015, pp. 7–12. doi: 10.1145/2751491.2751494.

[63] S. Rouhani and S. Rotbei, “Big data platforms: in the lens of selection and evaluation approach,” *Journal of Decision Systems*, pp. 1–30, Jan. 2021, doi: 10.1080/12460125.2020.1869432.

[64] G. Ilieva, “Decision Analysis for Big Data Platform Selection,” *Engineering Sciences*, vol. LVI, Jun. 2019, doi: 10.7546/EngSci.LVI.19.02.01.

[65] S. Uddin, M. Rahman, S. Hasan, I. Rana, and S. M. Allayear, “A Fuzzy TOPSIS Approach for Big Data Analytics Platform Selection,” p. 8, 2019.

[66] S. Singh, R. Verma, and S. Koul, “Big Data Platform Selection at a Hospital: A Rembrandt System Application,” in *Innovations in Computational Intelligence and Computer Vision*, Singapore, 2021, pp. 130–140. doi: 10.1007/978-981-15-6067-5_16.

[67] M. Volk, S. Bosse, D. Bischoff, and K. Turowski, “Decision-Support for Selecting Big Data Reference Architectures,” 2019, pp. 3–17. doi: 10.1007/978-3-030-20485-3_1.

[68] M. Jarke and Y. Vassiliou, “A framework for choosing a database query language,” *ACM Comput. Surv.*, vol. 17, no. 3, pp. 313–340, Sep. 1985, doi: 10.1145/5505.5506.

[69] G. Büyüközkan, E. Mukul, and M. Güler, “Business Intelligence System Selection with Hesitant Fuzzy Linguistic MCDM Methods,” in *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*, Oct. 2019, pp. 141–146. doi: 10.1109/ICDSBA48748.2019.00038.

[70] M. Hanine, O. Boutkhoum, A. Tikniouine, and T. Agouti, “Application of an integrated multi-criteria decision making AHP-TOPSIS methodology for ETL software selection,” *SpringerPlus*, vol. 5, no. 1, p. 263, Dec. 2016, doi: 10.1186/s40064-016-1888-z.

[71] D. Lehmann, D. Fekete, and G. Vossen, “Technology Selection for Big Data and Analytical Applications,” *Open Journal of Big Data*, vol. 3, no. 1, p. 25, 2017.

[72] M. Volk, D. Staegemann, D. Bischoff, and K. Turowski, *Applying Multi-Criteria Decision-Making for the Selection of Big Data Technologies*. 2021.

[73] H.-Y. Lin, P.-Y. Hsu, and G.-J. Sheen, “A fuzzy-based decision-making procedure for data warehouse system selection,” *Expert Systems with Applications*, vol. 32, no. 3, pp. 939–953, Apr. 2007, doi: 10.1016/j.eswa.2006.01.031.

[74] K. Kayabay, “DEVELOPING AN ARCHITECTURAL FRAMEWORK FOR FACILITATING TRANSFORMATION TOWARDS DATA-DRIVEN ORGANIZATIONS,” p. 161, 2022.

[75] “What is a Tech Stack? Choosing What Goes In Yours | Heap.” https://heap.io/topics/what-is-a-tech-stack (accessed May 16, 2022).

[76] M. Glinz, “On Non-Functional Requirements,” in *15th IEEE International Requirements Engineering Conference (RE 2007)*, Oct. 2007, pp. 21–26. doi: 10.1109/RE.2007.45.

[77] *Systems and software engineering. Systems and software quality requirements and evaluation (SQuaRE). System and software quality models*. 2011.

[78] B. Gezici and A. K. Tarhan, “Systematic literature review on software quality for AI-based software,” *Empir Software Eng*, vol. 27, no. 3, p. 66, May 2022, doi: 10.1007/s10664-021-10105-2.

[79] J. Horkoff, “Non-Functional Requirements for Machine Learning: Challenges and New Directions,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*, Sep. 2019, pp. 386–391. doi: 10.1109/RE.2019.00050.

[80] K. M. Habibullah and J. Horkoff, “Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry,” *arXiv:2109.00872 [cs]*, Sep. 2021, Accessed: Apr. 06, 2022. [Online]. Available: http://arxiv.org/abs/2109.00872

[81] M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender, “Explainability as a Non-Functional Requirement,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*, Sep. 2019, pp. 363–368. doi: 10.1109/RE.2019.00046.

[82] V. Garousi, M. Felderer, and M. V. Mäntylä, “The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature,” in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*, Limerick Ireland, Jun. 2016, pp. 1–6. doi: 10.1145/2915970.2916008.

[83] M. T. Çaldağ and E. Gökalp, “The maturity of open government data maturity: a multivocal literature review,” *Aslib Journal of Information Management*, vol. ahead-of-print, no. ahead-of-print, Jan. 2022, doi: 10.1108/AJIM-11-2021-0354.

[84] K. G. Gupta, S. K. Maity, A. Das, and S. Wandhekar, “Performance analysis of different distribution of Python and TensorFlow to efficiently utilize CPU on HPC Cluster,” in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Dec. 2021, pp. 1–6. doi: 10.1109/ICECET52533.2021.9698764.

[85] Y. Wu, W. Cao, S. Sahin, and L. Liu, "Experimental Characterizations and Analysis of Deep Learning Frameworks," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 372–377. doi: 10.1109/BigData.2018.8621930.

[86] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey," *Arch Computat Methods Eng*, Feb. 2019, doi: 10.1007/s11831-018-09312-w.

[87] Y. Tian, Z. Zeng, M. Wen, Y. Liu, T. Kuo, and S.-C. Cheung, "EvalDNN: A Toolbox for Evaluating Deep Neural Network Models," in *2020 IEEE/ACM 42nd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, Oct. 2020, pp. 45–48.

[88] G. Nguyen *et al.*, "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, Jun. 2019, doi: 10.1007/s10462-018-09679-z.

[89] M. Nara, B. R. Mukesh, P. Padala, and B. Kinnal, "Performance Evaluation of Deep Learning frameworks on Computer Vision problems," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, pp. 670–674. doi: 10.1109/ICOEI.2019.8862603.

[90] R. Munjal, S. Arif, F. Wendler, and O. Kanoun, "Comparative Study of Machine-Learning Frameworks for the Elaboration of Feed-Forward Neural Networks by Varying the Complexity of Impedimetric Datasets Synthesized Using Eddy Current Sensors for the Characterization of Bi-Metallic Coins," *Sensors*, vol. 22, no. 4, p. 1312, Feb. 2022, doi: 10.3390/s22041312.

[91] S. Mahon, S. Varrette, V. Plugaru, F. Pinel, and P. Bouvry, "Performance Analysis of Distributed and Scalable Deep Learning," in *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, May 2020, pp. 760–766. doi: 10.1109/CCGrid49817.2020.00-13.

[92] L. Liu, Y. Wu, W. Wei, W. Cao, S. Sahin, and Q. Zhang, "Benchmarking Deep Learning Frameworks: Design Considerations, Metrics and Beyond," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, Jul. 2018, pp. 1258–1269. doi: 10.1109/ICDCS.2018.00125.

[93] S. Kim, H. Wimmer, and J. Kim, "Analysis of Deep Learning Libraries: Keras, PyTorch, and MXnet," in *2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA)*, May 2022, pp. 54–62. doi: 10.1109/SERA54885.2022.9806734.

[94] A. Jain, A. A. Awan, H. Subramoni, and D. K. Panda, "Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for High-Performance Deep Learning on Frontera," in *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers*

(DLS), Nov. 2019, pp. 76–83. doi: 10.1109/DLS49591.2019.00015.

[95] H. Dai, X. Peng, X. Shi, L. He, Q. Xiong, and H. Jin, "Reveal training performance mystery between TensorFlow and PyTorch in the single GPU environment," *Sci. China Inf. Sci.*, vol. 65, no. 1, p. 112103, Jan. 2022, doi: 10.1007/s11432-020-3182-1.

[96] B. J. Erickson, P. Korfiatis, Z. Akkus, T. Kline, and K. Philbrick, "Toolkits and Libraries for Deep Learning," *J Digit Imaging*, vol. 30, no. 4, pp. 400–405, Aug. 2017, doi: 10.1007/s10278-017-9965-6.

[97] J. Zhang, Y. Zhang, and Y. Lu, "A Comparative Study of Deep Learning Frameworks Based on Short-term Power Load Forecasting Experiments," *J. Phys.: Conf. Ser.*, vol. 2005, no. 1, p. 012070, Aug. 2021, doi: 10.1088/1742-6596/2005/1/012070.

[98] M. M. Yapıcı and N. Topaloğlu, "Performance comparison of deep learning frameworks," *Computers and Informatics*, vol. 1, no. 1, pp. 1–11, 2021.

[99] L. Ferreira, A. Pilastri, C. M. Martins, P. M. Pires, and P. Cortez, "A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost," in *2021 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9534091.

[100] R. Elshawi, A. Wahab, A. Barnawi, and S. Sakr, "DLBench: a comprehensive experimental evaluation of deep learning frameworks," *Cluster Comput*, vol. 24, no. 3, pp. 2017–2038, Sep. 2021, doi: 10.1007/s10586-021-03240-4.

[101] X. Du *et al.*, "Comparative Study of Distributed Deep Learning Tools on Supercomputers," in *Algorithms and Architectures for Parallel Processing*, vol. 11334, J. Vaidya and J. Li, Eds. Cham: Springer International Publishing, 2018, pp. 122–137. doi: 10.1007/978-3-030-05051-1_9.

[102] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: a comparison of deep learning frameworks," *Educ Inf Technol*, vol. 25, no. 3, pp. 1951–1963, May 2020, doi: 10.1007/s10639-019-10068-4.

[103] K. Dinghofer and F. Hartung, "Analysis of Criteria for the Selection of Machine Learning Frameworks," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 373–377. doi: 10.1109/ICNC47757.2020.9049650.

[104] M. C. Chirodea, O. C. Novac, C. M. Novac, N. Bizon, M. Oproescu, and C. E. Gordan, "Comparison of Tensorflow and PyTorch in Convolutional Neural Network - based Applications," in *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Jul. 2021, pp. 1–6. doi: 10.1109/ECAI52376.2021.9515098.

[105] J. Han, E. Shihab, Z. Wan, S. Deng, and X. Xia, "What do Programmers Discuss

about Deep Learning Frameworks," *Empir Software Eng*, vol. 25, no. 4, pp. 2694–2747, Jul. 2020, doi: 10.1007/s10664-020-09819-6.

[106] A. Parvat, J. Chavan, S. Kadam, S. Dev, and V. Pathak, "A survey of deep-learning frameworks," in *2017 International Conference on Inventive Systems and Control (ICISC)*, Jan. 2017, pp. 1–7. doi: 10.1109/ICISC.2017.8068684.

[107] A. Adewumi, S. Misra, N. Omoregbe, and L. F. Sanz, "FOSSES: Framework for open-source software evaluation and selection," *Software: Practice and Experience*, vol. 49, no. 5, pp. 780–812, 2019, doi: https://doi.org/10.1002/spe.2682.

[108] T. Azim and S. Ahmed, "Open Source Knowledge Base for Machine Learning Practitioners," in *Composing Fisher Kernels from Deep Neural Models*, Cham: Springer International Publishing, 2018, pp. 47–59. doi: 10.1007/978-3-319-98524-4_5.

[109] S. Elbagir and J. Yang, "Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment," *Hong Kong*, p. 5, 2019.

[110] I. Stančin and A. Jović, "An overview and comparison of free Python libraries for data mining and big data analysis," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2019, pp. 977–982. doi: 10.23919/MIPRO.2019.8757088.

[111] K. Dubovikov, "PyTorch vs TensorFlow — spotting the difference," *Medium*, Oct. 15, 2018. https://towardsdatascience.com/pytorch-vs-tensorflow-spotting-the-difference-25c75777377b (accessed Sep. 12, 2022).

[112] "PyTorch vs TensorFlow: Deep Learning Framework Comparison," *CIO Insight*, Jul. 26, 2022. https://www.cioinsight.com/big-data/pytorch-vs-tensorflow/ (accessed Sep. 12, 2022).

[113] "Keras vs. PyTorch: Difference Between Keras & PyTorch," *upGrad blog*, Dec. 28, 2020. https://www.upgrad.com/blog/keras-vs-pytorch/ (accessed Sep. 12, 2022).

[114] "Pytorch Vs Tensorflow Vs Keras: Here are the Difference You Should Know," *Simplilearn.com*, Jul. 27, 2020. https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article (accessed Sep. 12, 2022).

[115] "Keras vs TensorFlow vs PyTorch | Top 10 Awesome Differences to Learn," *EDUCBA*, Jul. 05, 2020. https://www.educba.com/keras-vs-tensorflow-vs-pytorch/ (accessed Sep. 12, 2022).

[116] "Keras vs PyTorch," *GeeksforGeeks*, Feb. 04, 2020. https://www.geeksforgeeks.org/keras-vs-pytorch/ (accessed Sep. 12, 2022).

[117] "Keras vs TensorFlow vs PyTorch | Deep Learning Frameworks," *Edureka*, Dec.

05, 2018. https://www.edureka.co/blog/keras-vs-tensorflow-vs-pytorch/ (accessed Sep. 12, 2022).

[118] "Tensorflow vs. PyTorch vs. Keras for Deep Learning." https://www.dominodatalab.com/blog/tensorflow-pytorch-or-keras-for-deep-learning (accessed Sep. 12, 2022).

[119] K. Budek, "PyTorch vs. TensorFlow - a detailed comparison," *Tooploox*, Oct. 02, 2020. https://www.tooploox.com/pytorch-vs-tensorflow-a-detailed-comparison (accessed Sep. 12, 2022).

[120] R. R. Uppari, "Comparison Between KERAS Library and FAST.AI Library using Convolution Neural Network(Image Classification) Model.," p. 50.

[121] "PyTorch vs TensorFlow in 2022," *News, Tutorials, AI Research*, Dec. 14, 2021. https://www.assemblyai.com/blog/pytorch-vs-tensorflow-in-2022/ (accessed Sep. 12, 2022).

[122] "PyTorch documentation — PyTorch 1.12 documentation." https://pytorch.org/docs/stable/index.html (accessed Sep. 12, 2022).

[123] "Home," *OpenCV*. https://opencv.org/ (accessed May 27, 2022).

[124] "Weights & Biases." https://docs.wandb.ai/ (accessed Nov. 14, 2022).

[125] M. Waskom, "seaborn: statistical data visualization," *JOSS*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/joss.03021.

[126] "Streamlit Docs." https://docs.streamlit.io/ (accessed Nov. 14, 2022).

[127] M. Han and Y. Geum, "Roadmapping for Data: Concept and Typology of Data-Integrated Smart-Service Roadmaps," *IEEE Trans. Eng. Manage.*, pp. 1–13, 2020, doi: 10.1109/TEM.2020.3013295.

[128] M. Bogdan and A. Borza, "BIG DATA ANALYTICS AS A STRATEGIC CAPABILITY: A SYSTEMATIC REVIEW," p. 9, 2019.

[129] R. J. B. Hutton and G. Klein, "Expert decision making," *Syst. Engin.*, vol. 2, no. 1, pp. 32–45, 1999, doi: 10.1002/(SICI)1520-6858(1999)2:1<32::AID-SYS3>3.0.CO;2-P.

[130] S. Kirsch, "How Open Source is Driving the Future of Data Science," *RTInsights*, Apr. 02, 2021. https://www.rtinsights.com/how-open-source-is-driving-the-future-of-data-science/ (accessed Nov. 30, 2022).

APPENDICES

APPENDIX A

ODISEE Interview Questions

Data Science Technology Selection Interview Questions:

We have developed this set of questions as part of an ongoing master's thesis study at Middle East Technical University, Informatics Institute, to gather information for the developed Data Science Technology Selection method. Please answer the following questions:

Personal:

Name:

Education:

Organization:

Role:

Preliminary:

Data Science Workflow:

- Please explain each step of the end-to-end data science project in detail.
- Which algorithms, methods, and tests will be used in each step?

Details:

Organization:

- Where does this Data Science Project take place?
- What are the budget and constraints for this project?
- Does your organization have any open-source or commercial software preferences?
- Does your organization have an on-premises or cloud platform preference?
- How many people work in your analytics team? What are their roles?

Data:

- What kind of dataset are you planning to use? (Data type, dimensions, size)

- Does your dataset require any storage requirements?
- How frequently will the dataset be updated?
- Does your dataset have any privacy and security requirements?
- Where will the Data be stored in terms of physical location?

Infrastructure:

- What are the hardware resources at hand for this project? (CPU, GPU, RAM, HDD)
- Is speed in storage critical for you?
- If you plan to use a cloud vendor, is there a particular vendor preference?
- What are the required hardware specifications for your cloud infrastructure?
- Do you require CUDA programming?
- Do you require a High-Performance Computing Infrastructure?
- Do you require parallel programming?
- Is there any infrastructure alternative you may consider?

Software:

- What statistical methods, algorithms, models, and techniques do you require for each step?
- Do you need pre-trained models?
- What type of development platform is required? (Notebook-based or IDE)
- Do you require a collaborative development platform?

Environment:

- Does the domain require any infrastructure limitations and regulations?
- Does the domain require any processing limitations and regulations?
- Does the domain require any storage limitations and regulations?

Experience:

- What infrastructures is your team experienced with?
- What technologies is your team experienced with?
- Does your team have a preferred programming language, library, or framework for each workflow step?
- Is there a technology that your team would like to avoid?

APPENDIX B

Ethics Committee Approval Letter

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
www.ueam.metu.edu.tr

Sayı: 28620816 /

20 MAYIS 2022

Konu : Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi : İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın P.Erhan EREN

Danışmanlığını yürüttüğünüz Kerem Nazlıel'in "VERİ BİLİMİ PROJELERİ İÇİN TEKNOLOJİ SEÇİM YÖNTEMİ GELİŞTİRME" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **0284-ODTUİAEK-2022** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

Prof.Dr. Mine MISIRLISOY
İAEK Başkan

APPENDIX C

Analytic Hierarchy Process Templates

Analytic Hierarchy Template: n= 5 riteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Pairwise Comparison Matrix

Importance Comparison of Workflows Steps

	Modeling	Experiment Trackir	Data Preprocessin	Visualization	Storage	0
Modeling	1	1	1	1	1	1
Experiment Tracking	1	1	1	1	1	1
Data Preprocessing	1	1	1	1	1	1
Visualization	1	1	1	1	1	1
Storage	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

	AHP		Consistency check
1	0.200	20.0%	Consistency OK
2	0.200	20.0%	0%
3	0.200	20.0%	
4	0.200	20.0%	
5	0.200	20.0%	
6	0.000	0.0%	
7	0.000	0.0%	
8	0.000	0.0%	
9	0.000	0.0%	
10	0.000	0.0%	
11	0.000	0.0%	
12	0.000	0.0%	
13	0.000	0.0%	
14	0.000	0.0%	
15	0.000	0.0%	

Figure 12: AHP Template for Workflow Step Criticality Ranking

Analytic Hierarchy Template: n=	6	Criteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Pairwise Comparison Matrix

Data Storage Criteria

	Community & Docun	Performance Effici	Prior User Experie	Operability	Scalability	Fault Tolerance
Community & Documentation	1	1	1	1	1	1
Performance Efficiency	1	1	1	1	1	1
Prior User Experience	1	1	1	1	1	1
Operability	1	1	1	1	1	1
Scalability	1	1	1	1	1	1
Fault Tolerance	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

	AHP		Consistency check
1	0.167	16.7%	Consistency OK
2	0.167	16.7%	0%
3	0.167	16.7%	
4	0.167	16.7%	
5	0.167	16.7%	
6	0.167	16.7%	
7	0.000	0.0%	
8	0.000	0.0%	
9	0.000	0.0%	
10	0.000	0.0%	
11	0.000	0.0%	
12	0.000	0.0%	
13	0.000	0.0%	
14	0.000	0.0%	
15	0.000	0.0%	

Figure 13: AHP Template for Data Storage Step

Analytic Hierarchy Template: n=	4	Criteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Data Preprocessing Criteria

Pairwise Comparison Matrix

	Community & Docur	Operability	Prior User Experie	Performance Effici	0	0
Community & Documentation	1	1	1	1	5	1
Operability	1	1	1	1	4	1
Prior User Experience	1	1	1	1	2	1
Performance Efficiency	1	1	1	1	3	1
	1/5	1/4	1/2	1/3	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

	AHP		Consistency check
1	0.250	25.0%	Consistency OK
2	0.250	25.0%	0%
3	0.250	25.0%	
4	0.250	25.0%	
5	0.000	0.0%	
6	0.000	0.0%	
7	0.000	0.0%	
8	0.000	0.0%	
9	0.000	0.0%	
10	0.000	0.0%	
11	0.000	0.0%	
12	0.000	0.0%	
13	0.000	0.0%	
14	0.000	0.0%	
15	0.000	0.0%	

Figure 14: AHP Template for Data Preprocessing Step

Analytic Hierarchy Template: n=	5	Criteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8

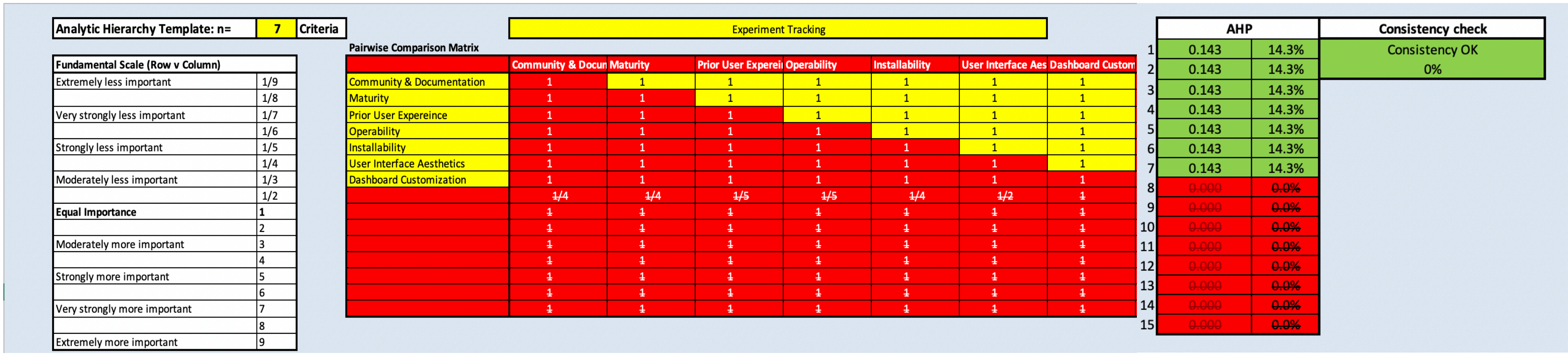
Pairwise Comparison Matrix

Modeling Criteria

	Community & Docum	Customizability	Prior User Experiei	Maturity	Performance Effici	0
Community & Documentation	1	1	1	1	1	1
Customizability	1	1	1	1	1	1
Prior User Experience	1	1	1	1	1	1
Maturity	1	1	1	1	1	1
Performance Efficiency	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

	AHP		Consistency check
1	0.200	20.0%	Consistency OK
2	0.200	20.0%	0%
3	0.200	20.0%	
4	0.200	20.0%	
5	0.200	20.0%	
6	0.000	0.0%	
7	0.000	0.0%	
8	0.000	0.0%	
9	0.000	0.0%	
10	0.000	0.0%	
11	0.000	0.0%	
12	0.000	0.0%	
13	0.000	0.0%	
14	0.000	0.0%	
15	0.000	0.0%	

Figure 15: AHP Template for Modeling Step



Analytic Hierarchy Template: n= 7 **Criteria**

Experiment Tracking

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	1
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Pairwise Comparison Matrix

	Community & Docun	Maturity	Prior User Experei	Operability	Installability	User Interface Aes	Dashboard Custom
Community & Documentation	1	1	1	1	1	1	1
Maturity	1	1	1	1	1	1	1
Prior User Expereince	1	1	1	1	1	1	1
Operability	1	1	1	1	1	1	1
Installability	1	1	1	1	1	1	1
User Interface Aesthetics	1	1	1	1	1	1	1
Dashboard Customization	1	1	1	1	1	1	1
	~~1/4~~	~~1/4~~	~~1/5~~	~~1/5~~	~~1/4~~	~~1/2~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~
	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~

	AHP	
1	0.143	14.3%
2	0.143	14.3%
3	0.143	14.3%
4	0.143	14.3%
5	0.143	14.3%
6	0.143	14.3%
7	0.143	14.3%
8	~~0.000~~	~~0.0%~~
9	~~0.000~~	~~0.0%~~
10	~~0.000~~	~~0.0%~~
11	~~0.000~~	~~0.0%~~
12	~~0.000~~	~~0.0%~~
13	~~0.000~~	~~0.0%~~
14	~~0.000~~	~~0.0%~~
15	~~0.000~~	~~0.0%~~

Consistency check
Consistency OK
0%

Figure 16: AHP Template for Experiment Tracking Step

Analytic Hierarchy Template: n=	5	Criteria

Fundamental Scale (Row v Column)	
Extremely less important	1/9
	1/8
Very strongly less important	1/7
	1/6
Strongly less important	1/5
	1/4
Moderately less important	1/3
	1/2
Equal Importance	**1**
	2
Moderately more important	3
	4
Strongly more important	5
	6
Very strongly more important	7
	8
Extremely more important	9

Pairwise Comparison Matrix

	Visualization Criteria					
	Community & Docun	Prior User Experiei	Operability	Plot Variety	User Interface Aes	0
Community & Documentation	1	1	1	1	1	1
Prior User Experience	1	1	1	1	1	1
Operability	1	1	1	1	1	1
Plot Variety	1	1	1	1	1	1
User Interface Aesthetics	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

	AHP		Consistency check
1	0.200	20.0%	Consistency OK
2	0.200	20.0%	0%
3	0.200	20.0%	
4	0.200	20.0%	
5	0.200	20.0%	
6	0.000	0.0%	
7	0.000	0.0%	
8	0.000	0.0%	
9	0.000	0.0%	
10	0.000	0.0%	
11	0.000	0.0%	
12	0.000	0.0%	
13	0.000	0.0%	
14	0.000	0.0%	
15	0.000	0.0%	

Figure 17: AHP Template for Visualization Step