AUTOMATED ANALYSIS OF CROSSING ACTIONS IN FOOTBALL COMMENTARY USING LARGE LANGUAGE MODELS

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

ANIL ERKUL

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

APRIL 2024

AUTOMATED ANALYSIS OF CROSSING ACTIONS IN FOOTBALL COMMENTARY USING LARGE LANGUAGE MODELS

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

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

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

Prof. Dr. Tuğba Taşkaya Temizel
Supervisor, **Department of Data Informatics, METU**

Examining Committee Members:

Prof. Dr. Altan Koçyiğit
Information Systems, METU

Prof. Dr. Tuğba Taşkaya Temizel
Department of Data Informatics, METU

Assoc. Prof. Dr. Kamil Demirberk Ünlü
Department of Industrial Engineering, Atılım University

Date: 19.04.2024

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 : Anıl Erkul

Signature :

ABSTRACT

AUTOMATED ANALYSIS OF CROSSING ACTIONS IN FOOTBALL COMMENTARY USING LARGE LANGUAGE MODELS

Erkul, Anıl
MSc., Department of Information Systems
Supervisor: Prof. Dr. Tuğba Taşkaya Temizel

April 2024, 173 pages

In football, well-executed crosses have a significant impact, directly creating about 15% of scoring opportunities in the top leagues of England, Italy, and France. This thesis focuses on the development of a comprehensive dataset sourced from the in-depth analysis of football commentary extracted from the English Premier League matches spanning the 2022-2023 Football season. The main goal is to gather and organize key details related to crossing actions. This involves collecting diverse information, such as the outcome of each cross, identifying the team and the player making the cross, and evaluating the sentiment or qualitative aspects of the executed crosses with human annotators. In the second phase of the thesis, large language models are used and fine-tuned to automatically identify, extract, and describe instances of crossing actions, along with their related details, in the extensive football commentary from Premier League matches. The results indicate that the large language models are useful to reveal crossing actions successfully.

Keywords: Football, Crossing, Labeling, Large Language Models, Text Classification

ÖZ

FUTBOL MAÇ ANLATIMLARINDAKİ ORTA AÇMA EYLEMLERİNİN BÜYÜK DİL MODELLERİ KULLANILARAK OTOMATİK ANALİZİ

Erkul, Anıl

Yüksek Lisans, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Prof. Dr. Tuğba Taşkaya Temizel

Nisan 2024, 173 sayfa

Futbolda İngiltere, İtalya ve Fransa gibi üst düzey liglerde gol fırsatlarının %15'i doğrudan başarılı bir şekilde yapılmış ortalar sonucunda gerçekleşmektedir. Bu tez, 2022-2023 futbol sezonu İngiltere Premier Lig maçlarına ait maç anlatımlarından oluşturulan kapsamlı bir veri setinin derinlemesine analiz edilmesine odaklanmaktadır. Ana hedef, orta eylemlerine yönelik temel göstergeleri toplamak ve organize etmektir. Bu doğrultuda; tezin ilk aşamasında maç anlatımlarındaki her bir ortanın sonucu, ortayı yapan takım ve oyuncu, ortaların duygu ve niteliksel yönlerinin insan tarafından tespiti ve etiketlenmesi gerçekleştirilmekte ve bir veri seti oluşturulmaktadır. İkinci aşamadaysa, oluşturulmuş olan veri seti ile ince ayar (fine-tune) yapılmış büyük dil modelleri ile, orta eylemlerinin otomatik bir şekilde tespiti ve özelliklerinin tanımlanması sağlanmıştır. Sonuçlar büyük dil modellerinin orta eylemlerinin tespiti ve karakteristiklerinin tespitinde başarılı bir şekilde kullanılabileceğini göstermektedir.

Anahtar Sözcükler: Futbol, Orta Eylemi, Etiketleme, Büyük Dil Modelleri, Metin Sınıflandırması

To My Father

Audentes Fortuna iuvat

ACKNOWLEDGMENTS

First and foremost, I extend my sincere gratitude to my thesis advisor Prof. Dr. Tuğba Taşkaya Temizel for her unwavering support, invaluable guidance during challenging moments, extensive knowledge, and enduring patience have been the cornerstone of my journey through the completion of this thesis and my academic pursuit. I would also like to extend my warmest thanks to the professors in the Graduate School of Informatics at Middle East Technical University for their valuable insights and contributions to my academic growth.

I would like to thank my jury members, Prof. Dr. Altan Koçyiğit and Assoc. Prof. Dr. Kamil Demirberk Ünlü for their valuable input and careful criticism of my work. The quality of this thesis has enhanced significantly with their expertise.

I want to express my gratitude to my father Tamer Erkul who is my pillar of strength. I credit my current self to his everlasting presence, wisdom, and love. I hope someday I will be a great dad just like him.

I would like to thank my girlfriend Büşra Emiroğlu for her enduring love, care, and patience as well as the immeasurable joy she has brought into my life.

Special thanks to my cousins Mert Erkul and Kaan Erkul for their consistent support and encouragement throughout the thesis process as they have always done.

I also want to thank my friends Deniz İşgenç, Furkan Yılmaz, Mert Karakaş, Ozan Ersan, and Ronay Pekgöz for their assistance, time, and motivation during the thesis process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BART	Bidirectional and Auto-Regressive Transformers
BERT	Bidirectional Encoder Representations from Transformers
FIFA	Fédération Internationale de Football Association
GPT	Generative Pre-trained Transformer
HELP	Human Expert Labeling Process
IAA	Inter-Annotator Agreement
KPI	Key Performance Indicator
LLM	Large Language Model
NLP	Natural Language Processing
Q&A	Question and Answer
RFID	Radio-Frequency Identification
UEFA	Union of European Football Associations
USD	United States Dollar
WOS	Web of Science

CHAPTER 1

INTRODUCTION

Football is often referred to as the world's most captivating sport, with an influence extending far beyond the green fields. With the rapid growth of its fame, power, and economic impact, winning matches has become more vital than ever. To do so, key performance indicators of scoring have been studied in recent years. One of those aspects is the cross delivery. They are the reasons for many scoring opportunities created in the matches and the number of crosses delivered in matches has been increasing recently (FIFA, 2023). Therefore, determining the key aspects of the crosses and identifying the players that can deliver better crosses is an important task.

The features of the crosses are mainly identified by spatial-temporal analysis or human notations by companies such as StatsBomb and Wyscout. Despite giving a good insight into the position of the players and ball, spatial-temporal analysis cannot determine the vitality of a challenge that a defender makes or the delivery of a good cross that was not touched by a teammate. These insights can be extracted when the match is watched, observed, and the execution is confirmed as a successful cross. Therefore, although data retrieval processes have rapidly increased in recent years, analyzing and interpreting them remains a challenging task. Moreover, it continues to demand significant time and human resources.

One of the real-time observations made during the matches, and crosses, is the match commentaries. In recent years, especially online commentaries have gained notable interest from fans because of their accessibility, ease of reading, and following their favorite teams. These commentaries are nearly written for all the matches played in the top leagues of each country. In other words, there is a bunch of insight and observations for each top-level match played now only to be analyzed.

The Large Language Models (LLMs) have rapid growth in recent years with an impact on a variety of tasks including classification and Natural Language Processing (NLP). They are used in various domains such as literature, medicine, design, and language translation. With the help of these models, the insights from the online commentaries can also be extracted and features of the crossing actions can be determined.

In this thesis, the live commentaries from the 304 matches of the Premier League 2022-2023 Season are used to extract features of the crosses using LLMs. To do so, these commentaries are labeled by 17 different annotators to determine the existence of a cross and its properties such as crossing player, outcome of the cross, and quality of the cross. This labeled dataset containing the commentaries is used as the ground truth to fine-tune the LLMs and observe their performances on this task.

1.1. Research Questions

The following research questions are aimed to be answered in this thesis:

Research Question 1: What is the reliability of online football commentaries for gaining insights into key performance indicators, such as crossing action, in football and how can a dataset be constructed through labeling of these actions?

- Research Question 1.1: How can crossing actions in a football match be found and annotated using online football commentaries?

- Research Question 1.2: How can players responsible for specific actions in football matches, such as crossing, be found and annotated using online football commentaries?

- Research Question 1.3: How can outcomes of the specific actions in football matches, such as crossing, be found and annotated using online football commentaries?

- Research Question 1.4: How can the quality of the specific actions in football matches, such as crossing, be identified using online football commentaries?

Research Question 2: How can labels that will be employed in Large Language Models be defined and validated to automate the extraction of key performance indicators?

Research Question 3: How effectively can Large Language Models extract key performance indicators in football, such as crossing action, from online football commentaries?

- Research Question 3.1: How can key performance indicators be detected for future matches using a dataset containing commentary from past events?

- Research Question 3.2: How can key performance indicators for a specific team be accurately identified when utilizing a dataset containing commentary from matches involving different teams?

1.2. Contribution of the Study

The contributions of this thesis are as follows:

- Proposing a novel and expanded framework for defining and categorizing the outcomes of the crosses in football.

- A new dataset with 19686 rows consisting of live football commentaries that supplemented with detailed labels capturing various aspects of crossing actions. This dataset also includes essential information such as time intervals, dates, participating teams, and match scores.

- LLM models can be used to detect the crossing actions in text-formatted datasets and extract their features which can be used by managers and decision-makers for analyzing their team and opponents, scouting, and improving training performance.

The examination of crossing actions in football is addressed within the existing literature (Mitrotasios et al., 2021),(Pulling et al., 2018). Nevertheless, there is a necessity for the identification and definition of new outcome categories to gain deeper insights into crossing actions. These additional categories not only enhance the precision of crossing notations but also enable tracking of subsequent events triggered by them.

The online football commentary has been introduced more than 15 years ago (Sabater et al., 2008). However, their usage in football analytics is not common. With the rapid increase and success of Large Language Models, this text data can be utilized in various aspects to extract key performance indicators in the football domain. In this thesis, a new dataset consisting of 19686 distinct football commentaries from 304 matches is constructed. This dataset consists of information from each commentary, including match scores, match dates, event time, and participating teams. Besides this information, a labeling process is handled with human annotators to extract information about the crossing actions in each commentary. The existence of a cross, the team and player who perform the crossing action, the outcome of the cross, the quality of the cross, and the existence of a secondary cross and its properties are all included in the dataset.

Finally, Large Language Models are fine-tuned with the newly created dataset capable of automatically labeling whether a given text contains a crossing action and adapted to classify various aspects of the crosses. The training of these models is conducted with open-source LLMs.

1.3. Organization of the Thesis

The organization of the thesis is as follows:

- Chapter 1 describes the overview of the thesis and outlines the contribution of this research.

- Chapter 2 describes the reasoning behind this research and provides a literature review of it.

- Chapter 3 covers the methodology followed in this research and provides the results.

- Chapter 4 consists of a discussion of the research questions and describes the limitations.

- Chapter 5 contains a conclusion that gives a summary of the results and findings of the research and describes future works which can be conducted.

CHAPTER 2

RELATED WORK

In this section, a review of related work covering the topics of the thesis is provided. First, a review of football and crossing actions' definitions, importance, and characteristics are provided. Later, the data analytics in the football domain related to the topic of this thesis is stated. After that, the process of labeling to generate ground truth for machine learning tasks is explained. Then, the Large Language Models (LLMs) definitions, a brief history, and popular models are discussed. Finally, the usage and applications of Natural Language Processing (NLP) in the football domain are stated.

2.1. Football and Crossing Action

Legendary coach Bill Shankly famously expressed that football is much more serious than a matter of life and death. Indeed, association football (soccer), or simply football, is a sport that extends its influence beyond the pitch and has shaped society since the 19th century.

The most essential aspect of football is goal scoring as it determines the difference between winning and losing. Therefore, the correlated elements with goal scoring such as shots on target, number of corners, and proportion of possession hold significant importance in football. (Wright et al., 2011). Another substantial action that leads to a score is a cross.

Cross can be defined as the attempted or successful play of the ball from a wide area into the penalty box to create an opportunity for a teammate to score. (Yamada & Hayashi, 2015) (Wu et al., 2020). In addition to their great potential to create scoring opportunities, crosses add an extra layer of challenge for the opposing defense and force them to form their defensive line deeper in the field. This allows the attacking team to get the momentum in the play.

With the knowledge of the importance of crossing, teams put a lot of effort into it. For instance, European clubs utilize crossing strategies with an average of 15 crosses per match (Stats Perform, 2017). About 15 percent of the scoring opportunities are created by crosses in the England, Italian, and French leagues which belong to the Big Five European Football Leagues (Worville, 2021). Furthermore, the number of goals scored from crosses during the FIFA 2022 (Fédération Internationale de Football Association) World Cup increased to 48, from the 25 goals scored in 2018 (FIFA, 2023).

The crosses can be delivered with two options in the flow of the game: there are set play crosses which are delivered after corners and free kicks and open play crosses (Vecer, 2014). These crosses can be in 6 types according to FIFA's Football Language (FIFA, 2021):

- Inswinging: ball with a path that bends towards the goal coming from wide positions.
- Outswinging: ball with a path that bends away from the goal coming from wide positions.
- Driven: ball with pace and strength whether on the ground in the air coming from wide positions.
- Lofted: ball looped into the box coming from wide positions.
- Cut-back: ball delivered diagonally back coming from wide positions.
- Push: ball delivered with the medium speed with a target and accuracy from wide positions.

The ability of a cross to change the direction of a game is significant, therefore delivering a successful cross is crucial in the game. A cross can be defined as successful if:

- The cross is delivered to a teammate and the opportunity to score is created.
- The cross receiver cannot touch the ball despite an opportunity to score is created.
- The cross is cleared by an opponent with a vital challenge (Kim et al., 2019)

The crossing skills of players and teams and the impacts of the crosses are evaluated from many perspectives. The mechanisms and characteristics of the crosses are investigated to determine the game plans for the decision makers (Mitrotasios et al., 2021),(Pulling et al., 2018), (Yamada & Hayashi, 2015), the game theoretic models are considered to get insight into the goal scoring correlation (Sarkar, 2018), statistical analysis of the impacts of crosses is considered (Vecer, 2014),(Pulling et al., 2018) and spatial-temporal data is used the inspect the crossing success (Wu et al., 2020).

The characteristics of a cross are crucial to determining its success. The crosser or the player who performs the crossing action (Yamada & Hayashi, 2015), (Wu et al., 2020), the game interval that the cross is made (Mitrotasios et al., 2021), (Partridge & Franks, 1989), (Pulling et al., 2018), (Wu et al., 2020), the outcome of the cross (Mitrotasios et al., 2021), (Pulling et al., 2018), the cross delivering area or position (Mitrotasios et al., 2021), (Pulling et al., 2018), (Yamada & Hayashi, 2015), (Wu et al., 2020), the delivery type of the cross (Mitrotasios et al., 2021), (Pulling et al., 2018), position of the receiver (Yamada & Hayashi, 2015), (Wu et al., 2020) and defensive proximity to crossing player (Mitrotasios et al., 2021), (Pulling et al., 2018), (Yamada & Hayashi, 2015), (Wu et al., 2020) are identified as foundational to the success of a cross.

2.2. Data Analytics in Football and Notational Analysis

Data analytics in sports conducts mathematical analysis of sports-related data including the utilization of data visualization, statistics, and data management. In general, the focus is on aspects appearing during the games such as event and action data (Alamar, 2013) (Link, 2018). This information can be used by coaches, decision-makers and trainers to improve the performance of athletes and analyze opponents. (Morgulev et al., 2018)

Football stands out as one of the most beloved sports globally with five billion fans worldwide and fostering a giant economy generating over 200 billion USD (CIES, 2018), (FIFA, 2021). Within the domain of sports analytics, football holds its ground as a cornerstone for data analytics. The main objectives of data analytics in football are:

- Analyzing opponents' tactics and players to develop a strategy for upcoming matches. (Link, 2018),
- Using statistics and performance variables of the players to find suitable players for the team (Ghar et al., 2021),
- Using biometric devices, cameras, and RFID tags, the location, pace, and route of players and the ball can be recorded as tracking data, enabling the analysis of physical activity, player behavior, and movement patterns. This data can be utilized to create tailored training programs (Morgulev et al., 2018), (Link, 2018),
- Data collected during the training and matches can be used to create decision-making systems to predict potential injuries (Piłka et al., 2023).

To evaluate the success of a team using different types of data obtained, Key Performance Indicators (KPIs) are used such as the number of shots on target, number of successful crosses, total passes, and possession rate in football (Herold et al., 2021). These KPIs help coaches to know where to direct the attack according to the opponent's weaknesses, where to create progression in attack, where opponents usually press and which players they use most frequently, what type of attack the opponent employs, and which generates the most danger (Almenara, 2021).

One of the main methods to determine the KPIs is notational analysis. Notational analysis is a method of forming records of the events to examine various performance indicators through a process (James, 2006). The decision-makers and coaches can benefit from this analysis at the individual and team levels (Hughes et al., 2012).

All teams have their own methodologies to identify the KPIs they require by notational analysis to improve their game, increase athletes' performances, and identify opponents better. All these methodologies contain certain difficulties of their own. The main challenges are human errors during the notation process and the objectivity of annotators. To solve these issues, working with multiple analysts for the notation process is highly encouraged (Arastey, 2018).

2.3. Labeling and Ground Truth in Machine Learning Projects

Machine learning has been a noticeable topic for a decade and its influence is prominently visible in every domain. As the saying "garbage in, garbage out" holds a profound truth, the quality of data upon which a machine learning model is trained is principal to its performance and accuracy. Thus, the quality of datasets to be used in machine learning projects is a major concern (Dai et al., 2018).

In conjunction, data labeling plays a significant role in determining data quality within machine learning projects which is the method of tagging the data with meaningful annotations or labels (Mahalle et al., 2023). This labeling output is used as the objectively correct labels and referred to as ground truth which will be used in machine learning models (Tecimer et al., 2021). The quality of this ground truth directly affects the models' performances and can be reduced it if noisy ground truth is supplied (Yilin Zhang et al., 2022).

As the significance of ground truth is expressed, it is important to understand how ground truth is added to a dataset. Therefore, how the contribution collectively is made should be described. One of the methodologies that can be followed is the Human Expert Labeling Process (HELP) by Aslan et al. (2017). This process consists of four main stages:

1) Planning: Preparation of training material, literature search on labels to be used in labeling, preparing a dictionary or definition of each label, and deciding the labeling tool.
2) Labeler Recruitment: Recruiting Labelers, training the labelers, evaluation of labels from sample data.
3) Labeling: Labeling of the dataset by recruited labelers, monitoring the labelers, and giving feedback to them.
4) Post-Labeling: Evaluating the labelers' agreement and forming the final labels, measure inter-annotators agreement.

The inter-annotator agreement (IAA) is a measure that reflects the agreement level of the labelers (Braylan et al., 2022). There are various metrics that show the IAA within a labeling assignment. The most basic metric is raw agreement which is found by simply counting the identical labels and reporting it as a percentage. However, to get more insight into the reliability of the labeling, a metric from the kappa/alpha family should be considered (Artstein, 2017). The most common measures are Cohen's Kappa and Krippendroff's alpha.

The main differences between Cohen's kappa and Krippendroff's alpha are their ability to measure different data types and the handling capabilities of the number of labelers. Cohen's Kappa can be used for categorical variables and for two labelers, on the other hand, Krippendroff's alpha can be used for different types of data including categorical, ordinal, interval, or ratio-level data. In addition, it can be used when there are more than two labelers (Cohen, 1960), (Hayes & Krippendorff, 2007).

Cohen's Kappa which is represented by κ ranges from -1 to +1 and measure IAA for categorical labels when there are two labelers (Cohen, 1960). +1 is the perfect agreement and 0 is the agreement if random guessing would have done. The negative values practically do not happen according to Cohen (1960), (McHugh, 2012). The agreement levels for each Kappa value are determined by McHugh (2012) and shared in Table 1.

Table 1 Kappa Values and Their Level of Agreement

Kappa Value	Level of Agreement	% of Data that are Reliable
0-.20	None	0-4%
.21-.39	Minimal	4-15%
.40-.59	Weak	15-35%
.60-.79	Moderate	35-63%
.80-.90	Strong	64-81%
Above .90	Almost Perfect	82-100%

Cohen's kappa can be calculated according to Equation (2-1) where P_a is the actual agreement which is the proportion that raters assigned the same label and P_e is the expected agreement that is the probability if raters randomly guessed the labels which is calculated by Equation (2-2). In this formula cm_1 and cm_2 represent the marginal values of columns 1 and 2 respectively and rm_1 and rm_2 represent marginal values of rows 1 and 2 respectively. The n value is the number of labeled data (McHugh, 2012).

$$\kappa = \frac{P_a - P_e}{1 - P_e} \tag{2-1}$$

$$P_e = \frac{\frac{(cm_1 \times \ rm_1)}{n} + \frac{(cm_2 \times \ rm_2)}{n}}{n} \tag{2-2}$$

There are also cases where the disagreement between raters for different categories is not the same, and the classes are ordinal. In this case, the disagreement in labeling tasks between different levels of orders has a different impact (Warrens, 2011). Therefore, a metric that considers the extent of disagreement between ratings is required for this task which is Cohen's weighted kappa (Cohen, 1968). There are many types of weighting patterns used such as linear and quadratic weights. Linear weights are used when it is assumed that a consistent rise in disagreement exists whereas quadratic weights are used to place more significance on substantial disagreements (Brennan & Prediger, 1981), (Ben-David, 2008).

The calculation of Cohen's weighted kappa is given in Equation (2-3). In this formula $P_{a_{ij}}$ is the actual agreement for category i by rater 1 and category j by rater 2 and $P_{e_{ij}}$ is the expected agreement for category i by rater 1 and category j by rater 2. Lastly,

w_{ij} is the weight determined to disagreement levels for category i by rater 1 and category j by rater 2. The weight calculations for linear and quadratic schemes are given in Equation (2-4) and Equation (2-5) respectively. C Is the total number of ordinal categories.

$$K_w = 1 - \frac{\sum w_{ij} P_{a_{ij}}}{\sum w_{ij} P_{e_{ij}}} \tag{2-3}$$

$$w_{ij} = 1 - \frac{|i - j|}{C - 1} \tag{2-4}$$

$$w_{ij} = 1 - \frac{(i - j)^2}{(C - 1)^2} \tag{2-5}$$

After the IAA is measured and inconsistent data is excluded, the final labels needed to be assigned. It can be done in various ways: assigning the most voted labels as the final level and reaching a consensus for each label are some useful methods (Aslan et al., 2017).

2.4. Large Language Models (LLMs)

Machine learning algorithms are used in various domains for a wide array of roles and applications for tasks like classification, forecasting, clustering, and Natural Language Processing (NLP) (Ahuja et al., 2019). Similarly, as academia and business practices continually advance and evolve, Large Language Models (LLMs) have found diverse applications across various domains (Hou et al., 2023) (Nagarhalli et al., 2021). For instance, LLMs derive significant benefits across key applications such as text generation, translation, conversational interfaces, question answering, and summarization (Radford et al., 2019), (Britz et al., 2018), (Devlin et al., 2018). This rapid development is initiated with the proposal of transformers by Vaswani et al. (2017). Following this contribution, many derivative models have been introduced such as BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018), GPT (Generative Pre-trained Transformer) (Radford et al., 2018), T5 (Raffel et al., 2019), BART (Lewis et al., 2019), DistilBERT (Sanh et al., 2020) and Longformer (Beltagy et al., 2020).

BERT is a bidirectional language model that learns context from both left and right directions of a word. With this bidirectional training concept, the model captures a deeper language understanding by considering the complete context of a word (Devlin et al., 2018). GPT is a series of language models introduced by OpenAI that predicts and generates text. The initial model proposed a transformer architecture that contains a sequential input process. This enabled the model to generate longer texts with better coherence (Radford et al., 2018). T5 which is the Text-to-Text Transfer Transformer is a model that treats all NLP tasks as a text-to-text problem and is enabled to combine diverse applications such as summarization, question answering, and translation into one unified structure. BART is a denoising autoencoder that is trained by introducing corruptions to text with random noise and learning a model to restore the original content from these modified versions (Lewis et al., 2019). DistilBERT is the condensed edition of BERT which enables a comparable performance level (keeping 97% of its language understanding capabilities) with significantly reduced parameters (reducing the size by 40%) and quicker inference speed (60% faster) (Sanh et al., 2020). Longformer is a model that processes longer text inputs more efficiently while keeping the same performance level by incorporating a self-attention approach that minimizes the computational cost for long documents (Beltagy et al., 2020).

After the introduction of GPT, OpenAI has released larger models called GPT-2 with 1.5 billion parameters (Solaiman et al., 2019), GPT-3 with 175 billion parameters (Brown et al., 2020) and, GPT-4 which is a brand-new model that is assumed to have over 1.7 trillion parameters allowing to have multi-model capabilities that process different data types than text like image documents (OpenAI et al., 2023), (Bastian, 2023). OpenAI also released a model called ChatGPT which handles conversation, question answering, text generation, and more in many domains like medical, education, software, and scientific research (OpenAI, 2022), (Liu et al., 2023). Meta AI also introduced an LLM called LLaMA that contains 65 billion parameters which outperformed GPT-3 on majority of assessments (Meta AI, 2023). No longer after that, the LLaMA 2 Model was introduced with LLAMA 2-Chat that is optimized for dialogue like ChatGPT (META GenAI, 2023). There are also open-source models shared publicly like Falcon-40 B and Falcon-180B that can be beneficial for public usage (Penedo et al., 2023).

2.5. Natural Language Processing (NLP) in Football Domain

NLP is a branch of artificial intelligence and computational linguistics that centers on the interaction between computers and human language (Jurafsky & Martin, 2000). It serves as a key tool among the many advanced technologies that football leverages.

NLP has a variety of applications in the football domain. It can be used to analyze the player performances and use this insight for scouting, planning tactics, and training players (Bera et al., 2023), summarizing and segmenting the events in football matches (Tang et al., 2018), analyzing specific news (Nguyen et al., 2014) and generating live commentary from the event data (Taniguchi et al., 2019).

CHAPTER 3

METHODOLOGY AND RESULTS

The importance of the crosses in football is mentioned in section 2.1. Furthermore, as mentioned in section 2.2, the notational analysis of the events can yield valuable insights into the performance of the players and teams. With these profundities, the notation and labeling of outcomes from a cross in football are crucial in determining the success of players and teams in utilizing their crossing skills.

The success of a cross relies on the characteristics it stated in section 2.1 such as its outcome, crossing position, and the opposing team's defenders' positions. Therefore, the notation of these features is highly advantageous for team decision-makers.

As outlined in section 2.1, a successful cross is achieved when it either reaches creating a scoring opportunity or remains untouched by a receiver despite a chance to or when it is cleared by an opponent through a vital challenge. While spatial-temporal analysis aids in identifying the first scenario (Wu et al., 2020), it falls on short in evaluating the latter two. Assessing the quality of the cross and the threat posed by the position demands human judgment to gain insights into these aspects. Such judgment can only be extracted when the play is watched, and observed, and the execution is confirmed as a successful cross.

This thesis proposes a new approach to determine the various features of a cross by extracting the information from the match commentaries in text format. In particular, the interval of the cross, the team who performs the crossing action, the player who makes the cross, the outcome of the cross, and the quality of the cross are the characteristics of the cross identified in this research. The extraction of these characteristics is intended to be facilitated by a Fine-Tuned Large Language Model (LLM), offering valuable insights to decision-makers in the football domain. To refine such a model, it is essential to create and label a carefully tailored dataset specifically designed for this purpose.

The method proposed for the purpose of this study comprises the following components:

1) Construction of the Raw Data
2) Labeling Process
3) Fine-Tuning the LLM Model

3.1. Construction of the Raw Data

The broadcast of sport competitions, especially football matches, has emerged as one of the most followed programs recently. (Lee et al., 2016). Historically, football commentaries have been delivered orally through radio and television. However, during the 2006 World Cup held in Germany, there was a notable emergence of written online commentaries on an international scale (Sabater et al., 2008). Like the commentaries on television and radio, online written commentaries portray diverse levels of excitement and emotions directly reflect emotions to the readers (Trouvain, 2011).

In this thesis, live commentaries sourced from goal.com are employed, focusing on matches from the Premier League 2022-2023 Season. The dataset consists of commentaries from a total of 304 matches, comprising 20.255 rows of commentary. The raw data includes information about each commentary's teams, match scores, match dates, and the specific minute when the commentary was made.

3.2. Labeling Process

Following the creation of the raw dataset detailed in the previous section, the labeling process is initiated to identify whether a commentary involves a crossing action. If such an action is identified, the labeling process determines the team executing the cross, the player who makes the cross, the cross' outcome, and its quality. To label the data, the Human Expert Labeling Process (HELP) introduced by Aslan et al. (2017) which is described in detail in section 2.3 is followed. The process contains four major stages: planning, labeler recruitment, labeling, and post-labeling.

3.2.1. Planning

Initially, the appropriate labels for annotating the commentaries are identified with a literature review. Subsequently, a label dictionary is formulated, encompassing the definitions for each label. These labels are then used on sample data to check their validity, and any new labels, not covered in the literature review, are incorporated. Following this, instructional materials for prospective labelers, along with accompanying handouts are prepared. After that, the selection of the labeling tool and the preparation of surveys for the labelers are finalized. Finally, the qualification of the prospective labelers is determined.

3.2.1.1. Literature Review on Labels

The commentary labeling process involves identifying whether there is a crossing action in the commentary. If there is, the process includes determining the team executing the cross, the player who makes the cross, the outcome of the cross, and its quality. Unlike the cross' outcome, the labeling tasks for the other features do not require a literature review as the labelers will determine labels as "Yes" or "No", Team 1" or "Team 2", the players' names and "Very Bad", "Bad", "Neutral", "Good" and "Very Good" as a quality respectively. The players subject to labeling are those included in the squad roster for each respective team during the Premier League 2022-2023 Season.

In contrast to other features, a systematic literature review was conducted to determine the accurate outcomes of crosses. Considering the lack of research focusing on the crosses in football, different KPIs and main match actions besides the crossing are also selected and investigated. The actions of outfield players and goalkeepers are investigated, and those involving delivery of the ball are considered in the review (Liu et al., 2013) (*Opta Event Definitions*, n.d.). Furthermore, the outcomes of "free kicks" and "corner kicks" are included as they entail the delivery of crosses as set pieces. Lastly, after the initial investigation of these parameters during the review, the keywords "notational analysis", "event", "effectiveness" and "performance analysis" keywords are included for the latter search. Consequently, key words in the review used are as follows:

- Cross
- Corner kick
- Free kick
- Shot
- Pass
- Transition
- Attack
- Possession
- Goalkeeper
- Defender/Defensive/Defence/Defense
- Action
- Event
- Effectiveness
- Performance analysis
- Notational analysis

The search strategy employed in the review was based on database searches, with inquiries conducted through the Web of Science (WOS), Scopus, Taylor & Francis Online, JSTOR, and ResearchGate databases, as well as Google Scholar in September 2023. The articles are selected using a systematic search approach as follows:

- The KPI parameters mentioned above (also gerund form versions of them like crossing, shooting, etc.) are included,
- The "outcome" keyword is included with the KPI parameter.
- "Football" or "Soccer" words are included (Tienza-Valverde et al., 2023)
- To focus on the "Cross" parameter, "cross-section" and "cross-sectional" keywords are excluded to narrow the scope.
- Topics related to "Medicine", "Psychology", "Biochemistry" and "Health Care" are excluded to filter articles covering injuries in football.
- Language is selected as "English".

After this search approach is applied in each database, a total of 15.015 documents are identified. After that, a review based on the title and abstract is conducted. Finally, these articles' complete texts are examined, and 42 articles are revealed that present possible outcomes of crossing.

The documents that include an event outcome are selected and the following variables for each are specified (Kitchenham et al., 2009):

- Reference (Author and Year)
- Title
- Document Type
- Journal or Institute Name
- Keywords of Document
- Founded Inquiry Key Words
- The described action outcomes in the document

The documents selected within the Systematic Literature Review and their variables are given in Table 2.

The outcomes for the following actions in football are derived in this systematic literature review:

- Cross Outcome
- Corner Kick Outcome
- Free kick Outcome
- Pass Outcome
- Event Outcome
- Shot Outcome
- Attacking-Offensive Action Outcome
- Goalkeeper Action Outcome
- Defender Action Outcome
- Transition Outcome
- Individual Action Outcome

After the review of these actions' outcomes, Table 3 is formed indicating the category of an outcome of the cross, definition of each category, the articles that this categorization and definition are made, and the attacking action defined in these articles. There are five main outcomes of a cross which are:

1) Cross on Target
2) Cross off Target
3) Defender Interception
4) Goalkeeper Interception
5) Referee Interception

Table 2 Reviewed Document List in Systematic Literature Review

#	Citation	Title	Context	Database	Type of Document	Journal/ Conference/ Institute	Keywords	Founded Keyword	Defined Outcome Parameter
1	(Alcock, 2010)	Analysis of direct free kicks in the women's football World Cup 2007	Describes the position and result of the free kicks during 2007 Women World Cup	WOS, Taylor & Francis Online	Journal Article	European Journal of Sport Science	Association football, soccer, set plays, performance analysis, video analysis	Goalkeeper Outcome, Free kick Outcome	Shot
2	(Almeida et al., 2014)	Effects of Match Location, Match Status and Quality of Opposition on Regaining Possession in UEFA Champions League	Describes the impact of match location, status, and the quality of opposition on regaining possession in UEFA Champions League matches.	ResearchGate	Journal Article	Journal of Human Kinetics	Soccer, notational analysis, situational variables, team performance, defensive strategies	Notational Analysis, Performance Analysis	Defensive Action, Goalkeeper Action
3	(Beare & Stone, 2019)	Analysis of Attacking Corner Kick Strategies in the FA Women's Super League 2017/2018	Describes how corner kicks are taken and their effectiveness during the 2017/2018 FA Women's Super League Season	WOS, Scopus, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	Performance analysis, set pieces, football, soccer	Corner Kick Outcome	Corner Kick
4	(Carmichael et al., 2000)	Team Performance: the Case of English Premiership Football	Describes the evaluation of team performance within the context of English Premiership Football, utilizing a new data source, analyzing various metrics and factors influencing overall team success in the league.	JSTOR	Journal Article	Managerial and Decision Economics	Not Determined	Cross Outcome	Event
5	(Castelão et al., 2014)	Comparison of tactical behavior and performance of youth soccer players in 3v3 and 5v5 small-sided games	Describes an assessment and contrast of the tactical conduct and effectiveness of soccer players in small-sided games	Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	soccer, tactical behavior, tactical performance, small-sided games	Action Outcome	Shot, Referee Interception, Shot, Defender Action
6	(Cordón-Carmona	What Is the Relevance in the Passing Action Between the Passer and the Receiver in	Describes the significance of passing interactions between the passer and receiver in elite La Liga soccer,	ResearchGate	Journal Article	International Journal of Environment	performance indicators, performance analysis,	Notational Analysis, Performan	Shot, Pass

Table 2 (cont.)

	et al., 2020)	Soccer? Study of Elite Soccer in La Liga	emphasizing the relevance of their connection in team dynamics and successful gameplay strategies.			al Research and Public Health	tactical behavior, soccer	ce Analysis	
7	(De Baranda & Lopez-Riquelme, 2012)	Analysis of Corner Kicks in Relation to Match Status in the 2006 World Cup	Describes the quantitative analysis of corner kicks including performance indicators of it such as effectiveness, goal zone and defense tactics during FIFA 2006 World Cup	Taylor & Francis Online	Journal Article	European Journal of Sport Science	Notational analysis, soccer, corner kick, World Cup, match status	Corner Kick Outcome, Notational Analysis	Corner Kick
8	(Decroos et al., 2019)	Actions Speak Louder Than Goals: Valuing Player Actions in Soccer	Describes a new language for player actions and a framework to valuate these actions	WOS	Conference Article	ACM SIGKDD International Conference on Knowledge Discovery & Data Mining	sports analytics, event stream data, soccer match data, valuing actions, probabilistic classification	Event Outcome, Defensive Outcome	Individual Action
9	(Delgado-Bordonau et al., 2013)	Offensive and defensive team performance: relation to successful and unsuccessful participation in the 2010 Soccer World Cup	Describes the effect of various KPIs on teams' success during the FIFA 2010 World Cup.	WOS	Journal Article	Journal of Human Sport and Exercise	soccer, game-related statistics, scoring effectiveness, first goal effect, match analysis	Defensive Outcome	Shot
10	(Gómez et al., 2018)	Analysis of Playing Styles According to Team Quality and Match Location in Greek Professional Soccer	Describes the identification of different play styles of Greek Superleague teams	WOS	Journal Article	International Journal of Performance Analysis in Sport	Football, factor analysis, styles of play, performance indicators, situational variables	Transition Outcome	Attack, Transition, Defender Action
11	(Hughes & Franks 2004)	Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport	Describes how notational analysis systems in sports enhance coaching methodologies and overall athletic performance.	ResearchGate	Book	Journal of Sports Science and Medicine	Not Determined	Notational Analysis, Defensive Outcome	Shot
12	(Hughes & Lovell, 2019)	Transition To Attack in Elite Soccer	Describes the strategic process of transitioning from defense to attack in football, emphasizing the critical tactics	WOS	Journal Article	Journal of Human Sport and Exercise	Champion's League, Transitions, Turnover, Soccer	Possession Outcome, Transition Outcome	Transition

Table 2 (cont.)

			and methods employed during this pivotal phase of play.						
13	(Kim et al., 2019)	Determining Unstable Game States to Aid the Identification of Perturbations in Football	Describes utilizing algorithms to detect unstable game states in football, facilitating the identification of disruptions or perturbations within match dynamics.	WOS	Journal Article	International Journal of Performance Analysis in Sport	Perturbations, unstable situations, game states, football	Transition Outcome	Cross, Shot
14	(Kubayi & Larkin, 2019)	Analysis of teams' corner kicks defensive strategies at the FIFA World Cup 2018	Describe the defense strategies and analysis of the corner kicks during FIFA 2018 World Cup	Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	set pieces, marking, defending, goals	Corner Kick Outcome	Goalkeeper, Defense, Shot
15	Kvesić et al., 2017)	Analysis of Crosses in the Croatian First Football League	Describes the ball clearance efficiency by the analysis of the matches from the First Croatian Football League	ResearchGate	Journal Article	Acta Kinesiologica	Not Determined	Cross Outcome	Defensive Action
16	(Lee & Mills, 2021)	Analysis of Corner Kicks at the FIFA Women's World Cup 2019 in Relation to Match Status and Team Quality	Describes the analysis and characteristics of corner kicks during FIFA women's World Cup 2019.	WOS, Scopus, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	Performance analysis, set pieces, football, soccer, corner kicks	Corner Kick Outcome	Corner Kick
17	(Link et al., 2016)	A Topography of Free Kicks in Soccer	Describes the performance variables for free kicks during 2013-2015 German Bundesliga League	Taylor & Francis Online	Journal Article	Journal of Sports Science	Performance analysis, free kick, soccer, topography, variography	Free kick outcome	Free Kick
18	(Liu et al., 2015)	Match Performance Profiles of Goalkeepers of Elite Football Teams	Describes the observation of the goalkeepers' performance during 2012/2013 Spanish First Division	WOS	Journal Article	International Journal of Sports Science & Coaching	Association Football, Performance Analysis, Soccer, Sport Analytics	Goalkeeper Outcome	Goalkeeper
19	(Liu et al., 2016)	Modelling Relationships Between Match Events and Match Outcome in Elite Football	Describes the construction of models that establish correlations between specific match events and the outcome in a football match, aiming to predict or	Taylor & Francis Online	Journal Article	European Journal of Sport Science	Notational analysis, performance indicators, situational variable, soccer	Notational Analysis	Event

Table 2 (cont.)

			understand how certain actions influence the overall game result.						
20	(Lorains et al., 2013)	Performance Analysis for Decision Making in Team Sports	Describes the design and test of a reliable method of analyzing decision-making performance in team sports and transfer of training into competition matches.	WOS	Journal Article	International Journal of Performance Analysis in Sport	Australian football, in-game measurement, decision making, transfer	Notational Analysis, Effectiveness	Pass
21	(Mara et al., 2012)	Attacking Strategies That Lead to Goal Scoring Opportunities in High Level Women's Football	Describes the specific attacking strategies employed in high-level women's football that effectively create goal-scoring opportunities, highlighting the tactical approaches utilized to generate scoring chances.	WOS	Journal Article	International Journal of Sports Science & Coaching	Association Football, Attacking Strategies, Performance Analysis, Women's Soccer	Attacking Outcome, Performance Analysis	Attack
22	(Mitrotasios et al., 2021)	Analysis of Corner Kick Success in Laliga Santander 2019/2020	Describes the effectiveness and identify KPIs associated with the outcomes of corner kicks during 2019/2020 Laliga Santander	ResearchGate, Scopus	Journal Article	European Journal of Human Movement	Key performance indicators, set plays, observational methodology	Corner Kick Outcome	Corner Kick
23	(Mitrotasios et al., 2021)	Analysis of Crossing Opportunities at the 2018 FIFA World Cup	Describes the investigation of open play crosses during FIFA 2018 World Cup.	ResearchGate	Journal Article	Montenegrin Journal of Sports Science and Medicine	cross outcome, match status, attacking, goal-scoring	Cross Outcome	Cross
24	(O'Donoghue et al., 2012)	Statistical Methods in Performance Analysis: An Example from international Soccer	Describes the utilization of statistical methods in performance analysis, using an international football example to show their application in evaluating player or team performance.	WOS	Journal Article	International Journal of Performance Analysis in Sport	nonparametric tests, generalization	Performance Analysis	Attack
25	(Peterson & Bruton, 2020)	A Review of the interaction Between the Striker and the Goalkeeper at the individual Tactical Level in Football	Describes the goalkeeper actions and the interactions between goalkeeper and striker	WOS	Journal Article	International Journal of Sports Science & Coaching	Association football, performance analysis, sport analytics, soccer	Performance Analysis	Goalkeeper Action

Table 2 (cont.)

26	(Pollard & Reep, 1997)	Measuring the Effectiveness of Playing Strategies at Soccer	Describes the methods to measure the effectiveness of team possession in football, aiming to quantify their impact on team performance and outcomes using a notational system.	JSTOR	Journal Article	Journal of the Royal Statistical Society	Logistic regression, Performance analysis, Soccer, Strategy	Event Outcome, Performance Analysis	Pass
27	(Pulling & Newton, 2017)	Defending Corner Kicks in the English Premier League: Near-Post Guard Systems	Describes the usage and investigation of near-post guard system when defending the corner kicks in English Premier League during 2015/2016 Season	WOS, Scopus, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	tactics, strategy, coaching, set play, set piece, performance analysis	Performance Analysis	Corner Kick
28	(Pulling et al., 2013)	Defending Corner Kicks: Analysis From the English Premier League	Describes the tactical behavior when defending the corner kicks in English Premier League	WOS, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	notational analysis, soccer, corner kicks, defending	Corner Kick Outcome	Corner Kick
29	(Pulling et al., 2018)	Analysis of Crossing at the 2014 FIFA World Cup	Describes the analysis of open play crosses during FIFA 2014 World Cup including delivery side, delivery type, defensive pressure, time of cross and delivery outcome of the crosses	WOS, Scopus, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	Performance analysis, tactics, strategy, coaching, attacking, defending	Cross Outcome	Cross
30	(Pulling, 2015)	Long Corner Kicks in the English Premier League: Deliveries into the Goal Area and Critical Area	Describes the dynamics of long corner kicks in the English Premier League, focusing on the types of deliveries into the goal area and critical zones, highlighting their significance in creating scoring opportunities.	ResearchGate, Scopus	Journal Article	Kinesiology	performance analysis, notational analysis, soccer	Corner Kick Outcome	Corner Kick
31	(Reilly, 2005)	Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance	Describes a comprehensive guide, employing a systematic approach to enhance performance through soccer match analysis.	Taylor & Francis Online	Book	Taylor & Francis	Not Determined	Shot Outcome	Shot
32	(Ruiz-Vanoye et al., 2017)	Motivation index to Improve the Soccer Performance	Describes the creation of a motivation index tailored for enhancing soccer performance by measuring and leveraging motivational factors.	ResearchGate	Journal Article	International Journal of Combinatorial Optimization	Sports Performance, statistical indicators, Motivation Index, soccer performance	Corner Kick Outcome	Corner Kick

Table 2 (cont.)

						Problems and Informatics			
33	(Sainz de Baranda et al., 2019)	Differences in the offensive and Defensive Actions of the Goalkeepers at Women's FIFA World Cup 2011	Describes the analysis of offensive and defensive strategies employed by goalkeepers in the Women's FIFA World Cup 2011 matches.	ResearchGate	Journal Article	Frontiers in Psychology	women's football, match statistics, notational analysis, performance indicators, soccer	Notational Analysis	Goalkeeper Action
34	(Sarmento et al., 2018)	Influence of Tactical and Situational Variables on offensive Sequences During Elite Football Matches	Describes how tactical and situational factors affect offensive sequences in football matches, examining their influence on team strategies and gameplay patterns.	ResearchGate	Journal Article	Journal of Strength and Conditioning Research	soccer, notational analysis, match analysis, goal scoring	Notational Analysis	Performance Analysis
35	(Stein et al., 2017)	How To Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects	Describes the step-by-step process of understanding team sport data, encompassing acquisition, data modeling, and research aspects to derive meaningful insights.	WOS	Journal Article	Data	sport analytics, visual analytics, high frequency spatial-temporal data	Performance Analysis	Event
36	(Strafford et al., 2019)	Comparative Analysis of the Top Six and Bottom Six Teams' Corner Kick Strategies in the 2015/2016 English Premier League	Describes the corner kick strategies used by the top six and bottom teams in English Premier League during 2015/2016 Season	WOS, Scopus, Taylor & Francis Online	Journal Article	International Journal of Performance Analysis in Sport	Soccer, observational methodology, performance analysis, set pieces	Attacking Outcome	Corner Kick
37	(Tenga et al., 2010)	Effect of Playing Tactics on Achieving Score-Box Possessions in A Random Series of Team Possessions from Norwegian Professional Soccer Matches	Describes how specific playing tactics influence the attainment of score-box possessions in random sequences of team possessions in Norwegian League matches.	Taylor & Francis Online	Journal Article	Journal of Sports Science	Validity, opponent interaction, logistic regression, soccer playing effectiveness, match-performance analysis	Possession Outcome	Possession
38	(Tijsen, 2018)	Analyzing offensive Player-And Team Performance in Soccer Using Position Data	Describe the offense players' and teams' performances by a method using the chance of goal scoring probabilities distributed to the parts of the football pitch.	Google Scholar	Bachelor Thesis	Leiden Institute of Advanced Computer Science (LIACS)	Not Determined	Event Outcome	Event

Table 2 (cont.)

39	(Turner & Sayers, 2010)	The influence of Transition Speed on Event Outcomes in a High-Performance Football Team	Describes how transition speed impacts event outcomes within a high-performance football team, highlighting the correlation between the pace of transitions and subsequent match events.	WOS	Journal Article	International Journal of Performance Analysis in Sport	performance analysis	Event Outcome, Performan ce Analysis	Transition
40	(van Maarsev een et al., 2017)	System For Notational Analysis in Small-Sided Soccer Games	Describes the development of a notational analysis system designed for small-sided football games, aiming to capture and analyze key performance indicators in this specific setting.	WOS	Journal Article	International Journal of Sports Science & Coaching	Association football, performance analysis, sport analytics	Notational Analysis, Performan ce Analysis	Event
41	(Yi et al., 2020)	Evaluation of the Technical Performance of Football Players in the UEFA Champions League	Describes how player performance is significantly influenced by team quality, opponent strength, and match outcome, highlighting key areas like goal scoring, passing, and organizing	WOS	Journal Article	International Journal of Environment al Research and Public Health	Technical performance profile, situational variable, playing position, football, soccer, match analysis	Defensive Outcome	Shot, Defensive Action, Attack
42	(Zileli & Söyler, 2020)	Analysis of Corner Kicks in FIFA 2018 World Cup	Describes the analysis of corner kicks in terms of direction, time interval, during FIFA 2018 World Cup.	WOS	Journal Article	Journal of Human Sport and Exercise	Football, Match analysis, Set play.	Attacking Outcome	Corner Kick

Table 3 Cross Outcome Categories Defined in Literature

Cross Outcome	**Definition**	**Authors**	**Action Outcomes Defined in Articles**
1. Cross on target, reaches a teammate	Cross is controlled by an attacker at the target and he made a shot with kick or head.	(Mara et al., 2012), (Beare & Stone, 2019), (van Maarseveen et al., 2017), (Hughes & Franks 2004), (Carmichael et al., 2000)	Attack, Pass, Corner Kick, Event
1.1. Cross Receiver Shoots or Heads	Cross is controlled by an attacker at the target and he made a shot with kick or head.	(Mara et al., 2012), (Link et al., 2016), (Liu et al., 2016), (Hughes & Franks 2004), (Sarmento et al., 2018), (Carmichael et al., 2000), (Cordón-Carmona et al., 2020), (Gómez et al., 2018), (Yi et al., 2020), (Castelão et al., 2014)	Attack, Free kick, Shot, Cross, Pass, Event
1.1.1. Cross Receiver Scores, Goal	Cross is controlled by an attacker at the target and he made a shot. The ball went over the goal line inside the dimensions of the goalposts. The referee awarded a goal.	(Mitrotasios et al., 2021), (Pulling et al., 2018), (Pulling, 2015), (Strafford et al., 2019), (De Baranda & Lopez-Riquelme, 2012), (Link et al., 2016), (Lee & Mills, 2021), (Gómez et al., 2018), (Stein et al., 2017), (Lorains et al., 2013), (Reilly, 2005), (Sarmento et al., 2018), (Pulling et al., 2013), (Cordón-Carmona et al., 2020), (Hughes & Franks 2004), (Hughes & Lovell, 2019), (Alcock, 2010), (Yi et al., 2020), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Free kick, Shot, Transition, Pass
1.1.2. Cross receiver's shot hits goalpost, woodwork	Cross is controlled by an attacker at the target and he made a shot. The ball hits goalpost, woodwork.	(De Baranda & Lopez-Riquelme, 2012), (Link et al., 2016), (Reilly, 2005), (van Maarseveen et al., 2017), (Gómez et al., 2018), (Strafford et al., 2019), (Beare & Stone, 2019)	Corner Kick, Shot, Attack, Free kick, Cross
1.1.3. Cross receiver's shot blocked by defender	Cross is controlled by an attacker at the target and he made a shot. The ball hits or blocked by a defender.	(De Baranda & Lopez-Riquelme, 2012), (Lee & Mills, 2021), (Pulling et al., 2013), (van Maarseveen et al., 2017), (Liu et al., 2016), (Cordón-Carmona et al., 2020), (Gómez et al., 2018), (Hughes & Franks 2004), (Link et al., 2016), (Alcock, 2010), (Yi et al., 2020)	Corner Kick, Shot, Attack, Free kick
1.1.4. Cross receiver's shot saved by goalkeeper	Cross is controlled by an attacker at the target and he made a shot. The shot is saved by goalkeeper.	(De Baranda & Lopez-Riquelme, 2012), (Lee & Mills, 2021), (Pulling, 2015), (Link et al., 2016), (Reilly, 2005), (Hughes & Lovell, 2019), (Pollard & Reep, 1997), (van Maarseveen et al., 2017), (Sainz de Baranda et al., 2019), (Kubayi & Larkin, 2019) (Sarmento et al., 2018), (Beare & Stone, 2019), (Cordón-Carmona et al., 2020), (Liu et al., 2015), (Alcock, 2010)	Corner Kick, Shot, Free kick, Cross, Pass
1.1.5. Cross receiver's shot goes out, attempt off target	Cross is controlled by an attacker at the target and he made a shot. Shot goes out, attempt off target.	(Mitrotasios et al., 2021), (Hughes & Franks 2004), (De Baranda & Lopez-Riquelme, 2012), (Pulling, 2015), (Pulling et al., 2013), (Link et al., 2016), (Tijsen, 2018), (Lee & Mills, 2021), (Beare & Stone, 2019), (Pulling & Newton, 2017), (Ruiz-Vanoye et al., 2017), (van Maarseveen et al., 2017),	Cross, Shot, Corner Kick, Free kick, Event, Goalkeeper Action, Pass

Table 3 (cont.)

	Ball not directed within the dimensions of the goal.	(Sarmento et al., 2018), (Cordón-Carmona et al., 2020), (Gómez et al., 2018), (Strafford et al., 2019), (Stein et al., 2017), (Almeida et al., 2014), (Alcock, 2010), (Kubayi & Larkin, 2019)	
1.2. Cross Receiver Passes or Crosses to a teammate	Cross is controlled by an attacker at the target and he passed/crossed the upcoming ball to a teammate.	(Link et al., 2016), (Lorains et al., 2013), (Lorains et al., 2013), (O'Donoghue et al., 2012), (Stein et al., 2017), (Strafford et al., 2019), (Carmichael et al., 2000), (Decroos et al., 2019), (Gómez et al., 2018), (Yi et al., 2020)	Free kick, Cross, Pass, Attack, Corner Kick, Individual Action
1.3. Cross Receiver dribbles	Cross is controlled by an attacker at the target and he starts dribbling the ball.	(Carmichael et al., 2000), (Yi et al., 2020)	Event
1.4. Cross Receiver cannot control or touch the ball, lost the possession	Cross reaches an attacker at the target and but he cannot control or touch the ball and team lost possession. Unsuccessful attacking action.	(Mara et al., 2012), (Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Carmichael et al., 2000), (Pulling et al., 2018), (Pulling & Newton, 2017), (Decroos et al., 2019), (Kim et al., 2019), (Link et al., 2016), (Yi et al., 2020)	Attack, Cross, Event, Individual Action, Free kick
2. Cross off target, not reaches a teammate	Cross does not reach an attacker.	(Hughes & Franks 2004), (Hughes & Lovell, 2019), (Ruiz-Vanoye et al., 2017), (Kvesić et al., 2017), (Lorains et al., 2013), (Cordón-Carmona et al., 2020)	Cross, Attack, Pass, Event
2.1. Team retains possession, ball is recycled	Cross does not reach an attacker at the target and the ball exits the box but attacking team retains possession, recycles the ball.	(De Baranda & Lopez-Riquelme, 2012), (Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Carmichael et al., 2000), (Pulling & Newton, 2017), (van Maarseveen et al., 2017), (Lorains et al., 2013), (Cordón-Carmona et al., 2020), (O'Donoghue et al., 2012), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Free kick, Event, Pas, Attack
2.2. Cross goes out of pitch/play	Cross does not reach an attacker at the target and ball goes out of pitch/play.	(Turner & Sayers, 2010), (Hughes & Franks 2004), (Pollard & Reep, 1997), (Lee & Mills, 2021), (Lee & Mills, 2021), (van Maarseveen et al., 2017), (Lorains et al., 2013), (Strafford et al., 2019)	Cross, Corner Kick, Pass
2.3. Cross has no contact in the box from both teams	Cross is not controlled by any player and the ball exited the box.	(Mitrotasios et al., 2021), (Pulling et al., 2013), (Pulling & Newton, 2017), (Lorains et al., 2013), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Pass
3. Defender Interception	Defender interception towards cross.	(Link et al., 2016), (Pollard & Reep, 1997), (Lee & Mills, 2021), (Hughes & Lovell, 2019), (van Maarseveen et al., 2017), (Decroos et al., 2019), (Gómez et al., 2018), (Strafford et al., 2019), (van Maarseveen et al., 2017), (Almeida et al., 2014)	Cross, Pass, Free kick, Shot, Corner Kick, Event, Defensive Action, Individual Action
3.1. Defensive block, tackle to crosser	Cross is blocked/intercepted/tackled by a defender immediately after ball	(Turner & Sayers, 2010), (Hughes & Franks 2004), (Link et al., 2016), (Pollard & Reep, 1997), (Hughes & Lovell, 2019), (Carmichael et al., 2000), (Ruiz-Vanoye et al., 2017), (Almeida et al., 2014), (Liu et al., 2016), (Decroos et al., 2019), (Gómez et al., 2018), (Yi et al., 2020)	Cross, Free kick, Shot, Attack, Event, Defensive Action, Individual Action

Table 3 (cont.)

	comes out of the foot of the attacker.		
3.2. Defender takes control of the ball upcoming from cross	Defender takes control of the ball upcoming from cross.	(De Baranda & Lopez-Riquelme, 2012), (Lee & Mills, 2021), (Hughes & Lovell, 2019), (Sarmento et al., 2018), (Liu et al., 2016), (Lorains et al., 2013), (Cordón-Carmona et al., 2020)	Corner Kick, Event, Pass
3.3. Defensive clearance to upcoming cross with kick, head etc.	Defender clears the ball upcoming from the cross with a kick, head etc.	(De Baranda & Lopez-Riquelme, 2012), (Mitrotasios et al., 2021), (Carmichael et al., 2000), (Pulling & Newton, 2017), (Decroos et al., 2019), (Zileli & Söyler, 2020), (Stein et al., 2017), (Yi et al., 2020)	Cross, Corner Kick, Event, Individual Action
3.3.1. To corner	Cross is cleared to corner by a defender.	(Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Pulling & Newton, 2017), (van Maarseveen et al., 2017), (Sarmento et al., 2018), (Cordón-Carmona et al., 2020), (Zileli & Söyler, 2020), (Kim et al., 2019), (Gómez et al., 2018), (Stein et al., 2017), (Pulling et al., 2018), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Free kick, Pass, Event
3.3.2. To throw in	Cross is cleared to throw in by a defender.	(Mitrotasios et al., 2021), (Stein et al., 2017), (Pulling et al., 2018)	Cross, Event
3.3.3. Out of box	Cross is cleared to out of box by a defender.	(Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Pulling & Newton, 2017), (Kubayi & Larkin, 2019)	Cross, Corner Kick
3.4. Defender's interception leads ball own goal	Defender's interception towards cross leads to own goal.	(Pulling et al., 2018), (Castelão et al., 2014)	Cross
4. Goalkeeper Interception	Goalkeeper interception towards cross.	(Ruiz-Vanoye et al., 2017), (Zileli & Söyler, 2020), (Peterson & Bruton, 2020), (Liu et al., 2015)	Cross, Corner Kick
4.1. Goalkeeper catches or gathers the ball upcoming from cross	The goalkeeper catches or gathers the ball and hold it in his hands from the cross.	(Hughes & Franks 2004), (Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Beare & Stone, 2019), (Pulling & Newton, 2017), (Ruiz-Vanoye et al., 2017), (Decroos et al., 2019), (Peterson & Bruton, 2020), (Stein et al., 2017), (Lee & Mills, 2021), (Liu et al., 2015), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Individual Action
4.2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	The goalkeeper clears the ball with punch, fist, slap, touch away.	(De Baranda & Lopez-Riquelme, 2012), (Mitrotasios et al., 2021), (Pulling, 2015), (Pulling et al., 2013), (Lee & Mills, 2021), (Beare & Stone, 2019), (Pulling & Newton, 2017), (Ruiz-Vanoye et al., 2017), (Sainz de Baranda et al., 2019), (Decroos et al., 2019), (Peterson & Bruton, 2020), (Stein et al., 2017), (Liu et al., 2015), (Kubayi & Larkin, 2019)	Cross, Corner Kick, Event, Individual Action
5. Referee Interception	Referee stops the game because of the violation of the rules.	(Pollard & Reep, 1997), (Cordón-Carmona et al., 2020), (Stein et al., 2017), (Cordón-Carmona et al., 2020)	Event, Shot, Pass

Table 3 (cont.)

1. Penalty	A player on the defending team committed a foul and the referee awarded a penalty.	(Beare & Stone, 2019), (Pulling & Newton, 2017), (Ruiz-Vanoye et al., 2017), (Sarmento et al., 2018), (Cordón-Carmona et al., 2020), (Decroos et al., 2019), (Stein et al., 2017), (Strafford et al., 2019), (Carmichael et al., 2000), (Kubayi & Larkin, 2019)	Corner Kick, Event, Pass, Event, Individual Action
2. Offside	Defending team win a free kick after a player on the attacking team is ruled offside during the cross.	(Turner & Sayers, 2010), (Pollard & Reep, 1997), (Lee & Mills, 2021), (Ruiz-Vanoye et al., 2017), (van Maarseveen et al., 2017), (Almeida et al., 2014), (Liu et al., 2016), (Stein et al., 2017), (Carmichael et al., 2000), (Decroos et al., 2019), (Hughes & Lovell, 2019), (Yi et al., 2020)	Cross, Corner Kick, Defensive Action, Goalkeeper Action, Event, Individual Action, Transition
3. Free kick for attacking team (direct/indirect)	The referee awarded a free kick to the attacking team.	(Pollard & Reep, 1997), (Lee & Mills, 2021), (Ruiz-Vanoye et al., 2017), (van Maarseveen et al., 2017), (Almeida et al., 2014), (Sarmento et al., 2018), (Liu et al., 2016), (Cordón-Carmona et al., 2020), (Decroos et al., 2019), (Gómez et al., 2018), (Gómez et al., 2018), (O'Donoghue et al., 2012), (Stein et al., 2017), (Strafford et al., 2019), (Hughes & Lovell, 2019), (Yi et al., 2020), (Castelão et al., 2014)	Corner Kick, Event, Shot, Pass, Attack, Individual Action, Transition
4. Free kick for defending team (direct/indirect)	The referee awarded a free kick to the defending team.	(Pulling et al., 2013), (Pollard & Reep, 1997), (Lee & Mills, 2021), (Beare & Stone, 2019), (Pulling & Newton, 2017), (Almeida et al., 2014), (Hughes & Lovell, 2019), (Yi et al., 2020), (Kubayi & Larkin, 2019), (Castelão et al., 2014)	Corner Kick, Shot, Transition

3.2.1.2. Label Dictionary

After the completion of the systematic literature review for outcome of a cross, labeling dictionary is created in Table 4. Consequently, it is the dictionary of cross outcome formed after literature review.

Table 4 Labeled Features Dictionary – Initial

Labeled Feature	**Definition of the Feature**	**Labels**
Is there a Cross	If the commentary contains a crossing action is labeled.	Yes, No
Crossing Team	If there is a crossing action, this feature will be labeled. The team which performs the crossing action is labeled.	Team_1, Team_2 (Changes depending on the match)
Crossing Player	If there is a crossing action, this feature will be labeled. The player which performs the crossing action is labeled.	Selected from the players that belongs to the squad labeled in "Crossing Team" feature.
Outcome of Cross	If there is a crossing action, this feature will be labeled. The outcome of the cross is labeled.	Selected from the lists of outcomes in Table 5
Quality of Cross	If there is a crossing action, this feature will be labeled. The quality of the cross is labeled.	Very Bad, Bad, Neutral, Good and Very Good

Table 5 Cross Outcome Dictionary – Initial

Categorized Outcome	**Definition**
The cross receiver scores, goal	Cross controlled by an attacker at the target and he made a shot. The ball went over the goal line inside the dimensions of the goalposts. The referee awarded a goal.
The cross receiver's shot hits goalpost, woodwork	Cross controlled by an attacker at the target and he made a shot. The ball hits goalpost, woodwork
The cross receiver's shot is blocked by defender	Cross controlled by an attacker at the target and he made a shot. Shot is blocked by defender
The cross receiver's shot saved by goalkeeper	Cross controlled by an attacker at the target and he made a shot. The shot is saved by goalkeeper.
The cross receiver's attempt off target, shot goes out	Cross controlled by an attacker at the target and he made a shot. Shot goes out, attempt off target. Ball not directed within the dimensions of the goal.
The cross receiver passes or crosses to a teammate	Cross is controlled by an attacker at the target and he passed/crossed the upcoming ball to a teammate.

Table 5 (cont.)

The cross receiver dribbles	Cross is controlled by an attacker at the target and he starts dribbling the ball.
The cross receiver cannot control or touch the ball, misses the ball, lost possession	Cross not controlled by an attacker at the target and team lost possession. Unsuccessful attacking action
Team retains possession, ball is recycled	Cross not controlled by an attacker at the target and the ball exits the box but attacking team retains possession, recycles the ball.
Cross goes out of pitch/play	Cross not controlled by an attacker at the target and ball goes out of pitch/play.
Cross has no contact in the box from both teams	Cross not controlled by any player and the ball exited the box
Defensive block, interception, tackle to crosser	Cross is blocked/intercepted/tackled by a defender immediately
Defender takes control of the ball from cross	Defender takes control of the ball coming from cross
Defensive clearance to cross - to corner	Cross is cleared to corner by a defender
Defensive clearance to cross - to throw in	Cross is cleared to throw in by a defender
Defensive clearance to cross - out of box	Cross is cleared to out of box by a defender
Defender scores an own goal	Defender's interception causes an own goal
Goalkeeper catches/gathers the ball from cross	The goalkeeper catches/gathers the ball and hold it in his hands from the cross
Goalkeeper clearance to cross - to corner	The goalkeeper clears the high ball with punch, fist, slap, touch away to corner
Goalkeeper clearance to cross - to throw in	The goalkeeper clears the high ball with punch, fist, slap, touch away to throw in
Goalkeeper clearance to cross - out of box	The goalkeeper clears the high ball with punch, fist, slap, touch away to out of box
Penalty	A player on the defending team committed a foul and the referee awarded a penalty.
Offside	Defending team win a free kick after a player on the attacking team is ruled offside during the cross
Free kick for attacking team (direct/indirect)	The referee awarded a free kick to the attacking team
Free kick for defending team (direct/indirect)	The referee awarded a free kick to the defending team
Not Determined	The action cannot be determined

In addition to the labels provided in the Table 5, “Not Sure” label has been included for all features in instances where the labelers are unable to determine the appropriate label to select.

3.2.1.3. Validation of Determined Labels

Once the labels are assigned, a sample dataset featuring commentary from two matches is employed for validation. Following the investigation, it was determined that subdividing the outcome of the cross would expedite labeling and support the validity of the cross outcomes. Furthermore, the need for new cross outcomes arises within the dataset. This necessity arises from both the absence of definitive labels to determine the outcome of a cross and the requirement for a more detailed elaboration on it. These measures aim to establish a more refined judgment regarding the quality of the outcomes. Lastly, it is observed that certain commentaries contain a second action that also involves a crossing action. Consequently, the labels are replicated to identify and assess the presence of this secondary cross, aiming to determine its features. As it is a repetitive action, the team is not asked to be labeled, it is assumed to be the same as the first cross’ team.

The following outcomes are observed during the investigation of the sample dataset and are added to cross outcome dictionary:

- New outcomes are defined and categorized under 1.1. Cross Receiver Shoots or Heads:
    - <u>1.1.6. Cross receiver's shot hits a teammate</u>: Cross is controlled by an attacker at the target, and he makes a shot. The hot is blocked by a teammate.

- New outcomes are defined and categorized under 1.2. Cross Receiver Passes or Crosses to a teammate:
    - <u>1.2.1. Passed teammate shoots and Scores, Goal</u>: Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate scores.
    - <u>1.2.2. Passed teammate shoots and misses or goalkeeper saves or defender blocks</u>: Cross is controlled by an attacker at the target, and he passes to a teammate. The teammate shoots and misses or the goalkeeper saves.
    - <u>1.2.3. Passed teammate passes to another teammate</u>: Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate passes to another teammate.
    - <u>1.2.4. Passed teammate cannot control the ball:</u> Cross is controlled by an attacker at the target, and he passes to a teammate. The teammate cannot control the ball.
    - <u>1.2.5. Pass is intercepted by defender or goalkeeper</u>: Cross is controlled by an attacker at the target, and he passes to a teammate. The Defense/Goalkeeper intercepts the passed ball.

- 1.2.6. Passed teammate passes to an opponent or out of play: Cross is controlled by an attacker at the target, and he passes to a teammate. It arrives at an opponent or goes out of play.

- New outcomes are defined and categorized under 3.3. Defensive clearance to upcoming cross with kick, head etc.:
  - 3.3.4. Direction is not mentioned: Cross is cleared by a defender. Direction is not mentioned.
  - 3.3.5. To Opponent: Cross is cleared to an opponent by a defender.
  - 3.3.6. Inside penalty area: Cross is cleared inside the penalty area by a defender.
  - 3.3.7. To a teammate: Cross is cleared to a teammate by a defender.

- New outcomes are defined and categorized under 3. Defender Interception:
  - 3.5. Defender's interception leads ball to opponent: Defender's interception towards cross find its way to an opponent.

- New outcomes defined and categorized under 4. Goalkeeper Interception:
  - 4.3. Goalkeeper's interception leads ball to the opponent: Goalkeeper's interception toward cross finds its way to an opponent.
  - 4.4. Goalkeeper's interception leads to ball own goal: Goalkeeper's interception towards cross leads to own goal.

- New outcomes are defined and categorized under 4.2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch:
  - 4.2.1. To corner: The goalkeeper clears the ball with a punch, fist, slap, touch away to corner.
  - 4.2.2. To throw in: The goalkeeper clears the ball with a punch, fist, slap, touch away to throw in.
  - 4.2.3. Out of box: The goalkeeper clears the ball with a punch, fist, slap, touch away to out of box.
  - 4.2.4. Direction is not mentioned: The goalkeeper clears the ball with punch, fist, slap, touch away. Direction is not mentioned.
  - 4.2.5. To Opponent: The goalkeeper clears the ball with punch, fist, slap, touch away to an opponent.
  - 4.2.6. Inside penalty area: The goalkeeper clears the ball with punch, fist, slap, touch inside penalty area by a defender.
  - 4.2.7. To a teammate: The goalkeeper clears the ball with punch, fist, slap, touch to a teammate by a defender.

After the insertion of these new features and dividing the outcome into subcategories, the dictionaries are formed as follows in Table 6 and for outcome of cross in Table 7.

Table 6 Labeled Features Dictionary - Final

Labeled Feature	**Definition of the Feature**	**Labels**
Is there a Cross	If the commentary contains a crossing action is labeled.	Yes, No
Crossing Team	If there is a crossing action, this feature will be labeled. The team which performs the crossing action is labeled.	Team_1, Team_2 (Changes depending on the match)
Crossing Player	If there is a crossing action, this feature will be labeled. The player which performs the crossing action is labeled.	Selected from the players that belongs to the squad labeled in "Crossing Team" feature.
Outcome of Cross – Category 1	If there is a crossing action, this feature will be labeled. The category 1 outcome of the cross is labeled.	1. Cross on Target, 2. Cross off target, 3. Defender Interception, 4. Goalkeeper Interception and 5. Referee Interception.
Outcome of Cross – Category 2	If there is a crossing action, this feature will be labeled. The category 2 outcome of the cross is labeled.	Selected from the lists of outcomes in Table 7
Outcome of Cross – Category 3	If there is a crossing action, this feature will be labeled. The category 3 outcome of the cross is labeled.	Selected from the lists of outcomes in Table 7
Quality of Cross	If there is a crossing action, this feature will be labeled. The quality of the cross is labeled.	Very Bad, Bad, Neutral, Good and Very Good
Is there a Second Cross	If the commentary contains a second crossing action is labeled.	Yes, No
Crossing player 2	If there is a second crossing action, this feature will be labeled. The player which performs the second crossing action is labeled.	Selected from the players that belongs to the squad labeled in "Crossing Team" feature.
Outcome of Cross 2– Category 1	If there is a second crossing action, this feature will be labeled. The category 1 outcome of the second cross is labeled.	1. Cross on Target, 2. Cross off target, 3. Defender Interception, 4. Goalkeeper Interception and 5. Referee Interception.

Table 6 (cont.)

Outcome of Cross 2– Category 2	If there is a second crossing action, this feature will be labeled. The category 2 outcome of the second cross is labeled.	Selected from the lists of outcomes in Table 7
Outcome of Cross 2– Category 3	If there is a crossing action, this feature will be labeled. The category 2 outcome of the cross is labeled.	Selected from the lists of outcomes in Table 7
Quality of Cross	If there is a crossing action, this feature will be labeled. The quality of the cross is labeled.	Very Bad, Bad, Neutral, Good and Very Good

Table 7 Cross Outcome Dictionary – Final

1. Cross on target, reaches a teammate	**Cross reaches an attacker at the target**
1. Cross Receiver Shoots/Heads	Cross is controlled by an attacker at the target, and he made a shot with kick or head.
1. Cross Receiver Scores, Goal	Cross is controlled by an attacker at the target, and he made a shot. The ball went over the goal line inside the dimensions of the goalposts. The referee awarded a goal.
2. Cross receiver's shot hits goalpost, woodwork	Cross is controlled by an attacker at the target, and he made a shot. The ball hits goalpost, woodwork
3. Cross receiver's shot blocked by defender	Cross is controlled by an attacker at the target, and he made a shot. The ball hits or blocked by a defender
4. Cross receiver's shot saved by goalkeeper	Cross is controlled by an attacker at the target, and he made a shot. The shot is saved by goalkeeper.
5. Cross receiver's shot goes out, attempt off target	Cross is controlled by an attacker at the target, and he made a shot. Shot goes out, attempt off target. Ball not directed within the dimensions of the goal.
6. Cross receiver's shot hits teammate	Cross is controlled by an attacker at the target, and he made a shot. Shot is blocked by a teammate
7. Not Sure	Cross is controlled by an attacker at the target, and he made a shot. The outcome of shoot cannot be determined
2. Cross Receiver Passes/Crosses to a teammate	Cross is controlled by an attacker at the target, and he passed/crossed the upcoming ball to a teammate.
1. Passed teammate shoots and Scores, Goal	Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate scores.
2. Passed teammate shoots and misses/goalkeeper saves/ defender blocks	Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate shoots and misses or goalkeeper saves or defender blocks.

Table 7 (cont.)

3. Passed teammate passes to another teammate	Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate passes to another teammate.
4. Passed teammate cannot control the ball	Cross is controlled by an attacker at the target, and he passed to a teammate. The teammate cannot control the ball.
5. Pass is intercepted by defender/goalkeeper	Cross is controlled by an attacker at the target, and he passed to a teammate. Defense/Goalkeeper intercepts the passed ball.
6. Pass to opponent, mis pass	Cross is controlled by an attacker at the target, and he passed to a teammate. It is a mis pass.
7. Not Sure	Cross is controlled by an attacker at the target, and he passed to a teammate. The outcome of pass cannot be determined.
3. Cross Receiver Dribbles	Cross is controlled by an attacker at the target, and he starts dribbling the ball.
4. Cross Receiver cannot control or touch the ball, lost possession	Cross reaches an attacker at the target, but he cannot control or touch the ball and team lost possession. Unsuccessful attacking action.
5. Not Sure	Cross reaches an attacker at the target, but the outcome of the next action cannot be determined.
2. Cross off target, not reaches a teammate	Cross does not reach an attacker.
1. Team retains possession, ball is recycled	Cross does not reach an attacker at the target and the ball exits the box but attacking team retains possession, recycles the ball.
2. Cross goes out of pitch/play	Cross does not reach an attacker at the target and ball goes out of pitch/play.
3. Cross has no contact in the box from both teams	Cross is not controlled by any player and the ball exited the box.
4. Not Sure	Cross does not reach an attacker but the outcome of the next action cannot be determined.
3. Defender Interception	Defender interception towards cross.
1. Defensive block, tackle to crosser	Cross is blocked/intercepted/tackled by a defender immediately after ball comes out of the foot of the attacker.
2. Defender takes control of the ball upcoming from cross	Defender takes control of the ball upcoming from cross.
3. Defensive clearance to upcoming cross with kick, head etc.	Defender clears the ball upcoming from the cross with a kick, head etc.
1. Direction is not mentioned	Cross is cleared by a defender. Direction not mentioned.
2. To corner	Cross is cleared to corner by a defender.

Table 7 (cont.)

3. To throw in	Cross is cleared to throw in by a defender.
4. Out of box	Cross is cleared to out of box by a defender.
5. To Opponent	Cross is cleared to an opponent by a defender.
6. Inside penalty area	Cross is cleared inside penalty area by a defender.
7. To a teammate	Cross is cleared to a teammate by a defender.
8. Not Sure	Cross is cleared by a defender, but direction cannot be determined.
4. Defender's interception leads ball to opponent	Defender's interception towards cross find its way to an opponent
5. Defender's interception leads ball own goal	Defender's interception towards cross leads to own goal
6. Not Sure	Defender makes and interception, but the outcome of the next action cannot be determined.
4. Goalkeeper Interception	Goalkeeper interception towards cross.
1. Goalkeeper catches/gathers the ball upcoming from cross	The goalkeeper catches/gathers the ball and hold it in his hands from the cross.
2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	The goalkeeper clears the ball with punch, fist, slap, touch away.
1. Direction is not mentioned	The goalkeeper clears the ball with punch, fist, slap, touch away. Direction not mentioned.
2. To corner	The goalkeeper clears the ball with punch, fist, slap, touch away to corner.
3. To throw in	The goalkeeper clears the ball with punch, fist, slap, touch away to throw in.
4. Out of box	The goalkeeper clears the ball with punch, fist, slap, touch away to out of box.
5. To Opponent	The goalkeeper clears the ball with punch, fist, slap, touch away to an opponent.
6. Inside penalty area	The goalkeeper clears the ball with punch, fist, slap, touch inside penalty area by a defender.
7. To a teammate	The goalkeeper clears the ball with punch, fist, slap, touch to a teammate by a defender.
8. Not Sure	The goalkeeper clears the ball with punch, fist, slap, touch away but direction cannot be determined.
3. Goalkeeper's interception leads ball to opponent	Goalkeeper's interception towards cross find its way to an opponent.
4. Goalkeeper's interception leads ball own goal	Goalkeeper's interception towards cross leads to own goal.
5. Not Sure	Goalkeeper makes and interception, but the outcome of the next action cannot be determined.
5. Referee Interception	Referee stops the game.

Table 7 (cont.)

1. Penalty	A player on the defending team committed a foul and the referee awarded a penalty.
2. Offside	Defending team win a free kick after a player on the attacking team is ruled offside during the cross.
3. Free kick for attacking team (direct/indirect)	The referee awarded a free kick to the attacking team.
4. Free kick for defending team (direct/indirect)	The referee awarded a free kick to the defending team.
5. Not Sure	Referee stops the game but the outcome cannot be determined.
6. Not Sure	The outcome of cross cannot be determined.

Table 7 (cont.)

3.2.1.4. Training Material

To train the prospective labelers, the following procedures are decided to be followed:

1) A user manual that explains the purpose of this study, how the labeling will be done, the dictionary of the labels, and the detailed description of the survey document is prepared. The User Manual is prepared in Turkish as all the labelers were from Turkey. The user manual that was shared with the labelers is given in Appendix A.
2) Before the labeling process, a demo is conducted to show the labelers how the labeling process will be done. During the meetings, this demo is planned to be presented, and the detail of the survey is intended to be explained to the prospective labelers.
3) A sample survey is created and shared with the labelers. After the labeling process, the sample is evaluated, and feedback about the identified mistakes and misunderstandings is explained.

3.2.1.5. Labeling Tool

A spreadsheet has been formed featuring convenient drop-down menus for each label to assist labelers in making the process easier. Furthermore, each team's squad, the cross' outcome dictionary, and a list of slang words commonly encountered in football commentaries are also shared with the labelers. Lastly, to further aid understanding, full match replays are provided to labelers. This resource allows them to clarify any confusing commentaries by watching the matches and making labeling decisions. The instructions on using this spreadsheet to label the commentaries are comprehensively outlined in the accompanying handout provided in APPENDIX A in Turkish directives.

3.2.1.6. Qualifications of Labelers

To ensure an effective labeling process, specific qualifications were identified for potential labelers. Prospective labelers are expected to have an excellent grasp of English, demonstrate a keen interest in and consistent viewership of football matches spanning over 15 years, and hold at least a bachelor's degree. These criteria aim to ensure their proficiency, dedication, and analytical abilities necessary for establishing cause-and-effect relationships accurately in the football domain.

3.2.2. Labeler Recruitment

Following the completion of the planning phase, the labeler recruitment phase is started. Scheduled meetings are held with potential labelers to explain the study's objectives and outline the expected contributions from them. After that user manuals are shared, and a survey demo is done to show how labeling is done. At the end of the meeting, a sample survey is shared with labelers to evaluate their performance and give feedback about their work. Lastly, the final labelers are recruited for the task.

3.2.2.1. Scheduled Meetings with Prospective Labelers

Table 7 (cont.)

Following the identification of the desired qualifications for labelers, online and face-to-face meetings are organized with 20 potential candidates that satisfy the conditions. These sessions serve to communicate the objectives of the labeling process, provide a comprehensive description of the commentaries, offer useful tips for effective labeling, and showcase a demo using an example survey, guiding the prospective labelers through the process. At the end of the meeting, a sample survey and a user manual handout is distributed to the candidates.

3.2.2.2. Evaluation of Sample Surveys and Selection of the Labelers

Following the assessment of the sample surveys distributed during the meetings and considering the workload associated with the labeling process, a total of 16 labelers were selected for the task. Feedback on the evaluation of the sample surveys and their corresponding labels is provided to enhance the labelers' performance in their assigned task to ensure label performance validity, a minimum of two labelers are assigned to label the same dataset. Therefore, all commentaries are also labeled by the author of the thesis.

3.2.3. Labeling

Once the labelers are chosen, the labeling phase has started. Before assigning the commentaries to the labelers, a dataset preprocess is conducted to reduce the workload on the labelers. After that, the commentaries are assigned and distributed to the labelers. Regular online meetings are held for quick Q&A sessions with the labelers throughout the process.

3.2.3.1. Preprocessing

Before assigning the commentaries to the labelers, it is decided to conduct a preprocessing step to the dataset. During the examination of initial samples, it became evident that certain commentary texts do not exhibit a crossing action and can be filtered out accordingly. The filtered data comprises the following properties:

- Commentaries that declare the half-time and match scores. These are comments that are typically characterized by being typed in full capital letters during stoppage times like 45+2, 90+5, etc. Ex: *"HALF-TIME: SOUTHAMPTON 2-1 CHELSEA.", "FULL-TIME: BOURNEMOUTH 0-0 WOLVES."*
- Commentaries that declare the match or the second half is started. These are the comments that are noted in the first minute of the first and second halves and contain phrases like: "kick-off", "underway",etc. Ex: *"Forest kicks off the first half.", "After a short delay and a change of referee, Kane gets this match underway at the London Stadium!", "Neves resumes proceedings for Wolves, and we're back underway at Molineux!", We are underway in Leicester!*
- Commentaries explaining a substitution of players in the field. These are the commentaries that contain phrases like "comes off", "replacement", "substitute", "taken off", "comes on", "takes his place", "enter", "change", "makes way", "brought on" etc. Ex: *"Final changes for Palace now, as Eze is taken off and replaced by Ebiowei, who makes his Premier League debut.", "De Cordova-Reid makes way for Duffy, who will shore up the Fulham defence for the final minute."*

Table 7 (cont.)

- Commentaries that give general information about the clubs' seasonal and historical statistics. These are the commentaries that contain phrases like: "Premier League", "season", "19xx", "league games" etc. Ex: *"West Ham have attempted 21 crosses in this match, higher than their season average of 20.3 in the Premier League. So, it is fair to say Bournemouth's defence have been tested and, for the most part, have handled the pressure.", "Leicester have lost nine of their last 11 Premier League games against Manchester City (W2), and have lost their last three in a row."*
- Commentaries that contain managers' names which contain certain tactical changes in the matches, teams' statistics, and substitutions. Ex: *"Brentford have rocked City in the early stages of the contest. The Bees have the lead and could be further ahead. Guardiola needs a response from his side.", "Klopp swaps his right-backs as Alexander-Arnold is withdrawn for Gomez."*
- Commentaries that describe an injury or a treatment for a player. These are the commentaries that contain phrases like "injury", "cramp", "treatment" etc. Ex: *"Lallana is on the ground and is receiving treatment.", "Koulibaly is struggling with cramp and will have to make way here. The defender might just be feeling the pace of the Premier League."*
- Commentaries that show referee decisions like booking and pausing. These are the commentaries that contain phrases like "book", "yellow card", "pause" etc. Ex: *"De Cordova-Reid goes into the book for time-wasting.", "Foden receives a yellow card after a clumsy challenge from the winger.", "The game is paused for a cooling break."*

These commentaries' "Is there a Cross" attribute are labeled as "No" and filtered out from the dataset before being sent to the labelers. After this, number of commentaries remaining for labeling has decreased to 13722 entries.

3.2.3.2. Assigning Commentaries to the Labelers

The allocation of commentaries to labelers is based on their daily workload. Additionally, labelers are given the opportunity to express their preferences regarding the specific teams' commentaries they would like to label. In conclusion, the amount of data and their preferred team or teams of the labelers are shared in Table 8. It should be noted that 13722 commentaries are also labeled by the author to check their validity. The duration of the survey was set as 1.5 months.

Table 8 Number of Commentary and Teams Assigned to each Labeler

Labeler No	**Teams**	**Labeled Data**
Labeler 1	Bournemouth, Southampton	1150
Labeler 2	Everton, Manchester United	1143
Labeler 3	Aston Villa, West Ham United	1138
Labeler 4	Wolverhampton	48
Labeler 5	Leicester City, Liverpool	1147
Labeler 6	Nottingham Forrest	547

Table 7 (cont.)

Labeler 7	Manchester City, Wolverhampton	390
Labeler 8	Newcastle United	594
Labeler 9	Fulham, Leeds United	1141
Labeler 10	Manchester City	354
Labeler 11	Arsenal, Chelsea, Manchester United	1579
Labeler 12	Arsenal, Liverpool, Manchester United	1456
Labeler 13	Wolverhampton	346
Labeler 14	Chelsea, Liverpool	399
Labeler 15	Tottenham, Brighton	1145
Labeler 16	Brentford, Crystal Palace	1145
TOTAL		**13722**

3.2.3.3. Q&A's with Labelers During the Labeling Process

During the labeling process, quick Q&A sessions are done with the labelers. In these meetings, the questions that labelers have are answered. The frequently asked questions with the answers or the precautions that are taken during these meetings are as follows:

- Q: There are players who make the crosses that are not listed in the teams' squads
  A: Some players are transferred at the mid-season and not included in the initially defined squads. These are added to the squads and shared with all the labelers.
- Q: Are long balls considered as cross?
  A: No, only the balls that are targeted into the box are considered as crosses.
- Q: Are corners and free kicks considered as cross?
  A: Yes, they do. They are set play crosses.
- Q: Some players that makes the cross and the detailed outcome can be determined through match replay. Should I label accordingly?
  A: No, you shouldn't. The focus is on the sentences not on the videos.
- Q: After a commentary that contains a scoring action, there comes a one that also gives detail about that goal. Should I label them both?
  A: These commentaries are decided to be merged after the labeling process. These ones can be passed or labeled as "No" to gain some time.

3.2.4. Post-Labeling

Once the labeling is completed after 1.5 months, the datasets are collected from the labelers. It took 2 months for the author to complete the labeling of 13722 commentaries. Five hour is set for each day on average for the labeling process for two months. The labeling duration of the commentaries that do not contain a crossing action took between 0.5 and 1 minutes. On the other hand, labeling duration each commentary that contains a crossing action took between 2 and 3 minutes. It can go up to 5 minutes if a commentary is hard to understand and confirmation from the live footage is necessary.

The labelers and the authors Inter Annotator Agreements (IAAs) measurements are calculated using raw agreement and Cohen's Kappa (Cohen, 1960). After that, the labels

Table 7 (cont.)

that are different between the two labelers are filtered. A meeting with each of the labelers is organized and these are labeled with a consensus.

After adding the initially filtered commentaries that are described in section 3.2.3.1 and consolidating the 1138 rows commentaries that describes a goal in two separate rows, the final dataset is formed comprising 19686 rows.

3.2.4.1. Inter Annotator Agreements (IAA)

As mentioned in Section 2.3, the IAA is calculated by raw agreement and Cohen's Kappa metrics given in Table 1 for all labeling tasks and the agreement levels are identified using Table 1. For the labels "Crossing Team", "Crossing Player", "Outcome of Cross - Category 1", "Outcome of Cross - Category 2", "Outcome of Cross - Category 3", "Quality of Cross" and "Is There a Second Cross" labels, to compare the agreement levels on crosses, just the commentaries that are labeled as "Yes" at "Is There a Cross" label are considered in the process. Furthermore, same methodology is followed for "Second Cross - Crossing Player", "Outcome of Second Cross - Category 1", "Outcome of Second Cross - Category 2", "Outcome of Second Cross - Category 3" and "Quality of Second Cross" as just the commentaries that are labeled as "Yes" at "Is There a Second Cross" label are considered in the process. In addition, "Outcome of Cross - Category 2" labels agreements are calculated over the commentaries that both labelers have labeled "Outcome of Cross - Category 1" as the same one and similarly "Outcome of Cross - Category 3" labels agreements are calculated over the commentaries that both labelers have labeled "Outcome of Cross - Category 2" as the same one. This procedure followed in second crosses as well.

In addition, for the cross quality labels covering cross and second cross, Cohen's quadratic weighted kappa is used to calculate IAA with Equation (2-3)

3.2.4.1.1. IAA for "Is There a Cross" Label

In this labeling task, labelers are asked to decide whether the commentary contains a crossing action with "Yes" or "No" labels. Before the calculation of the IAA metrics, the following preprocessing steps have been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.

- The commentaries containing the "Not Sure" labels are filtered out.

The commentaries that raw agreement and Cohen's Kappa with its level of agreement equivalent metrics for IAA are given in Table 49.

The main reasons of the disagreements are founded as follows:

- Commentaries containing long ball.

Table 7 (cont.)

- Commentaries containing set pieces such as freekicks, throw-ins and corners.
- Commentaries containing lob passes.
- Commentaries containing passes made in the penalty box.
- Commentaries containing cut back or pull back.

3.2.4.1.2. IAA for "Crossing Team" Label

In this labeling task, labelers tag the team performing the crossing action as whether there is a crossing action. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

As the matches are independent events from each other and the two teams are changing in each match, the IAA metrics are initially calculated separately for each match. The IAA metrics for each labeler per match is given in

Table 50.

The weighted average of the raw agreement and Cohen's Kappa with its level of agreement equivalent metrics with respect to number of labeled data for each labeler for "Crossing Team" Label IAA are given in Table 51.

For "Crossing Team" Label both raw agreement and kappa metrics are high, and the level of agreement is "Almost Perfect" for the majority. The main reasons of the disagreements are founded as follows:

- Commentaries containing the nicknames of teams.
- Commentaries containing the teams that have players with same surname.

3.2.4.1.3. IAA for "Crossing Player" Label

In this labeling task, labelers tag the player performing the crossing action as whether there is a crossing action. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

Just like "Crossing Team" label, the IAA metrics are initially calculated separately in each match for "Crossing Player" Label. The IAA metrics for each labeler per match is given in Table 52.

The weighted average of the raw agreement and Cohen's Kappa with its level of agreement equivalent metrics with respect to number of labeled data for each labeler for "Crossing Player" Label IAA are given in Table 53.

For "Crossing Player" Label both raw agreement and kappa metrics are high, and the level of agreement is "Strong" for the majority. The main reasons of the disagreements are founded as follows:

- Labelers misjudge the player who sends the cross and who receives it.
- Commentaries containing the just the nationality of the players.
- Commentaries containing the teams that have players with same surname.

- Commentaries containing two crosses and the player’s name in second cross is labeled.

- Labelers mis click the players name from the drop-down list.

3.2.4.1.4. IAA for "Outcome of Cross – Category 1" Label

In this labeling task, labelers tag the outcome of the cross' category 1 if there is a crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

The commentaries that raw agreement and Cohen's Kappa with its level of agreement equivalent metrics for IAA are given in Table 54.

For "Outcome of Cross – Category 1" label both raw agreement is high and kappa metric is intermediate, and level of agreement is "Moderate" for the majority. The main reasons of the disagreements are founded as follows:

- Labelers misjudges the shot saved by goalkeeper and goalkeeper's interception to cross.
- Labelers misjudges the interception opponent's role between defender and goalkeeper.
- Labelers misjudges the player who is making the first contact as an attacker or defense.
- Commentaries that the attacking player cannot touch the ball and labeled as "Cross off target".

3.2.4.1.5. IAA for "Outcome of Cross – Category 2" Label

In this labeling task, labelers tag the outcome of the cross' category 2 if there is a crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.
- Only the commentaries which are labeled as same at "Outcome of Cross – Category 1" label by both labelers are selected.

- The commentaries containing the "Not Sure" labels are filtered out.

In this case, the categories/labels changes with "Outcome of Cross – Category 1". Therefore, the IAA metrics are calculated separately in for each "Outcome of Cross – Category 1. The IAA metrics for each labeler per match is given in

Table 55.

For "Outcome of Cross – Category 2" Label, the raw agreement and kappa metrics are distributed for different cross outcome categories 1 as follows:

- <u>1. Cross on target, reaches a teammate:</u> high raw agreement, low kappa metric and "weak" level of agreement for the majority. The main reasons of the disagreements are founded as follows:
    - Labelers generally considered "1. Cross Receiver Shoots/Heads" for outcome category 2. This is due to the flicks of the cross receiver is misjudged as a shoot rather than a pass.
    - Labelers conflict choosing between "1. Cross Receiver Shoots/Heads" and "4. Cross Receiver cannot control or touch the ball, lost possession" when the cross receiver cannot connect properly and shoots wide.
- <u>2. Cross off target, not reaches a teammate:</u> intermediate raw agreement, low kappa metric and "weak" level of agreement for the majority. The main reasons of the disagreements are found as follows:
    - Labelers conflict choosing between "1. Team retains possession, ball is recycled" and "3. Cross has no contact in the box from both teams" when the crossed ball is not connected by any player in the box, but a teammate catches the ball somehow.
    - Labelers conflict choosing between "2. Cross goes out of pitch/play" and "3. Cross has no contact in the box from both teams" when the crossed ball is not connected by any player in the box and goes out of play.
- <u>3. Defender Interception:</u> intermediate raw agreement, low kappa metric and "minimal" level of agreement for the majority. The main reasons of the disagreements are found as follows:
    - Labelers conflict choosing between "1. Defensive block, tackle to crosser" and "3. Defensive clearance to upcoming cross with kick, head etc." when the cross is blocked, and the ball goes out for a corner.
    - Labelers conflict choosing between "3. Defensive clearance to upcoming cross with kick, head etc." and "4. Defender's interception leads ball to

opponent" when the defender clears the ball and top ends up to an opponent and the defender's interception directly leaded to opponent and cause an attack.

- 4. Goalkeeper Interception: high raw agreement, intermediate kappa metric and "moderate" level of agreement for the majority. The main reason of the disagreements is found as follows:

  - Labelers conflict choosing between "2. Goalkeeper clearance to the ball from cross with punch, fist, slap," and "3. Goalkeeper's interception leads ball to opponent" when the goalkeeper clears the ball and top ends up to an opponent and the goalkeeper's interception directly leaded to opponent and cause an attack.

- 5. Referee Interception: high raw agreement, high kappa metric and "strong" level of agreement for the majority. The main reason of the disagreements is found as follows:

  - Labelers conflict choosing between "3. Free kick for attacking team (direct/indirect)" and "4. Free kick for defending team (direct/indirect)" after referee stops the game.

3.2.4.1.6. IAA for "Outcome of Cross – Category 3" Label

In this labeling task, labelers tag the outcome of the cross' category 3 if there is a crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.

- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.

- Only the commentaries which are labeled as same at "Outcome of Cross – Category 2" label by both labelers are selected.

- The commentaries containing the "Not Sure" labels are filtered out.

In this case, the categories/labels changes with "Outcome of Cross – Category 2". Therefore, the IAA metrics are calculated separately in for each "Outcome of Cross – Category 2. The IAA metrics for each labeler per match is given in Table 56.

For "Outcome of Cross – Category 3" Label, the raw agreement and kappa metrics are distributed for different cross outcome 2 categories as follows:

- <u>1. Cross Receiver Shoots/Heads (Under Cross Outcome Category 1: 1. Cross on target, reaches a teammate):</u> high raw agreement, high kappa metric and "strong" level of agreement for the majority. The main reasons of the disagreements are found as follows:

    - Labelers conflict choosing between "2. Cross receiver's shot hits goalpost, woodwork" and "5. Cross receiver's shot goes out, attempt off target" when the goal post is mentioned in the commentary, but the ball has no connection to it after a shot.

    - Labelers conflict choosing between "3. Cross receiver's shot blocked by defender" and "1. Cross Receiver Scores, Goal" when commentaries containing a goal action, but it is not caused directly by a cross but after a deflection from the shot.

- <u>2. Cross Receiver Passes/Crosses to a teammate (Under Cross Outcome Category 1: 1. Cross on target, reaches a teammate):</u> intermediate raw agreement, intermediate kappa metric and "moderate" level of agreement for the majority. The main reason of the disagreements is founde as follows:

    - Labelers conflict choosing between "2. Passed teammate shoots and misses/goalkeeper saves/defender blocks" and "5. Pass is intercepted by defender/goalkeeper" when the cross receiver passes to a teammate and he the shot is intercepted.

- <u>3. Defensive clearance to upcoming cross with kick, head (Under Cross Outcome Category 1: 3. Defender Interception):</u> low raw agreement, very low kappa metric and "minimal" level of agreement for the majority. The main reasons of the disagreements are found as follows:

    - Labelers conflict choosing between "1. Direction is not mentioned" and "2. To corner" when a commentary contains a phrase "ball goes behind".

    - Labelers conflict choosing between "1. Direction is not mentioned" and "4. Out of box" when a commentary contains a phrase "cleared away".

- <u>2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch (Under Cross Outcome Category 1: 4. Goalkeeper Interception):</u> low raw agreement, low kappa metric and "weak" level of agreement for the majority. The main reasons of the disagreements are found as follows:

    - Labelers conflict choosing between "1. Direction is not mentioned" and "2. To corner" when a commentary contains a phrase "ball goes behind".

    - Labelers conflict choosing between "1. Direction is not mentioned" and "4. Out of box" when a commentary contains a phrase "cleared away".

3.2.4.1.7. *IAA for "Quality of Cross" Label*

In this labeling task, labelers tag the quality of the crossing action if there is a crossing action. Before the calculation of the IAA metrics, some preprocessing has applied as follows:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

The labels for this task were "Very Good", "Good", "Neutral", "Bad" and "Very Bad" in an ordinal manner. Due to that, quadratic weighted Cohen's kappa metric is used as an IAA since the similarity between "Very Good" and "Good" is different than "Very Good" and "Bad".

The commentaries that raw agreement and quadratic weighted Cohen's kappa with its level of agreement equivalent metrics for IAA are given in

Table 57.

For "Quality of Cross" label both raw agreement and kappa metrics are intermediate, and level of agreement is "Moderate" for the majority. The main reasons of the disagreements are found as follows:

- Some labelers tend to give "Very Good" label only to the commentaries that contains goal.
- Some labelers tend to give "Very Bad" and "Bad" labels to the commentaries if the cross is off target even it was a dangerous one.
- Some adjectives like "dangerous", "brilliant", "perfect" etc. are not considered as "Very Good".

3.2.4.1.8. *IAA for "Is There a Second Cross" Label*

In this labeling task, labelers asked to decide if the commentary contains a second crossing action with "Yes" or "No" labels. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Cross" label by both labelers are selected.

- The commentaries containing the “Not Sure” labels are filtered out.

The commentaries that raw agreement and Cohen’s Kappa with its level of agreement equivalent metrics for IAA are given in

Table 58.

For “Is There a Second Cross” label raw agreement metric is very high and kappa metrics intermediate, and level of agreement is “Moderate” for the majority. The main reasons of the disagreements are found as follows:

- The amount of “Second Cross action” is relatively low in the dataset. The labelers tend to label “No” for most of the crossing actions. However, labelers had difficulties finding the commentaries that contain the second action.

3.2.4.1.9. IAA for “Second Cross - Crossing Player” Label

In this labeling task, labelers tag the player who performs the second crossing action if there is a second crossing action. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.

- Only the commentaries which are labeled as “Yes” at “Is There a Second Cross” label by both labelers are selected.

- The commentaries containing the “Not Sure” labels are filtered out.

The IAA metrics are initially calculated separately in each match for “Second Crossing Player” Label just like the first one. The IAA metrics for each labeler per match is given in Table 59.

The weighted average of the raw agreement and Cohen’s Kappa with its level of agreement equivalent metrics with respect to number of labeled data for each labeler for “Crossing Player” Label IAA are given in Table 60.

For “Crossing Player” Label both raw agreement and kappa metrics are high, and level of agreement is “Almost Perfect” for the majority. The main reasons of the disagreements are found as follows:

- Labelers misjudge the player who sends the cross and who receives it.

- Labelers misjudge the crossing sequence of the players and label the player who receive the cross directly.

- Commentaries containing the teams that have players with same surname.

3.2.4.1.10. IAA for "Outcome of Second Cross – Category 1" Label

In this labeling task, labelers tag the outcome of the second cross' category 1 if there is a second crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Second Cross" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

The commentaries that raw agreement and Cohen's Kappa with its level of agreement equivalent metrics for IAA are given in Table 61.

For "Outcome of Second Cross – Category 1" label both raw agreement and kappa metrics are intermediate, and level of agreement is "Moderate" for the majority. The main reasons of the disagreements are found as same with the first cross outcome.

3.2.4.1.11. IAA for "Outcome of Second Cross – Category 2" Label

In this labeling task, labelers tag the outcome of the second cross' category 2 if there is a second crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.
- Only the commentaries which are labeled as "Yes" at "Is There a Second Cross" label by both labelers are selected.
- Only the commentaries which are labeled as same at "Outcome of Second Cross – Category 1" label by both labelers are selected.
- The commentaries containing the "Not Sure" labels are filtered out.

In this case, the categories/labels changes with "Outcome of Second Cross – Category 1". Therefore, the IAA metrics are calculated separately in for each "Outcome of Second Cross – Category 1. The IAA metrics for each labeler per match is given in

Table 62.

For "Outcome of Second Cross – Category 2" Label, the raw agreement and kappa metrics are distributed for different cross outcome categories 1 as follows:

- 1. Cross on target, reaches a teammate: very high raw agreement, very high kappa metric and "Almost Perfect" level of agreement for the majority besides the no record cases.

- 2. Cross off target, not reaches a teammate: very high raw agreement, very high kappa metric and "Almost Perfect" level of agreement for the majority besides the no record cases.

- 3. Defender Interception: intermediate raw agreement and kappa metric and "moderate" level of agreement for the majority. The main reasons of the disagreements are found as follows:

    - Labelers conflict choosing between "1. Defensive block, tackle to crosser" and "3. Defensive clearance to upcoming cross with kick, head etc." when the cross is blocked, and the ball goes out for a corner.

    - Labelers conflict choosing between "3. Defensive clearance to upcoming cross with kick, head etc." and "4. Defender's interception leads ball to opponent" when the defender clears the ball and top ends up to an opponent and the defender's interception directly leaded to opponent and cause an attack.

- 4. Goalkeeper Interception: very high raw agreement, very high kappa metric and "Almost Perfect" level of agreement for the majority besides the no record cases.

- 5. Referee Interception: very high raw agreement, very high kappa metric and "Almost Perfect" level of agreement for the majority besides the no record cases.

3.2.4.1.12. IAA for "Outcome of Second Cross – Category 3" Label

In this labeling task, labelers tag the outcome of second the cross' category 3 if there is a second crossing action from the first level of categories given in Table 7. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.

- Only the commentaries which are labeled as "Yes" at "Is There a Second Cross" label by both labelers are selected.

- Only the commentaries which are labeled as same at "Outcome of Second Cross – Category 2" label by both labelers are selected.

- The commentaries containing the "Not Sure" labels are filtered out.

In this case, the categories/labels changes with “Outcome of Second Cross – Category 2”. Therefore, the IAA metrics are calculated separately in for each “Outcome of Second Cross – Category 2. The IAA metrics for each labeler per match is given in Table 63.

For “Outcome of Second Cross – Category 3” Label, the raw agreement and kappa metrics are distributed for different second cross outcome 2 categories as follows:

- <u>1. Cross Receiver Shoots/Heads (Under Cross Outcome Category 1: 1. Cross on target, reaches a teammate):</u> very high raw agreement, very high kappa metric and “almost perfect” level of agreement for the majority besides the no record cases.

- <u>2. Cross Receiver Passes/Crosses to a teammate (Under Cross Outcome Category 1: 1. Cross on target, reaches a teammate):</u> only one record for this label with “perfect” agreement.

- <u>3. Defensive clearance to upcoming cross with kick, head (Under Cross Outcome Category 1: 3. Defender Interception):</u> only 7 records for this label with intermediate raw agreement and “weak” kappa agreement.

- <u>2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch (Under Cross Outcome Category 1: 4. Goalkeeper Interception):</u> only one record for this label with “none” agreement.

3.2.4.1.13. IAA for “Quality of Second Cross” Label

In this labeling task, labelers tag the quality of the crossing action if there is a crossing action. Before the calculation of the IAA metrics, the following preprocessing step has been applied:

- The former or latter commentaries of a scoring action describing the goal in detail which will be merged at the formation of final dataset are filtered out.

- Only the commentaries which are labeled as “Yes” at “Is There a Second Cross” label by both labelers are selected.

- The commentaries containing the “Not Sure” labels are filtered out.

The labels for this task were “Very Good”, “Good”, “Neutral”, “Bad” and “Very Bad” in an ordinal manner just like the first cross. Due to that, quadratic weighted Cohen’s kappa metric is used as an IAA since the similarity between “Very Good” and “Good” is different than “Very Good” and “Bad”.

The commentaries that raw agreement and quadratic weighted Cohen’s kappa with its level of agreement equivalent metrics for IAA are given in Table 64.

For “Quality of Cross” label both raw agreement and kappa metrics are low, and level of agreement is “Weak” for the majority. The main reasons of the disagreements are founded as follows:

- Some labelers tend to give “Very Good” label only to the commentaries that contains goal.

- Some labelers tend to give “Very Bad” and “Bad” labels to the commentaries if the cross is off target even it was a dangerous one.

Some adjectives like “dangerous”, “brilliant”, “perfect” etc. are not considered as “Very Good”.

3.2.4.2. Consensus and Final Label Formation

After collecting each dataset from the labelers, a spread sheet is created for each of them to compare their labels with the author’s labels. In this spreadsheet, two new columns are added to each labeling task which are the “controlling column” and “final label column”. The controlling columns checks whether there is a difference between labelers and author’s tags. If there is, this cell in this column is framed as “0” which indicates that the labels are not same including the labels “Not Sure” which were excluded in calculating IAA. If there is not, the cell in this column is framed as “1” and their mutual tag is saved to “final label column”. A sample from spreadsheet formed for Labeler 11 is given in Figure 1.

After this operation is applied to all labels, a meeting is organized with each of the labelers. In these meetings, all the labels that are framed as in the “controlling column” are reviewed with the labelers and a consensus tried to be reached. In case of disagreements and not reaching a consensus, A third party is chosen from a pool of 16 labelers and their opinion is used to make the final decision.

This consensus process took up to three hours with each labeler and it took one week to complete all the consensus process with all the labelers.

After labeling phase is completed, the final dataset with 19686 rows is formed to be used in LLMs.

Match No	Team_1	Team_2	Score	Match_Date	Min	Comment	Is there a Cross? (Anıl)	Is there a Cross? (Labeler 11)	Is there a Cross? (Check)	Is there a Cross? (Final)	Crossing Team (Anıl)	Crossing Team (Labeler 11)	Crossing Team (Check)	Crossing Team (Final)	Crossing Player (Anıl)	Crossing Player (Labeler 11)	Crossing Player (Check)	Crossing Player (Final)	Outcome of Cross - Category 1 (Anıl)	Outcome of Cross - Category 1 (Labeler 11)	Outcome of Cross - Category 1 (Check)	Outcome of Cross - Category 1 (Final)	Outcome of Cross - Category 2 (Anıl)	Outcome of Cross - Category 2 (Labeler 11)	Outcome of Cross - Category 2 (Check)	Outcome of Cross - Category 2 (Final)	Outcome of Cross - Category 3 (Anıl)	Outcome of Cross - Category 3 (Labeler 11)	Outcome of Cross - Category 3 (Check)	Outcome of Cross - Category 3 (Final)	Quality of Cross (Anıl)	Quality of Cross (Labeler 11)	Quality of Cross (Check)	Quality of Cross (Final)
12	Arsenal	Leicester_City	4 - 2	13 Aug 2022	3'	Jesus looks to chase down Xhaka's long pass over the top, but Evans holds firm to shield it out of play.	No	Yes	Checked	No		Arsenal	Checked	0		Midfielder - Granit Xhaka - 34	Checked	0		2. Cross off target, not reaches a teammate	Checked	0		2. Cross goes out of pitch/play	Checked	0			Checked	0		Neutral	Checked	0
12	Arsenal	Leicester_City	4 - 2	13 Aug 2022	11'	Close! Arsenal win a free-kick near the touchline. Odegaard swings in a dangerous cross and Martinelli nips in at the front post to head it into the side netting!	Yes	Yes	1	Yes	Arsenal	Arsenal	1	Arsenal	Midfielder - Martin Odegaard - 8	Midfielder - Martin Odegaard - 8	1	Midfielder - Martin Odegaard - 8	1. Cross on target, reaches a teammate	1. Cross on target, reaches a teammate	1	1. Cross on target, reaches a teammate	1. Cross Receiver Shoots/Heads	1. Cross Receiver Shoots/Heads	1	1. Cross Receiver Shoots/Heads	5. Cross receiver's shot goes out, attempt off target	5. Cross receiver's shot goes out, attempt off target	1	5. Cross receiver's shot goes out, attempt off target	Very Good	Very Good	1	Very Good
12	Arsenal	Leicester_City	4 - 2	13 Aug 2022	20'	Off the post! Saka curls a deep cross towards the back post and Xhaka arrives to send a header crashing off the post from close range!	Yes	Yes	1	Yes	Arsenal	Arsenal	1	Arsenal	Forward - Bukayo Saka - 7	Forward - Bukayo Saka - 7	1	Forward - Bukayo Saka - 7	1. Cross on target, reaches a teammate	1. Cross on target, reaches a teammate	1	1. Cross on target, reaches a teammate	1. Cross Receiver Shoots/Heads	1. Cross Receiver Shoots/Heads	1	1. Cross Receiver Shoots/Heads	5. Cross receiver's shot goes out, attempt off target	2. Cross receiver's shot hits goalpost, woodwork	Checked	2. Cross receiver's shot hits goalpost, woodwork	Good	Very Good	Checked	Very Good
12	Arsenal	Leicester_City	4 - 2	13 Aug 2022	89'	Gabriel can't quite latch onto Tierney's deep free-kick and it trickles behind.	Yes	No	Checked	Yes	Arsenal		Checked	Arsenal	Defender - Kieran Tierney - 3		Checked	Defender - Kieran Tierney - 3	2. Cross off target, not reaches a teammate		Checked	2. Cross off target, not reaches a teammate	2. Cross goes out of pitch/play		Checked	2. Cross goes out of pitch/play			Checked	0	Neutral		Checked	Neutral
32	Chelsea	Leicester_City	2 - 1	27 Aug 2022	39'	Castagne tees up Praet to cross from the right, and he does so, with another deflection off Cucurella winning Leicester a corner.	Yes	Yes	1	Yes	Leicester_City	Leicester_City	1	Leicester_City	Midfielder - Dennis Praet - 26	Defender - Timothy Castagne - 27	Checked	Midfielder - Dennis Praet - 26	3. Defender Interception	1. Cross on target, reaches a teammate	Checked	3. Defender Interception	3. Defensive clearance to upcoming cross with kick, head	3. Cross Receiver Dribbles	Checked	3. Defensive clearance to upcoming cross with kick, head	2. To corner		Checked	2. To corner	Neutral	Very Good	Checked	Neutral
32	Chelsea	Leicester_City	2 - 1	27 Aug 2022	47'	GOOOOOAAAAAALLLLLLLL!!!!! Chelsea lead!!!! Cucurella finds Sterling on the edge of the box, who takes on Amartey and shoots. It takes a nick off the Foxes defender, sending it looping over Ward and into the back of the net!!!! 1-0 to the hosts, and what a start to the second half!	No	No	1	No			1	0			1	0			1	0			1	0			1	0			1	0
32	Chelsea	Leicester_City	2 - 1	27 Aug 2022	47'	A good ball across from Cucurella gave Sterling plenty of space, and Amartey didn't get close enough until it was too late!	No	No	1	No			1	0			1	0			1	0			1	0			1	0			1	0

Figure 1 Sample from the spread sheet created for Labeler 11 to reach consensus on the labels

3.3. Fine-Tuning the Large Language Models (LLMs)

After the final dataset is formed, the large language models (LLMs) are used to identify crossing characteristics in football from commentaries.

In this thesis, SportsBERT model (Srinivasan, n.d.) which is a BERT model trained from scratch with specific focus on sports articles is fine-tuned to determine the crossing characteristics. The training corpus for SportsBERT model includes roughly 8 million news articles in English scraped from the web related to sports between 2016 and 2020 covering news from soccer, football, basketball, tennis etc. The architecture used in this model is the BERT base uncased architecture. The model was trained on four V100 GPUs. It is a Masked Language Model based transformers model and the primary task of the model is to fill in missing masked tokens. However, in this thesis the model is used in text classification tasks as the author suggested it can be used for such a case.

The SportsBERT model is fine-tuned with full fine-tuning approach with the datasets that are formed after the labeling process. In other words, the layers in the entire model are adjusted during the training process. Initially, Parameter-Efficient Fine-Tuning (PEFT) approach is used to for the fine-tuning process, but the performance metrics of the fine-tuned models were lower than the models that were trained with full fine-tuning approach. For full fine-tuning, the Tesla V100 GPU supplied by Google Colaboratory is used.

To evaluate the performance of the finetuned model, k-Fold cross-validation is applied on the football commentary datasets for different tasks. Three different splitting techniques are considered during the cross-validation process for obtaining train-validation-test datasets:

1) Random Splitting of the Dataset: The dataset in randomly split into train-validation-test datasets with 0.6-0.2-0.2 ratios respectively.

2) Splitting Dataset Based on Matches: The dataset is split into train-validation-test datasets with respect to the matches of each commentary. There are 304 matches in the datasets and these matches are split into train-validation-test matches with 0.6-0.2-0.2 ratios respectively. Thus, 182-61-61 is the distribution of matches for train-validation-test datasets, which also roughly divides the dataset into train-validation-test datasets with 0.6-0.2-0.2 ratios.

3) Splitting Dataset Based on Teams: The dataset is split into train-validation-test datasets with respect to the teams. There are 20 different teams in the datasets and these teams are split into train-validation-test to get roughly 0.6-0.2-0.2 ratios for them. Thus, 2 teams are selected for test datasets and 2 teams are selected for validation datasets and 16 for training dataset which leads to 10 folds.

The cross-validation technique is used to measure the performance for the models for the following 4 tasks:

1) Crossing Check: The existence of a crossing action in each commentary is considered.

2) Crossing Player: The player who makes the crossing action in each commentary that contains crossing action is considered (commentaries that contains more than one crossing action are eliminated).

3) Crossing Outcome: The outcome of the crossing in each commentary that contains crossing action is considered (commentaries that contains more than one crossing action and labeled as "Not Sure" by the labelers are eliminated).

4) Crossing Quality: The quality of the crossing in each commentary that contains crossing action is considered (commentaries that contains more than one crossing action are eliminated).

3.3.1. Crossing Check Classification with finetuned LLM Models:

The cross validation is applied on the whole dataset containing 19686 unique commentaries to determine the existence of crossing action for the classification task.

Optuna library is used to tune the hyperparameters. As the authors of the BERT model suggested, the following hyperparameters are considered as the input:

- Batch Size: 16, 32, 64 (The batch sizes below 16 are not considered because of the cost of the model)
- Learning Rate: 1e-5, 3e-5, 5e-5

The dataset split is done with match split method by applying 20 epochs to find the best model. The hyperparameters that resulted with least test loss with 0.108 are founded as:

- batch_size: 32,
- learning_rate: 5e-05

After finding the best hyperparameters, the cross-validation procedure is conducted with the best hyperparameters. 20 epochs are considered to observe the loss curve for the first fold. The training and evaluation loss vs step (one update of model's parameters during training process) is given in Figure 2. Each epoch consists of 370 batches with shape [32,128]. It can be seen that, the model tends to overfit after 5 epochs (step = 1850)

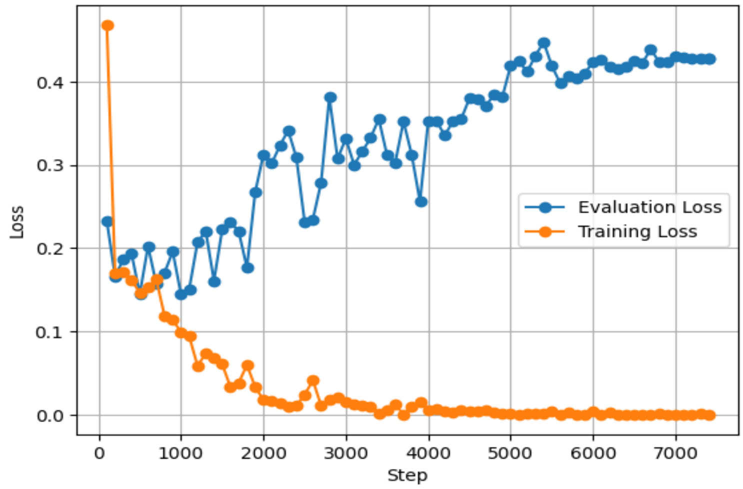


Figure 2 Training and Validation Loss vs Step Graph of the Model in Fold 1

The F1 scores and accuracies vs step are given in Figure 3 and Figure 4 respectively.

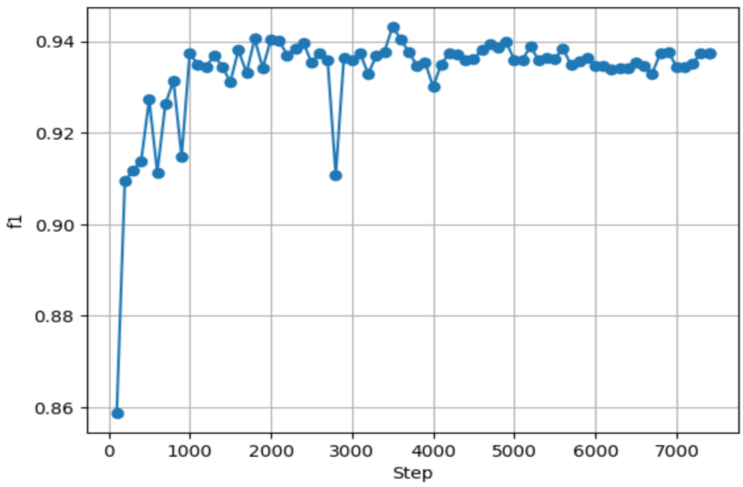


Figure 3 Test F1 score vs step graph of the finetuned model in Fold 1

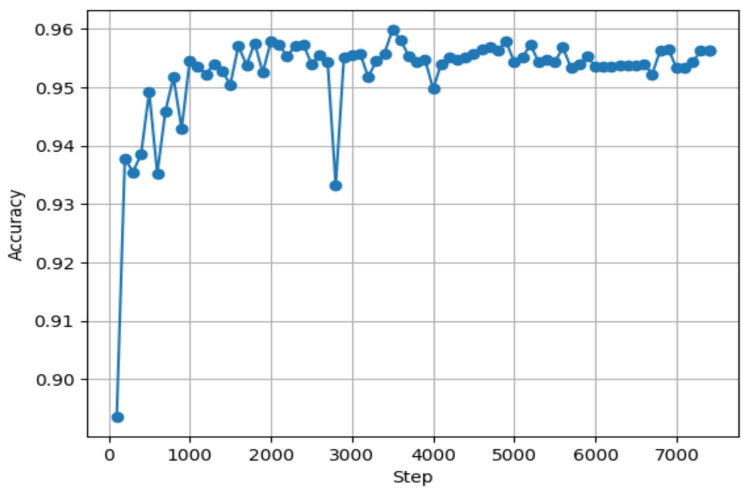


Figure 4 Test Accuracy vs Step Graph of the finetuned model in Fold 1

As seen in Figure 2, Figure 3 and Figure 4; the loss, F1 score and accuracy vs step graphs all overfit after 5 epochs, the cross validation is applied with 10 epochs to observe the characteristics for crossing check classification task.

0.6-0.2-0.2 ratios for train-validation-test datasets are considered and k=5-fold cross-validation is applied for random and match based splitting methods. The dataset is randomly divided 5 subsets (or folds) of approximately equal size for all strategies. To get 0.6-0.2-0.2 ratios for train-validation-test datasets, k=10-fold cross-validation is applied for team based splitting method.

The model is then trained 5 times for random, and match based splitting strategies and 10 times for team-based strategies, each time using a different fold as the test set and the remaining folds as the training and validation set and the performance metrics accuracy and F1 score are calculated for each iteration. Also, loss result for each iteration is depicted for all strategies. The confusion matrices for each iteration are also shared for each strategy. Only the first fold's loss graphs and confusion matrices are shared for each strategy below.

3.3.1.1. Crossing Check Classification with Random Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for random splitting method. In other words, 11811 commentaries are used for training, 3937 commentaries are used for validation and 3938 commentaries are used for test datasets. The loss, accuracy and F1 score vs step graphs in Fold 1 for crossing check classification with random splitting strategy are shared in Figure 5, Figure 6 and Figure 7 respectively.

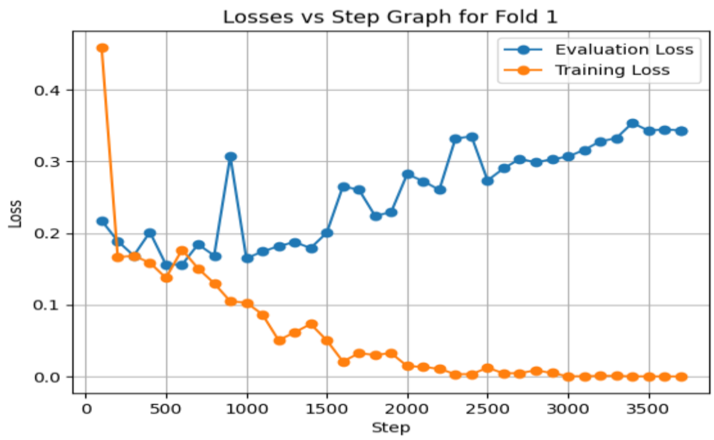


Figure 5 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Check Classification with Random Splitting Strategy

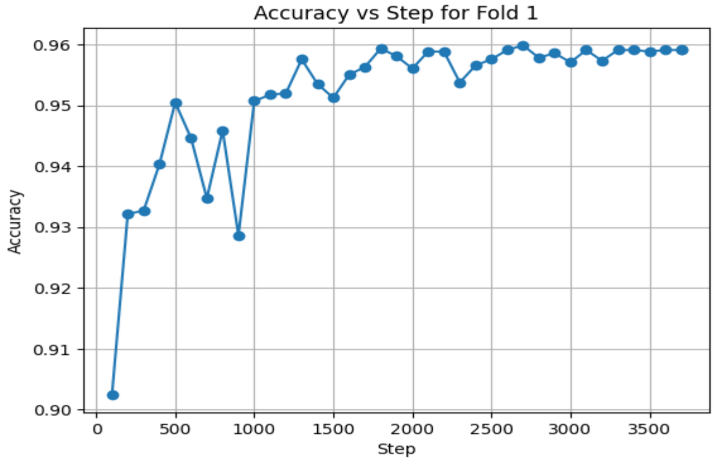


Figure 6 Accuracy vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Random Splitting Strategy

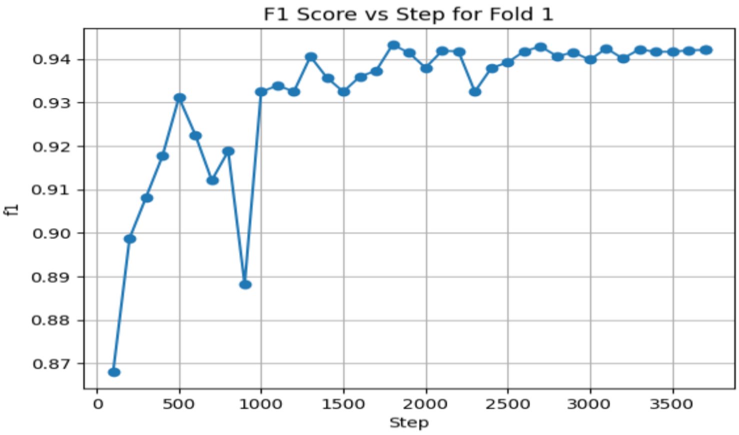


Figure 7 F1 Score vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Random Splitting Strategy

The class distribution percentages for label “No” to label “Yes” are 78%-22%, 77%-23% and 78%-22% in Fold 1 train, validation, and test datasets respectively which are approximately same.

The confusion matrix formed with test dataset in Fold 1 for crossing check classification with Random Splitting Strategy is given in Table 9 in which the rows represent the actual labels and columns represents the predicted labels.

Table 9 Confusion Matrix of test dataset for Fold 1 in Crossing Check Classification with Random Splitting Strategy

	No	**Yes**
No	3004	85
Yes	75	774

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, F1 score, precision, and recall results for each fold, are given in Table 10. The mean values and standard deviation of these metrics are also shown in

Table 10. The average accuracy for all folds is calculated as 95.9% and the average F1 score is calculated as 0.941.

Table 10 Performance metrics of fine-tuned model for each fold in Crossing Check Classification with Random Splitting Strategy

fold	accuracy	F1 score	precision	recall
1	0.959	0.940	0.938	0.942
2	0.958	0.938	0.935	0.941
3	0.957	0.937	0.933	0.940
4	0.961	0.946	0.951	0.941
5	0.962	0.946	0.947	0.945
Mean	**0.959**	**0.941**	**0.941**	**0.942**
Std Dev	**0.002**	**0.004**	**0.008**	**0.002**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “No”. Using this label as the as the classifier, the performance metrics are founded as in Table 11. It can be seen that, the average accuracy is increased by 18.1% and F1 score is increased by 0.503 using the fine-tuned model.

Table 11 Performance metrics of baseline model for each fold in Crossing Check Classification with Random Splitting Strategy

fold	accuracy	F1 score	precision	recall
1	0.784	0.440	0.392	0.500
2	0.786	0.440	0.393	0.500
3	0.787	0.441	0.394	0.500
4	0.762	0.432	0.381	0.500
5	0.770	0.435	0.385	0.500
Mean	**0.778**	**0.438**	**0.389**	**0.500**
Std Dev	**0.011**	**0.004**	**0.006**	**0.000**

3.3.1.2. Crossing Check Classification with Match Based Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for match based splitting method. In other words, 0.6 of the 304 matches, 182 matches’ commentaries are used for training, 61 matches’ commentaries are used for validation and 61 matches’ commentaries are used for test datasets. The loss, accuracy and F1 score vs step graphs in Fold 1 for crossing check classification with match based splitting strategy are shared in Figure 8, Figure 9 and Figure 10 respectively.

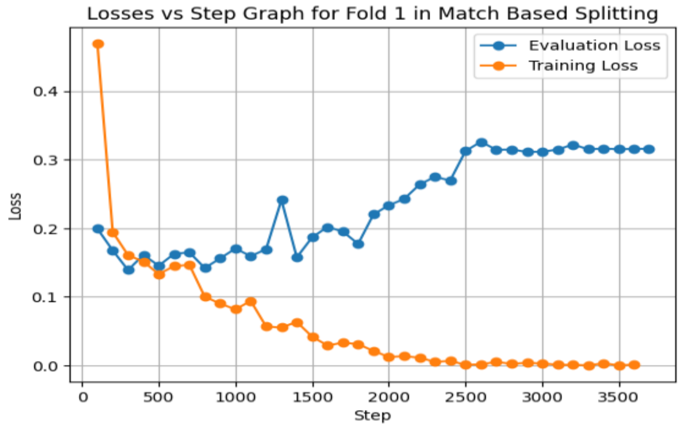


Figure 8 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Check Classification with Match Based Splitting Strategy

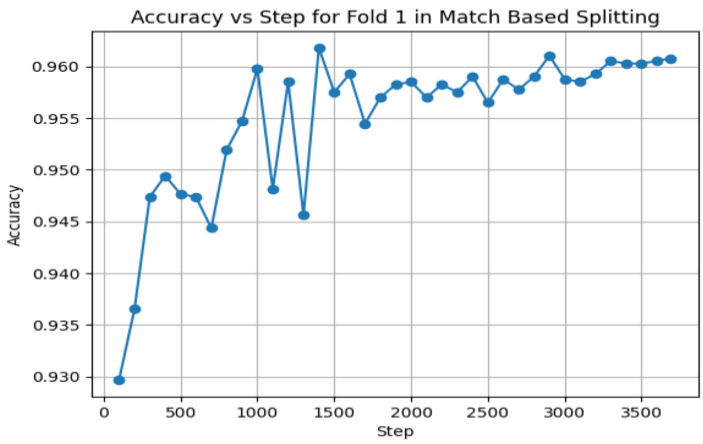


Figure 9 Accuracy vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Match Based Splitting Strategy

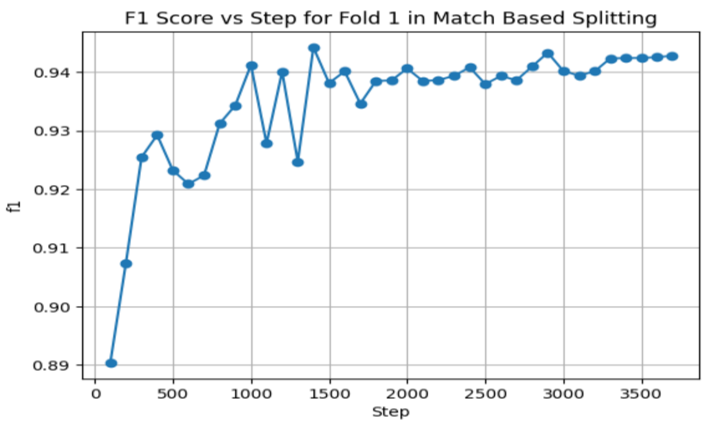


Figure 10 F1 Score vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Match Based Splitting Strategy

The class distribution percentages for label “No” to label “Yes” are 77%-23%, 78%-22% and 79%-21% in Fold 1 train, validation, and test datasets respectively which are approximately same.

The confusion matrix formed with test dataset in Fold 1 for crossing check classification with Match Based Splitting Strategy is given in Table 12 in which the rows represent the actual labels and columns represents the predicted labels.

Table 12 Confusion Matrix of test dataset for Fold 1 in Crossing Check Classification with Match Based Splitting Strategy

	No	**Yes**
No	3058	64
Yes	82	744

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, F1 score, precision, and recall results for each fold, are given in Table 13. The mean values and standard deviation of these metrics are also shown in Table 13. The average accuracy for all folds is calculated as 96.2% and the average F1 score is calculated as 0.945.

Table 13 Performance metrics of fine-tuned model for each fold in Crossing Check Classification with Match Based Splitting Strategy

fold	**accuracy**	**F1 score**	**precision**	**recall**
1	0.963	0.944	0.947	0.940
2	0.960	0.944	0.947	0.942
3	0.964	0.949	0.944	0.955
4	0.961	0.943	0.945	0.940
5	0.962	0.944	0.943	0.944
Mean	**0.962**	**0.945**	**0.945**	**0.944**
Std Dev	**0.002**	**0.003**	**0.002**	**0.006**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in match based splitting strategy for each fold is founded as “No”. Using this label as the as the classifier, the performance metrics are founded as in Table 14. It can be seen that, the average accuracy is increased by 18.4% and F1 score is increased by 0.507 using the fine-tuned model.

Table 14 Performance metrics of baseline model for each fold in Crossing Check Classification with Match Based Splitting Strategy

fold	**accuracy**	**F1 score**	**precision**	**recall**
1	0.791	0.442	0.395	0.500
2	0.763	0.433	0.381	0.500
3	0.774	0.436	0.387	0.500
4	0.780	0.438	0.390	0.500
5	0.782	0.439	0.391	0.500
Mean	**0.778**	**0.438**	**0.389**	**0.500**
Std Dev	**0.010**	**0.003**	**0.005**	**0.000**

3.3.1.3. Crossing Check Classification with Team Based Splitting Strategy:

There are 20 different teams in the datasets and these teams are split into train-validation-test to get roughly 0.6-0.2-0.2 ratios for them. Thus, two teams are selected for the test datasets and two teams are selected for the validation datasets and sixteen for training dataset which leads to 10 folds. The teams that are tested in each fold is given as below:

- Fold 1: ['Arsenal', 'Tottenham Hotspur']
- Fold 2: ['Aston Villa', 'Nottingham Forest']
- Fold 3: ['Chelsea', 'Fulham']
- Fold 4: ['Brentford', 'Liverpool']
- Fold 5: ['Southampton', 'West Ham United']
- Fold 6: ['Bournemouth', 'Manchester United']
- Fold 7: ['Leeds United', 'Wolverhampton Wanderers']
- Fold 8: ['Brighton', 'Manchester City']
- Fold 9: ['Everton', 'Leicester City']
- Fold 10: ['Crystal Palace', 'Newcastle United']

The loss, accuracy and F1 score vs step graphs in Fold 1 for crossing check classification with team based splitting strategy are shared in Figure 11, Figure 12, and Figure 13 respectively.

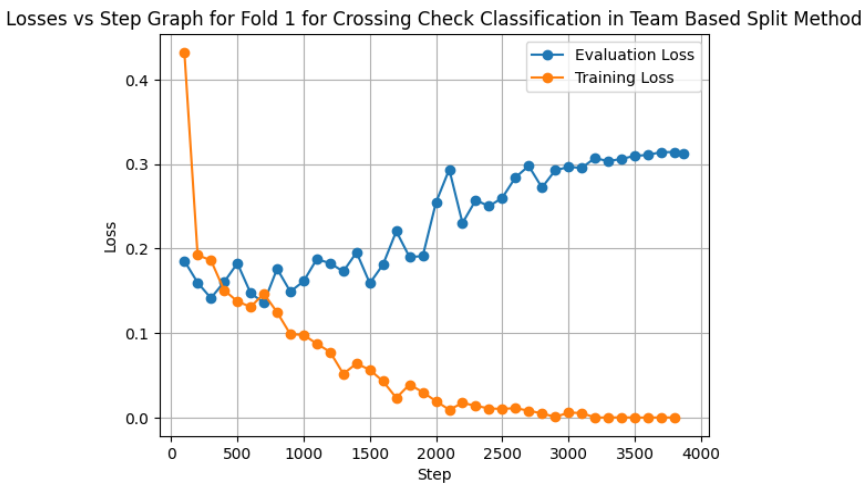


Figure 11 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Check Classification with Team Based Splitting Strategy

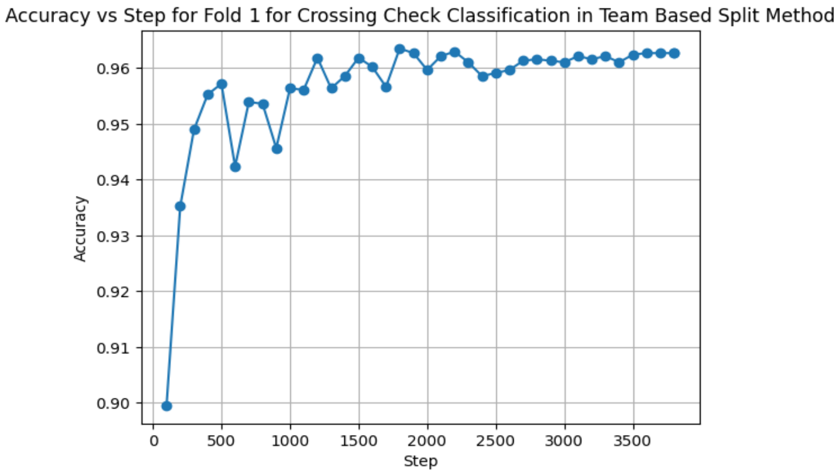


Figure 12 Accuracy vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Team Based Splitting Strategy

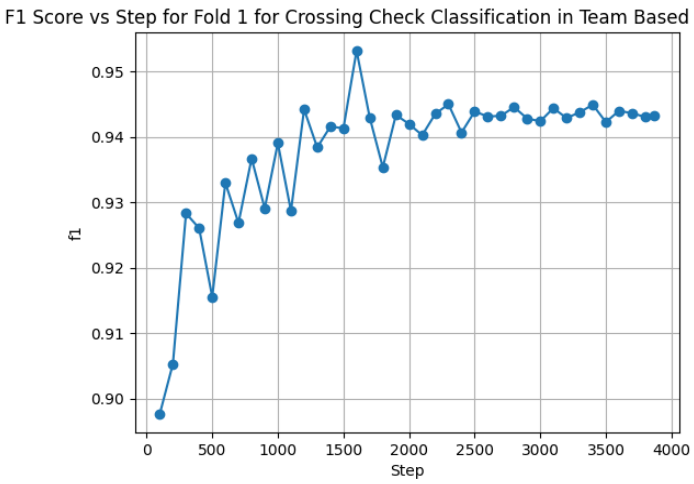


Figure 13 F1 Score vs Step Graph of the fine-tuned model in Fold 1 for Crossing Check Classification with Team Based Splitting Strategy

The class distribution percentages for label “No” to label “Yes” are all 78%-22% in Fold 1 train, validation, and test datasets.

The confusion matrix formed with test dataset in Fold 1 for crossing check classification with Team Based Splitting Strategy is given in Table 15 in which the rows represent the actual labels and columns represents the predicted labels.

Table 15 Confusion Matrix of test dataset for Fold 1 in Crossing Check Classification with Team Based Splitting Strategy

	No	**Yes**
No	2771	75
Yes	68	719

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, F1 score, precision, and recall results for each fold, are given in Table 16. The mean values and standard deviation of these metrics are also shown in

Table 16. The average accuracy for all folds is calculated as 96.1% and the average F1 score is calculated as 0.943.

Table 16 Performance metrics of fine-tuned model for each fold in Crossing Check Classification with Team Based Splitting Strategy

fold	accuracy	F1 score	precision	recall
1	0.961	0.942	0.941	0.944
2	0.959	0.941	0.943	0.940
3	0.962	0.947	0.948	0.947
4	0.956	0.938	0.940	0.936
5	0.968	0.954	0.958	0.950
6	0.960	0.943	0.942	0.943
7	0.962	0.943	0.946	0.940
8	0.961	0.941	0.946	0.937
9	0.960	0.942	0.943	0.940
10	0.959	0.941	0.942	0.939
Mean	**0.961**	**0.943**	**0.945**	**0.942**
Std Dev	**0.003**	**0.004**	**0.005**	**0.004**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in team based splitting strategy for each fold is founded as “No”. Using this label as the as the classifier, the performance metrics are founded as in Table 17. It can be seen that, the average accuracy is increased by 18.3% and F1 score is increased by 0.505 using the fine-tuned model.

Table 17 Performance metrics of baseline model for each fold in Crossing Check Classification with Team Based Splitting Strategy

fold	accuracy	F1 score	precision	recall
1	0.783	0.439	0.392	0.500
2	0.777	0.437	0.389	0.500
3	0.766	0.434	0.383	0.500
4	0.771	0.435	0.386	0.500
5	0.776	0.437	0.388	0.500
6	0.773	0.436	0.387	0.500
7	0.785	0.440	0.392	0.500
8	0.789	0.441	0.395	0.500
9	0.781	0.438	0.390	0.500
10	0.778	0.438	0.389	0.500
Mean	**0.778**	**0.438**	**0.389**	**0.500**
Std Dev	**0.007**	**0.002**	**0.003**	**0.000**

3.3.1.4. Crossing Check Classification Results of the Strategies:

The average metric results after cross validation applied for each strategy on test datasets are given in Table 18. The standard deviations for all metrics are low (<0.01) so it can be said that the average results of each strategy are nearly same. Overall, for all strategies the average accuracy is founded as 96.1% and macro F1 score is founded as 0.943.

Table 18 Average metric results of fine-tuned models for crossing player classification after cross validation applied for each strategy on test datasets

Strategy	accuracy	Macro F1 score	precision	recall
Random Split	0.959	0.941	0.941	0.942
Match Split	0.962	0.945	0.945	0.944
Team Split	0.961	0.943	0.945	0.942
Mean	**0.961**	**0.943**	**0.944**	**0.943**
Std Dev	**0.010**	**0.020**	**0.020**	**0.010**

When the test datasets are investigated, the common main reasons for mislabeling are observed as below:

- Some commentaries explain the actions that will happen in the following minutes. These commentaries generally are the ones that is written after a foul is conducted. The model tends to label these actions as "Yes" even the action did not happen yet:
  - Ex: "Aurier looks to take on Jones down the right flank and draws the foul from the midfielder. **Gibbs-White will send a cross into the Liverpool box**."
- Some set piece actions which does not explicitly declare the crossing actions and probably is a shooting or a passing action are labeled as "Yes".
  - Ex: "Maddison takes a free-kick from close to 30-yards out and fires it at Neto, who claims the ball easily."
  - Ex: "United force another corner down their right. This one's less successful, but the momentum's undoubtedly with the hosts now."
  - Ex: "Rashford shows good strength to hold off Alexander-Arnold and win a free-kick for United just outside the box. Eriksen swings it towards goal and it deflects off Liverpool's wall to hand United an early corner."
- Squaring the ball is not a type of crossing and the model labeled such commentary as "Yes".
  - "OVER! Summerville sends the ball over from close range. The substitute passed it into Bamford who pushed it wide to Aaronson. **The American then squared it back to Summerville** who could not find the target."
- Some set piece action that involves crossing are mislabeled as "No".

- Ex: “1-0 CRYSTAL PALACE!!!!! Eze sends in a brilliant free-kick and the ball deflects off Walker, onto Stones, and into the back of the net. Perfect start for Palace!”
- Ex: “Roca's curling corner is collected by Ederson.”

- Some crossing actions that are explained with words like lift, high ball etc. are mislabeled as “No”.
  - Ex: “It's better from Leicester here as Tielemans gets his head up and tries to **lift it into the box.** Maddison makes a great run off the back of Laporte and almost reaches it, but Ederson just gets there first.”

3.3.2. Crossing Player Classification with finetuned LLM Models:

The cross validation is applied on the dataset that contains only crossing actions to determine the crossing player for the classification task. In preliminary findings, it is observed that the model can select one of the players who makes the cross for the commentaries that have more than one crossing action. Therefore these 247 commentaries containing more than one crossing action are eliminated from the dataset before training for different folds. The final dataset containing crossing actions becomes 4124 instances.

All the players that have license during 2022-23 Premier League season are saved as a unique label. Also, as the commentaries only describe the surnames of the players, the players’ surnames are determined as the labels.

Optuna library is used to tune the hyperparameters. As the authors of the BERT model suggested, the following hyperparameters are considered as the input:

- Batch Size: 8, 16, 32, 64
- Learning Rate: 1e-5, 3e-5, 5e-5

The dataset split is done with match split method by applying 20 epochs to find the best model. The hyperparameters that resulted with least test loss with 0.128 are founded as:

- batch_size: 8,
- learning_rate: 5e-05

After finding the best hyperparameters, the cross-validation procedure is conducted with the best hyperparameters. 20 epochs are considered for the training process of all strategies. Each epoch consists of 310 batches with shape [8,128].

As 0.6-0.2-0.2 ratios for train-validation-test datasets are considered, k=5-fold cross-validation is applied for random split and match based split methods. The dataset is randomly divided 5 subsets (or folds) of approximately equal size for all strategies. To get 0.6-0.2-0.2 ratios for train-validation-test datasets, k=10-fold cross-validation is applied for team based split method.
The model is then trained 5 times for random and match based splitting strategies and 10 times for team based strategies, each time using a different fold as the test set and

the remaining folds as the training and validation set and the performance metrics accuracy and macro F1 score are calculated for each iteration. Also, loss characteristics for each iteration is depicted for all strategies. Only the first fold's loss graphs and confusion matrices are shared for each strategy below.

3.3.2.1. Crossing Player Classification with Random Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for random splitting method. In other words, 2474 commentaries are used for training, 825 commentaries are used for validation and 825 commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing check classification with random splitting strategy are shared in Figure 14, Figure 15 and Figure 16 respectively.

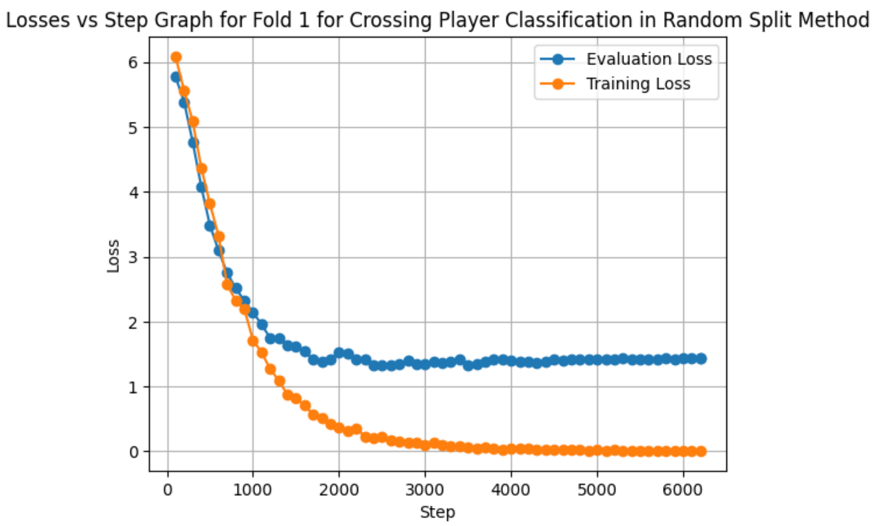


Figure 14 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Random Splitting Strategy

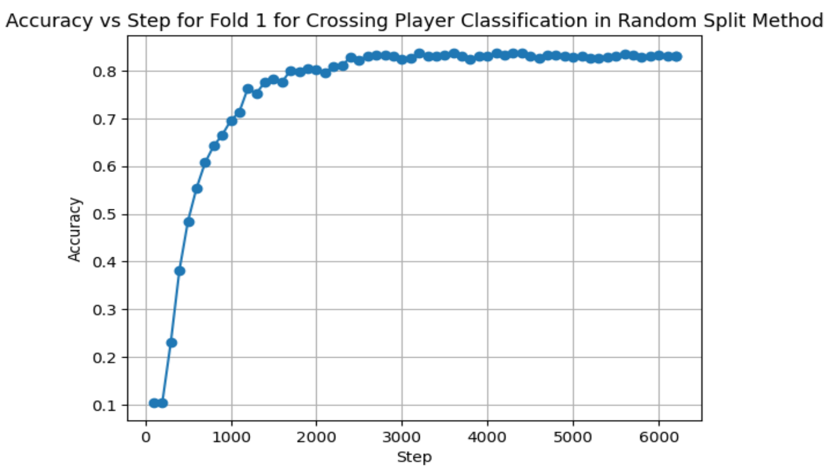


Figure 15 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Random Splitting Strategy

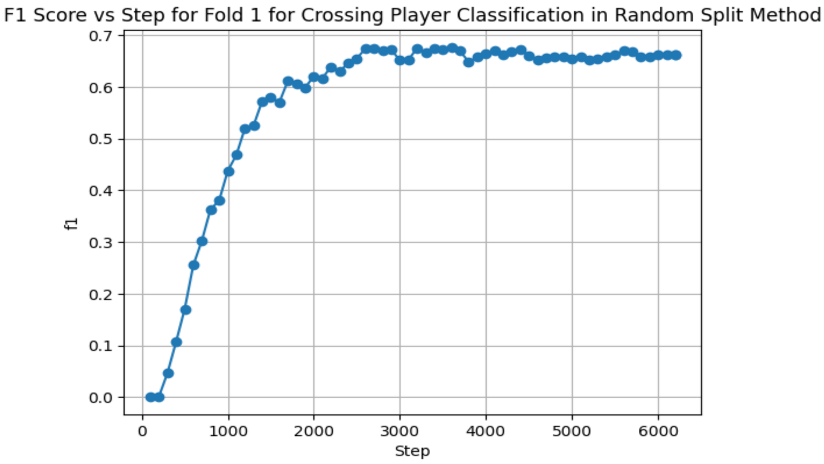


Figure 16 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Random Splitting Strategy

The confusion matrix formed with test dataset in Fold 1 for crossing player classification with Match Based Splitting Strategy is given in Table 21 in which the rows represent the actual labels and columns represents the predicted labels. The confusion matrix represents only the players that have records at least 5 times.

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, macro F1 score, precision, and recall results for each fold, are given in Table 19. The mean values and standard deviation of these metrics are also shown in Table 19. The average accuracy for all folds is calculated as 84.3% and the average macro F1 score is calculated as 0.684.

Table 19 Performance metrics of fine-tuned model for each fold in Crossing Player Classification with Random Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.844	0.668	0.673	0.693
2	0.836	0.661	0.668	0.688
3	0.833	0.685	0.694	0.710
4	0.841	0.703	0.701	0.736
5	0.859	0.703	0.706	0.728
Mean	**0.843**	**0.684**	**0.688**	**0.711**
Std Dev	**0.010**	**0.020**	**0.017**	**0.021**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Not Determined”. Using this label as the as the classifier, the performance metrics are founded as in Table 20. It can be seen that, the average accuracy is increased by 73.1% and macro F1 score is increased by 0.683 using the fine-tuned model.

Table 20 Performance metrics of baseline model for each fold in Crossing Player Classification with Random Splitting Strategy

fold	accuracy	Macro F1 score	precision	recall
1	0.122	0.001	0.001	0.004
2	0.109	0.001	0.000	0.004
3	0.101	0.001	0.000	0.004
4	0.115	0.001	0.001	0.004
5	0.114	0.001	0.001	0.005
Mean	**0.112**	**0.001**	**0.000**	**0.004**
Std Dev	**0.008**	**0.000**	**0.000**	**0.000**

Table 21 Confusion Matrix of test dataset for Fold 1 in Crossing Player Classification with Random Splitting Strategy (players that have record at least 5 times)

	Not Determined	Saka	Trossard	Xhaka	Luiz	Tavernier	Mbeumo	Jensen	March	Gross	Estupinan	Ayew	Zaha	Olise	Eze	McNeil	Gray	Iwobi	Kebano	Willian	Pereira	Robinson	Harrison	Barnes	Maddison	Robertson	Alexander-Arnold	Grealish	De Bruyne	Foden	Shaw	Saint-Maximin	Trippier	Johnson	Gibbs-White	Ward-Prowse	Son	Perisic	Porro	Bowen	Benrahma	Rice	Cresswell	Podence
Not Determined	75	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	3	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1
Saka	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Trossard	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Xhaka	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Luiz	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tavernier	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mbeumo	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jensen	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
March	0	0	0	0	0	0	0	0	7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gross	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Estupinan	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ayew	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zaha	0	0	0	0	0	0	0	0	0	0	0	1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Olise	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Eze	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
McNeil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gray	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Iwobi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kebano	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Willian	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pereira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Robinson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Harrison	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Barnes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maddison	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Robertson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alexander-Arnold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Grealish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
De Bruyne	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Foden	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Shaw	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
Saint-Maximin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
Trippier	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0
Johnson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
Gibbs-White	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
Ward-Prowse	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0
Son	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
Perisic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0
Porro	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
Bowen	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0
Benrahma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
Rice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0
Cresswell	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
Podence	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5

3.3.2.2. Crossing Player Classification with Match Based Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for match based splitting method. In other words, 0.6 of the 304 matches, 182 matches' commentaries are used for training, 61 matches' commentaries are used for validation and 61 matches' commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing player classification with match based splitting strategy are shared in Figure 17, Figure 18 and Figure 19 respectively.

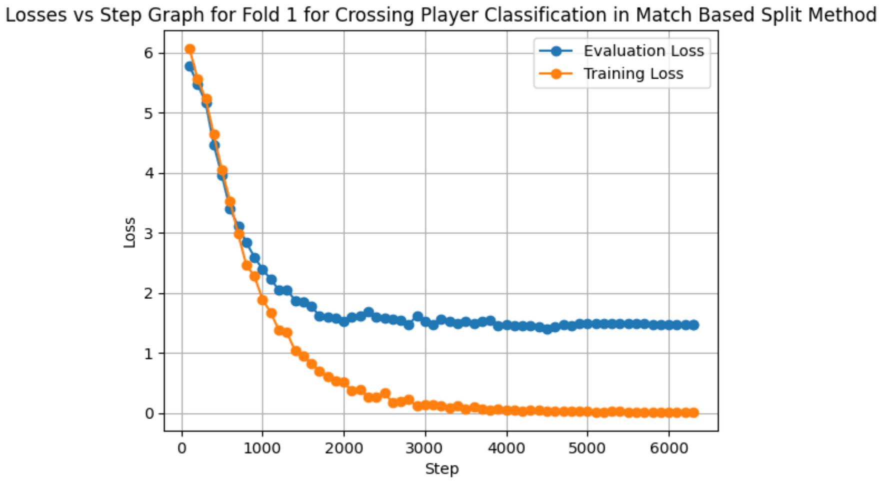


Figure 17 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Match Based Splitting Strategy

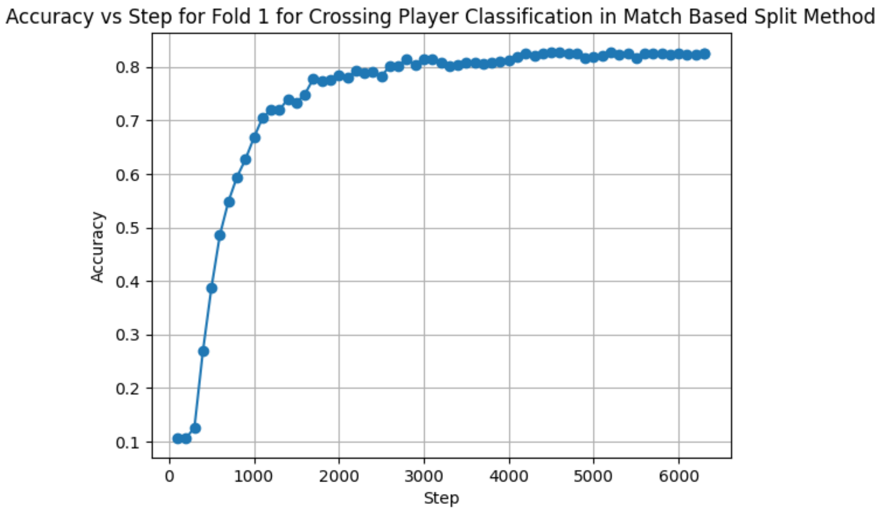


Figure 18 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Match Based Splitting Strategy

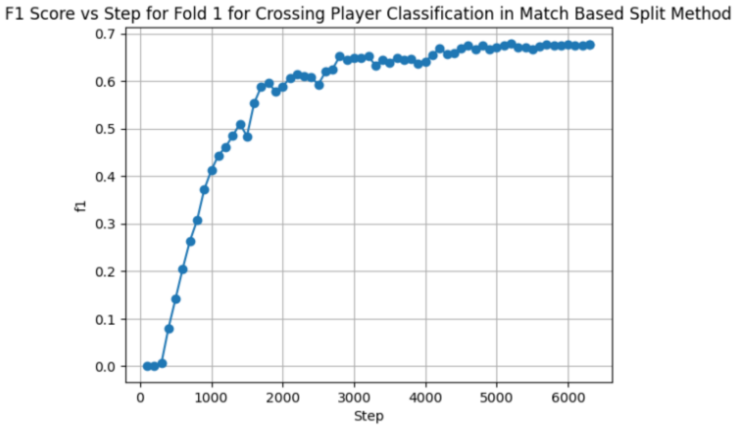


Figure 19 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Match Based Splitting Strategy

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, macro F1 score, precision, and recall results for each fold, are given in Table 22. The mean values and standard deviation of these metrics are also shown in Table 22. The average accuracy for all folds is calculated as 82.4% and the average macro F1 score is calculated as 0.652.

Table 22 Performance metrics of fine-tuned model for each fold in Crossing Player Classification with Match Based Splitting Strategy

fold	**accuracy**	**macro F1 score**	**precision**	**recall**
1	0.813	0.624	0.625	0.659
2	0.824	0.651	0.649	0.683
3	0.801	0.648	0.659	0.679
4	0.834	0.647	0.642	0.685
5	0.848	0.691	0.691	0.718
Mean	**0.824**	**0.652**	**0.653**	**0.685**
Std Dev	**0.018**	**0.024**	**0.025**	**0.022**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Not Determined”. Using this label as the as the classifier, the performance metrics are founded as in

Table 23. It can be seen that, the average accuracy is increased by 71.2% and macro F1 score is increased by 0.651 using the fine-tuned model.

Table 23 Performance metrics of baseline model for each fold in Crossing Player Classification with Match Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.114	0.001	0.001	0.005
2	0.114	0.001	0.001	0.005
3	0.115	0.001	0.001	0.005
4	0.120	0.001	0.001	0.005
5	0.098	0.001	0.000	0.005
Mean	**0.112**	**0.001**	**0.001**	**0.005**
Std Dev	**0.008**	**0.000**	**0.000**	**0.000**

3.3.2.3. Crossing Player Classification with Team Based Splitting Strategy:

There are 20 different teams in the datasets and these teams are split into train-validation-test to get roughly 0.6-0.2-0.2 ratios for them. Thus, 2 teams are selected for test datasets and 2 teams are selected for validation datasets and 16 for training dataset which leads to 10 folds. The teams that are tested in each fold is given as below:

- Fold 1: ['Arsenal', 'Tottenham Hotspur']
- Fold 2: ['Aston Villa', 'Nottingham Forest']
- Fold 3: ['Chelsea', 'Fulham']
- Fold 4: ['Brentford', 'Liverpool']
- Fold 5: ['Southampton', 'West Ham United']
- Fold 6: ['Bournemouth', 'Manchester United']
- Fold 7: ['Leeds United', 'Wolverhampton Wanderers']
- Fold 8: ['Brighton', 'Manchester City']
- Fold 9: ['Everton', 'Leicester City']
- Fold 10: ['Crystal Palace', 'Newcastle United']

The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing player classification with team based splitting strategy are shared in Figure 20, Figure 21 and Figure 22 respectively.

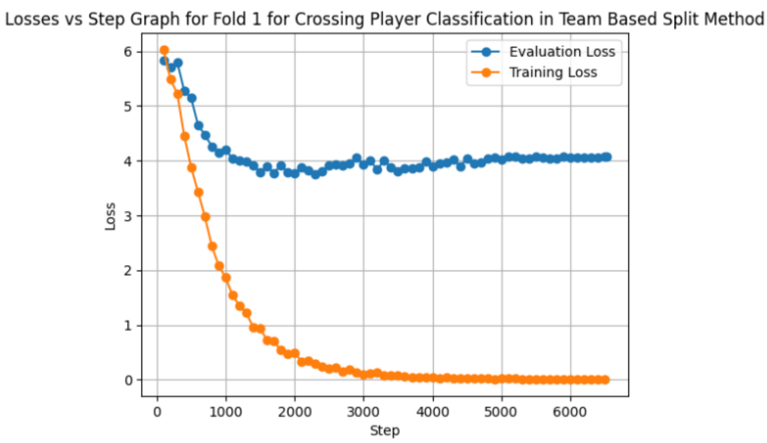


Figure 20 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Team Based Splitting Strategy

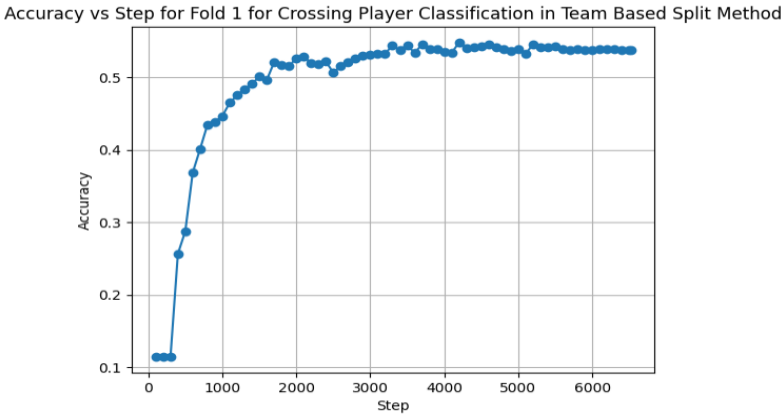


Figure 21 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Team Based Splitting Strategy

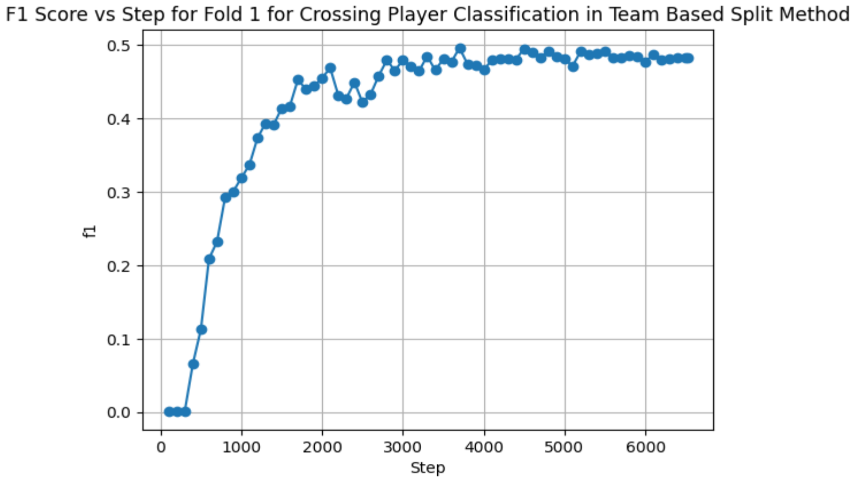


Figure 22 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Player Classification with Team Based Splitting Strategy

The performance metrics of the fine-tuned model calculated for the test dataset, including accuracy, macro F1 score, precision, and recall results for each fold, are given in Table 24. The mean values and standard deviation of these metrics are also shown in Table 24. The average accuracy for all folds is calculated as 45.5% and the average macro F1 score is calculated as 0.404.

Table 24 Performance metrics of fine-tuned model for each fold in Crossing Player Classification with Team Based Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.404	0.388	0.373	0.453
2	0.556	0.441	0.418	0.527
3	0.524	0.446	0.430	0.519
4	0.400	0.374	0.346	0.471
5	0.453	0.455	0.431	0.535
6	0.498	0.385	0.367	0.453

Table 24 (cont.)

7	0.451	0.376	0.365	0.447
8	0.352	0.327	0.305	0.405
9	0.478	0.465	0.444	0.526
10	0.431	0.384	0.362	0.454
Mean	**0.455**	**0.404**	**0.384**	**0.479**
Std Dev	**0.061**	**0.045**	**0.045**	**0.045**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Not Determined”. Using this label as the as the classifier, the performance metrics are founded as in Table 25. It can be seen that, the average accuracy is increased by 44.2% and macro F1 score is increased by 0.403 using the fine-tuned model.

Table 25 Performance metrics of baseline model for each fold in Crossing Player Classification with Team Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.116	0.001	0.001	0.006
2	0.104	0.001	0.001	0.005
3	0.093	0.001	0.001	0.005
4	0.137	0.001	0.001	0.006
5	0.092	0.001	0.001	0.006
6	0.112	0.001	0.001	0.005
7	0.141	0.001	0.001	0.006
8	0.107	0.001	0.001	0.007
9	0.124	0.001	0.001	0.006
10	0.107	0.001	0.001	0.006
Mean	**0.113**	**0.001**	**0.001**	**0.006**
Std Dev	**0.016**	**0.000**	**0.000**	**0.000**

3.3.2.4. Crossing Player Classification Results of the Strategies:

The average metric results after cross validation applied for each strategy on test datasets are given in Table 26. The high standard deviations for all metrics are caused by the low prediction performance of the team split strategy. The results obtained from the model of team split strategy was expected as model was not learning anything about the players that was not mentioned even once. Overall, for all strategies the average accuracy is founded as 70.7% and macro F1 score is founded as 0.58.

Table 26 Average metric results of fine-tuned models for crossing player classification after cross validation applied for each strategy on test datasets

Strategy	accuracy	Macro F1 score	precision	recall
Random Split	0.843	0.684	0.688	0.711
Match Split	0.824	0.652	0.653	0.685
Team Split	0.455	0.404	0.384	0.479
Mean	**0.707**	**0.580**	**0.575**	**0.625**
Std Dev	**0.219**	**0.153**	**0.166**	**0.127**

When the test datasets are investigated, the common main reasons for mislabeling are observed as below:

- The model can make mistakes about the crossing player and receiving player:
    - Ex: "Kane finds him in possession down the right and looks up in an attempt to find Son. The Englishman's cross is too powerful and bounces behind for a Forest goal-kick." (Actual: Kane, Predicted: Son)
- The model can make mistakes about the crossing player and the defending players:
    - Ex: "CHANCE! Young gets the better of De Cordova-Reid down the left and swings a good cross into the box. He picks out Watkins in the middle, but Robinson gets across just in time to block his shot." (Actual: Young, Predicted: De Cordova-Reid)

3.3.3. Crossing Outcome Classification with finetuned LLM Models:

The cross validation is applied on the dataset that contains only crossing actions to determine the crossing outcome for the classification task. In preliminary findings, it is observed that the model can select one of the outcomes for the commentaries that have more than one crossing action. Therefore these 247 commentaries containing more than one crossing action are eliminated from the dataset before training for different folds. Furthermore, eight commentaries that were labeled as "Not Sure" by the labelers are eliminated too. The final dataset containing crossing actions becomes 4116 instances.

Optuna library is used to tune the hyperparameters. As the authors of the BERT model suggested, the following hyperparameters are considered as the input:

- Batch Size: 8, 16, 32, 64
- Learning Rate: 1e-5, 3e-5, 5e-5

The dataset split is done with match split method by applying 20 epochs to find the best model. The hyperparameters that resulted with least test loss with 0.494 are founded as:

- batch_size: 8,
- learning_rate: 5e-05

After finding the best hyperparameters, the cross-validation procedure is conducted with the best hyperparameters. 20 epochs are considered for the training process of all strategies. Each epoch consists of 309 batches with shape [8,128].

As 0.6-0.2-0.2 ratios for train-validation-test datasets are considered, k=5-fold cross-validation is applied for random split and match based split strategies. The dataset is randomly divided 5 subsets (or folds) of approximately equal size for all strategies. To get 0.6-0.2-0.2 ratios for train-validation-test datasets, k=10-fold cross-validation is applied for team based split method.
The model is then trained 5 times for random and match based splitting strategies and 10 times for team based strategies, each time using a different fold as the test set and the remaining folds as the training and validation set and the performance metrics accuracy and macro F1 score are calculated for each iteration. Also, loss characteristics for each iteration is depicted for all strategies. Only the first fold's loss graphs and confusion matrices are shared for each strategy below to observe the performance of the fine-tuned model for crossing outcome classification.

3.3.3.1. Crossing Outcome Classification with Random Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for random splitting method. In other words, 2470 commentaries are used for training, 823 commentaries are used for validation and 823 commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing check classification with random splitting strategy are shared in Figure 23, Figure 24 and Figure 25 respectively.

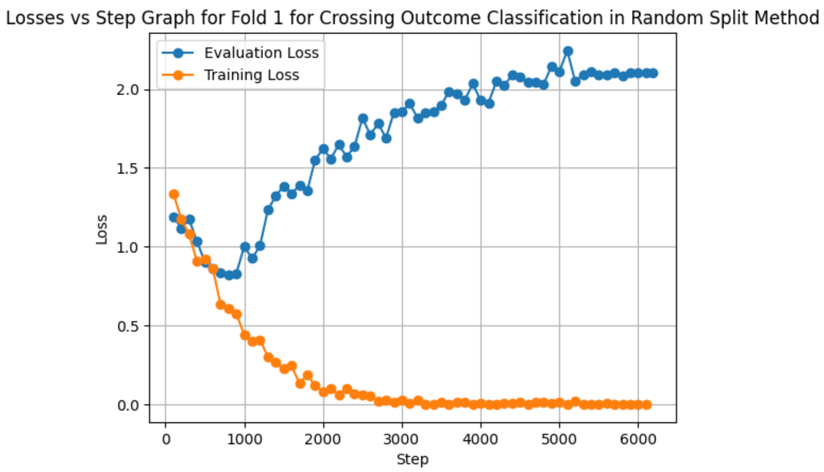


Figure 23 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Random Splitting Strategy

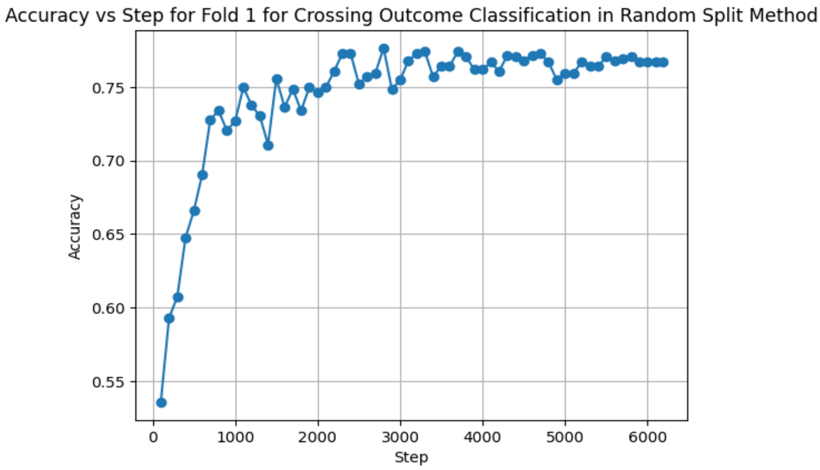


Figure 24 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Random Splitting Strategy

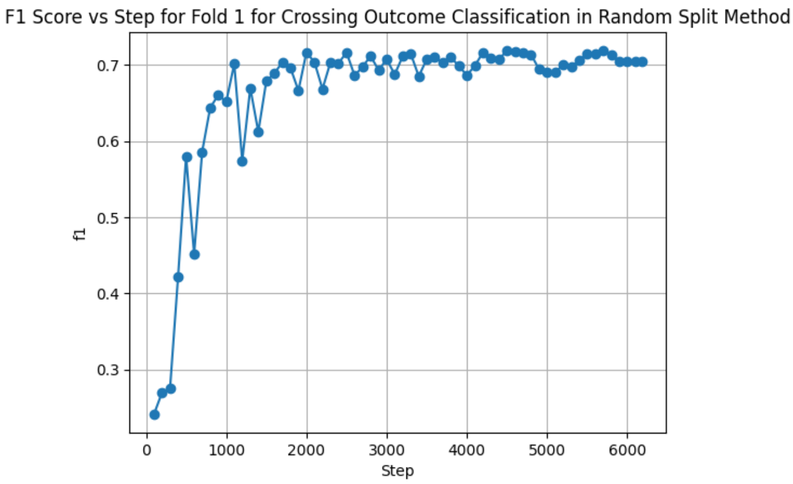


Figure 25 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Random Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing outcome classification with Random Splitting Strategy is given in Figure 26. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing outcome classification with Random Splitting Strategy is given in Table 27 in which the rows represent the actual labels and columns represents the predicted labels.

Table 27 Confusion Matrix of test dataset for Fold 1 in Crossing Outcome Classification with Random Splitting Strategy

	Cross on target, reaches a teammate	**Cross off target, not reaches a teammate**	**Defender Int.**	**Goalkeeper Int.**	**Referee Int.**
Cross on target, reaches a teammate	285	4	35	10	2
Cross off target, not reaches a teammate	11	53	13	7	1
Defender Interception	36	10	225	19	7
Goalkeeper Interception	11	3	11	56	0
Referee Interception	5	2	5	2	11

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 28. The mean values and standard deviation of these metrics are also shown in in Table 28. The average accuracy for all folds is calculated as 77.8% and the average macro F1 score is calculated as 0.705.

Table 28 Performance metrics of fine-tuned model for each fold in Crossing Outcome Classification with Random Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.765	0.679	0.691	0.672
2	0.808	0.748	0.769	0.730
3	0.763	0.685	0.728	0.661
4	0.776	0.705	0.741	0.681
5	0.776	0.709	0.706	0.717
Mean	**0.778**	**0.705**	**0.727**	**0.692**
Std Dev	**0.018**	**0.027**	**0.030**	**0.030**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Cross On Target”. Using this label as the as the classifier, the performance metrics are founded as in Table 29. It can be seen that, the average accuracy is increased by 34.9% and macro F1 score is increased by 0.585 using the fine-tuned model.

Table 29 Performance metrics of baseline model for each fold in Crossing Outcome Classification with Random Splitting Strategy

fold	accuracy	Macro F1 score	precision	recall
1	0.408	0.116	0.082	0.200
2	0.459	0.126	0.092	0.200
3	0.420	0.118	0.084	0.200
4	0.422	0.119	0.084	0.200
5	0.436	0.121	0.087	0.200
Mean	**0.429**	**0.120**	**0.086**	**0.200**
Std Dev	**0.020**	**0.004**	**0.004**	**0.000**

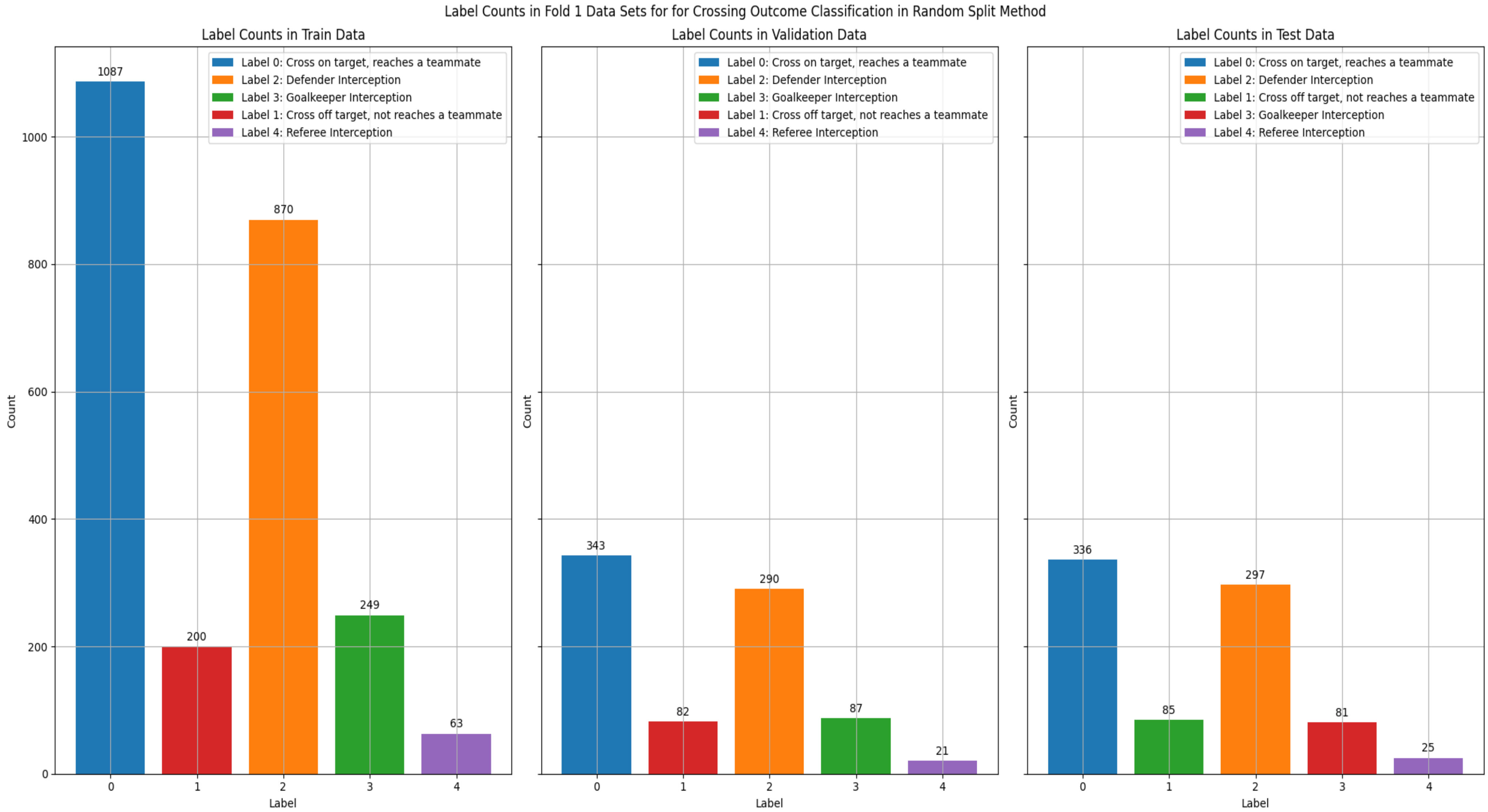


Figure 26 Label distributions of train-validation-test datasets in Fold 1 for Crossing Outcome Classification with Random Splitting Strategy

3.3.3.2. Crossing Outcome Classification with Match Based Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for match based splitting method. In other words, 0.6 of the 304 matches, 182 matches' commentaries are used for training, 61 matches' commentaries are used for validation and 61 matches' commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing outcome classification with match based splitting strategy are shared in Figure 27, Figure 28 and Figure 29 respectively.

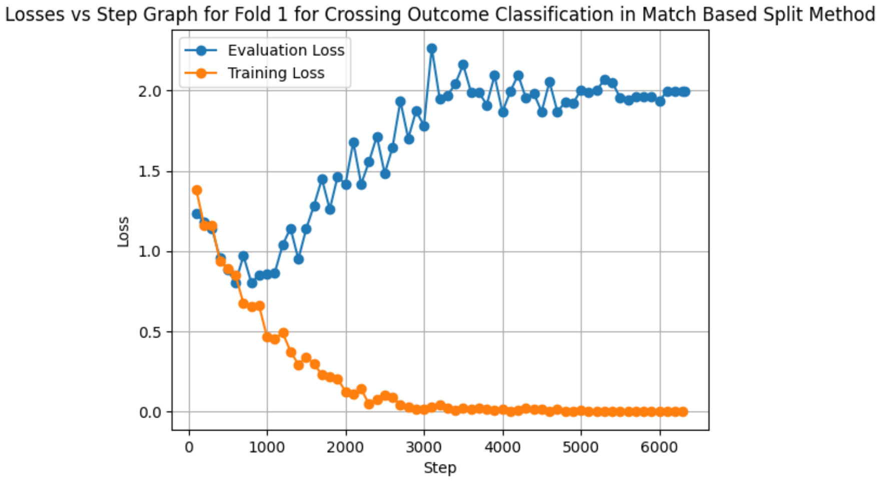


Figure 27 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Match Based Splitting Strategy

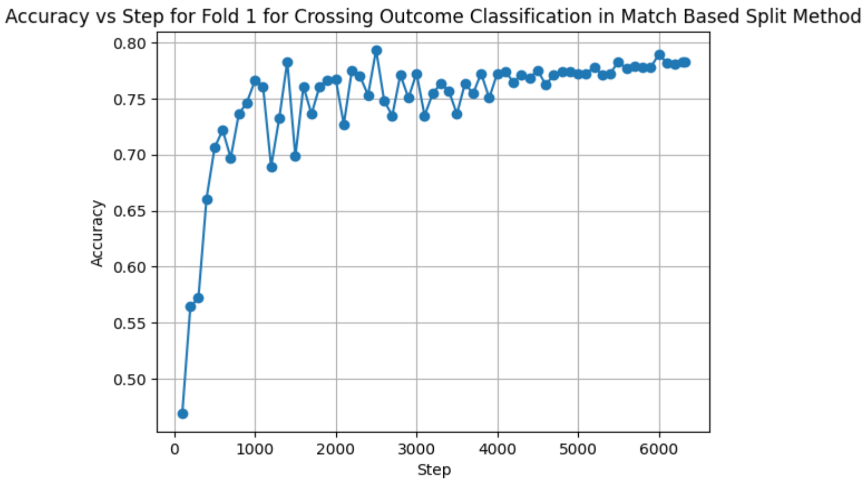


Figure 28 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Match Based Splitting Strategy

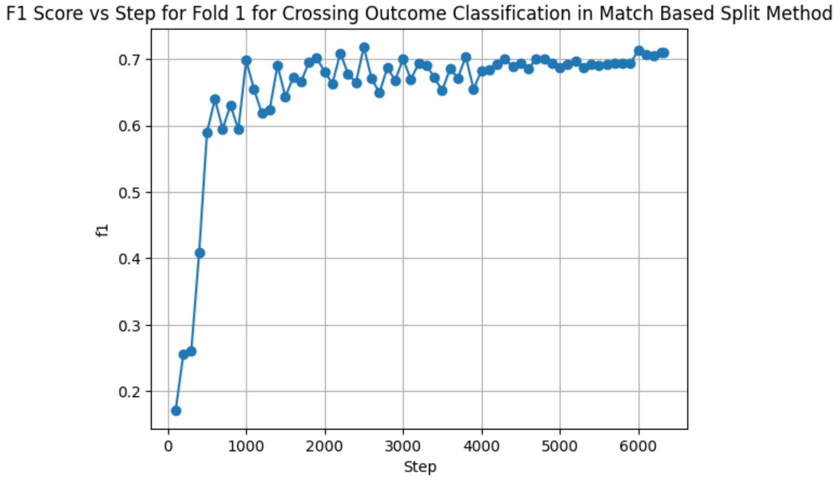


Figure 29 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Match Based Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing outcome classification with Match Based Splitting Strategy is given in Figure 30. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing outcome classification with Match Based Splitting Strategy is given in Table 30 in which the rows represent the actual labels and columns represents the predicted labels.

Table 30 Confusion Matrix of test dataset for Fold 1 in Crossing Outcome Classification with match Based Splitting Strategy

	Cross on target, reaches a teammate	**Cross off target, not reaches a teammate**	**Defender Int.**	**Goalkeeper Int.**	**Referee Int.**
Cross on target, reaches a teammate	307	18	32	11	2
Cross off target, not reaches a teammate	5	54	9	2	2
Defender Interception	33	14	227	7	4
Goalkeeper Interception	10	2	6	62	1
Referee Interception	2	1	2	1	10

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 31. The mean values and standard deviation of these metrics are also shown in in Table 31. The average accuracy for all folds is calculated as 77.6% and the average macro F1 score is calculated as 0.704.

Table 31 Performance metrics of fine-tuned model for each fold in Crossing Outcome Classification with Random Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.801	0.730	0.712	0.753
2	0.786	0.719	0.736	0.706
3	0.777	0.702	0.765	0.676
4	0.757	0.665	0.670	0.664
5	0.758	0.705	0.735	0.680
Mean	**0.776**	**0.704**	**0.724**	**0.696**
Std Dev	**0.019**	**0.025**	**0.035**	**0.035**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Cross On Target”. Using this label as the as the classifier, the performance metrics are founded as in Table 32. It can be seen that, the average accuracy is increased by 34.7% and macro F1 score is increased by 0.584 using the fine-tuned model.

Table 32 Performance metrics of baseline model for each fold in Crossing Outcome Classification with Random Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.449	0.124	0.090	0.200
2	0.429	0.120	0.086	0.200
3	0.415	0.117	0.083	0.200
4	0.430	0.120	0.086	0.200
5	0.422	0.119	0.084	0.200
Mean	**0.429**	**0.120**	**0.086**	**0.200**
Std Dev	**0.013**	**0.002**	**0.003**	**0.000**

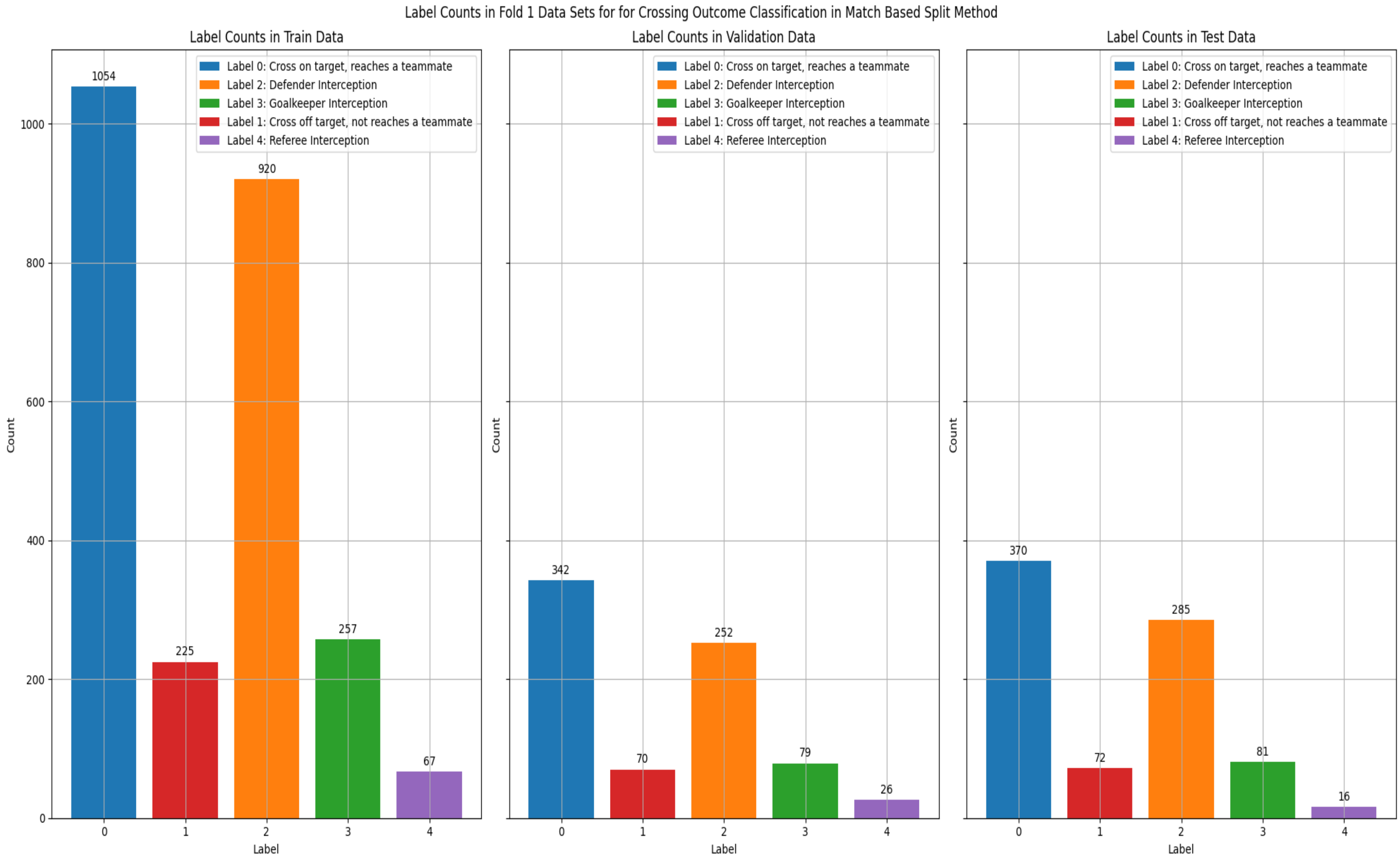


Figure 30 Label distributions of train-validation-test datasets in Fold 1 for Crossing Outcome Classification with Random Splitting Strategy

3.3.3.3. Crossing Outcome Classification with Team Based Splitting Strategy:

There are 20 different teams in the datasets and these teams are split into train-validation-test to get roughly 0.6-0.2-0.2 ratios for them. Thus, 2 teams are selected for test datasets and 2 teams are selected for validation datasets and 16 for training dataset which leads to 10 folds. The teams that are tested in each fold is given as below:

- Fold 1: ['Arsenal', 'Tottenham Hotspur']
- Fold 2: ['Aston Villa', 'Nottingham Forest']
- Fold 3: ['Chelsea', 'Fulham']
- Fold 4: ['Brentford', 'Liverpool']
- Fold 5: ['Southampton', 'West Ham United']
- Fold 6: ['Bournemouth', 'Manchester United']
- Fold 7: ['Leeds United', 'Wolverhampton Wanderers']
- Fold 8: ['Brighton', 'Manchester City']
- Fold 9: ['Everton', 'Leicester City']
- Fold 10: ['Crystal Palace', 'Newcastle United']

The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing outcome classification with team based splitting strategy are shared in Figure 31, Figure 32 and Figure 33 respectively.

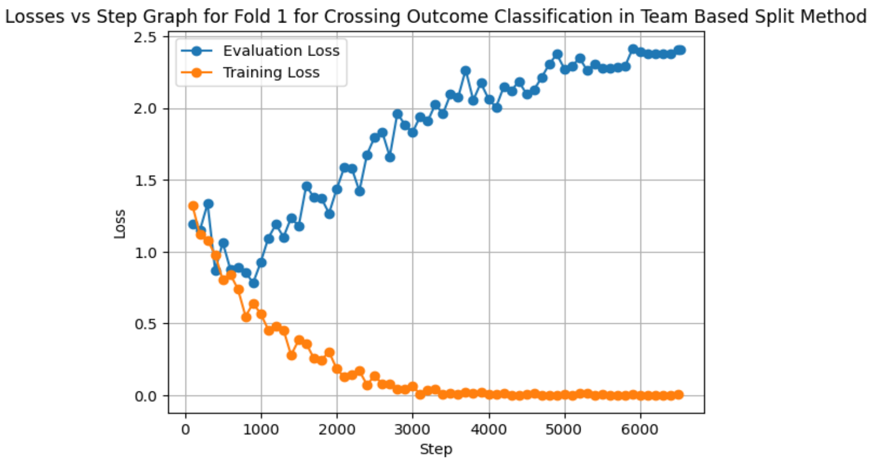


Figure 31 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Team Based Splitting Strategy

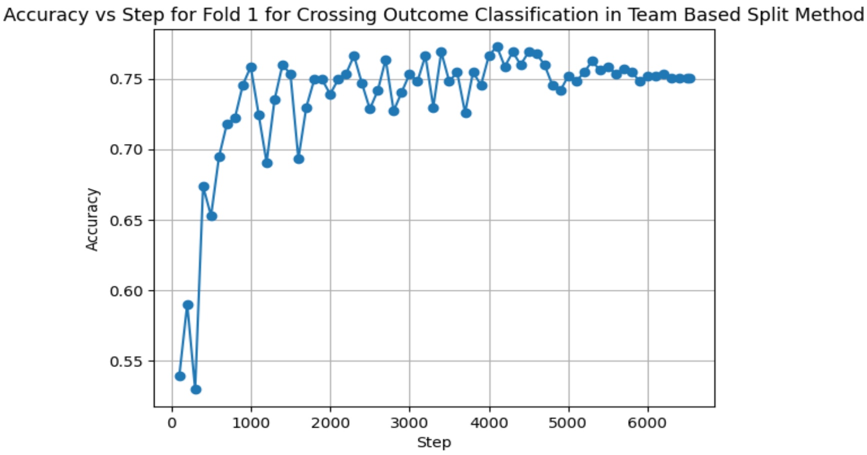


Figure 32 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Team Based Splitting Strategy

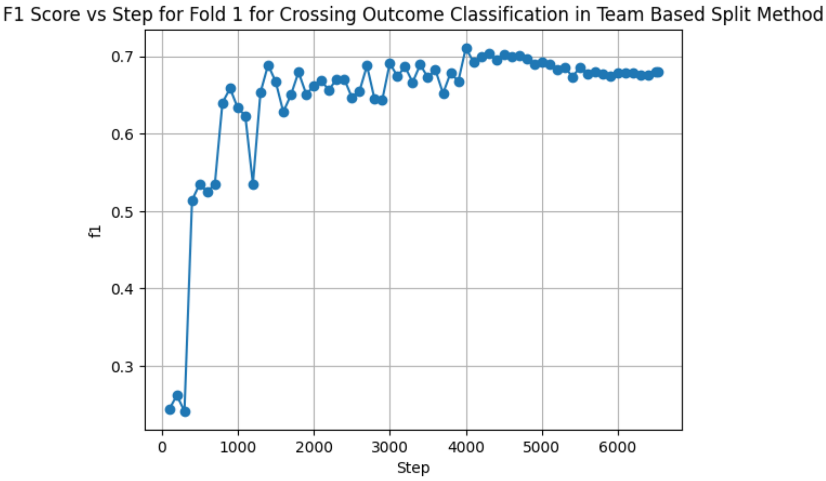


Figure 33 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Outcome Classification with Team Based Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing outcome classification with Team Based Splitting Strategy is given in Figure 34. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing outcome classification with Match Based Splitting Strategy is given in Table 33 in which the rows represent the actual labels and columns represents the predicted labels.

Table 33 Confusion Matrix of test dataset for Fold 1 in Crossing Outcome Classification with Team Based Splitting Strategy

	Cross on target, reaches a teammate	**Cross off target, not reaches a teammate**	**Defender Interception**	**Goalkeeper Interception**	**Referee Interception**
Cross on target, reaches a teammate	250	10	40	6	7
Cross off target, not reaches a teammate	14	45	12	1	0
Defender Interception	33	7	227	8	6
Goalkeeper Interception	9	2	10	38	0
Referee Interception	1	2	5	1	15

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 34. The mean values and standard deviation of these metrics are also shown in Table 34. The average accuracy for all folds is calculated as 77.6% and the average macro F1 score is calculated as 0.704.

Table 34 Performance metrics of fine-tuned model for each fold in Crossing Check Classification with Team Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.768	0.700	0.702	0.700
2	0.740	0.669	0.704	0.644
3	0.771	0.675	0.700	0.656
4	0.798	0.710	0.732	0.692
5	0.778	0.681	0.703	0.666
6	0.812	0.731	0.746	0.720
7	0.774	0.673	0.726	0.642
8	0.778	0.724	0.727	0.724
9	0.772	0.686	0.719	0.663
10	0.785	0.710	0.719	0.704
Mean	**0.778**	**0.696**	**0.718**	**0.681**
Std Dev	**0.019**	**0.022**	**0.015**	**0.031**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in team based splitting strategy for each fold is founded as "Cross On Target". Using this label as the as the classifier, the performance metrics are founded as in Table 29. It can be seen that, the average accuracy is increased by 34.8% and macro F1 score is increased by 0.576 using the fine-tuned model.

Table 35 Performance metrics of baseline model for each fold in Crossing Check Classification with Team Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.418	0.118	0.084	0.200
2	0.414	0.117	0.083	0.200
3	0.431	0.120	0.086	0.200
4	0.431	0.120	0.086	0.200
5	0.414	0.117	0.083	0.200
6	0.452	0.125	0.090	0.200
7	0.418	0.118	0.084	0.200
8	0.453	0.125	0.091	0.200
9	0.445	0.123	0.089	0.200
10	0.422	0.119	0.084	0.200
Mean	**0.430**	**0.120**	**0.086**	**0.200**
Std Dev	**0.015**	**0.003**	**0.003**	**0.000**

3.3.3.4. Crossing Outcome Classification Results of the Strategies:

The average metric results after cross validation applied for each strategy on test datasets are given in

Table 36. The standard deviations for all metrics are low (<0.01) so it can be said that the average results of each strategy are nearly same. Overall, for all strategies the average accuracy is founded as 77.7% and macro F1 score is founded as 0.701.

Table 36 Average metric results of fine-tuned models for crossing outcome classification after cross validation applied for each strategy on test datasets

Strategy	**accuracy**	**Macro F1 score**	**precision**	**recall**
Random Split	0.776	0.704	0.724	0.696
Match Split	0.776	0.704	0.724	0.696
Team Split	0.778	0.696	0.718	0.681
Mean	**0.777**	**0.701**	**0.722**	**0.691**
Std Dev	**0.001**	**0.005**	**0.003**	**0.009**

- When examining the confusion matrices presented in Table 27, Table 30 and Table 33, it becomes evident that the most frequent mislabels by the fine-tuned model occur between the "cross on target" and "defender interception" labels. When these commentaries are investigated, it becomes apparent that the model often misinterprets the roles of the attacking and defending teams.
    - Ex: Another free-kick for Arsenal and this one is chipped in by Xhaka. He's looking for Odegaard on the edge of the box, but it's nodded straight back to him by Mepham. (Actual: Defender Interception, Predicted: Cross on Target)
    - Ex: ALMIRON EQUALISES! It's Saint-Maximin again with some brilliant work down the left and he whips another dangerous cross into the box. Willock misses it, but Almiron leans forward just enough to glance it in off his thigh. The flag went up, but he was onside and VAR says it stands! 1-1! City just haven't been able to deal with Saint-Maximin's threat in this game so far and he dribbled away from two players to get his cross in this time. (Actual: Cross on Target, Predicted: Defender Interception)

- Another common mislabels by the fine-tuned model occur between the "goalkeeper interception" and "defender interception" labels. Since the model does not know the roles of the players as a goalkeeper or a defender, if the action is not a goalkeeper-based action like punch or gather, the model can make a mistake in labeling.

    - Ex: “The low cross from Tavernier is smashed away by Wissa. (Actual: Defender Interception, Predicted: Goalkeeper Interception)

    - Ex: Aaronson attracts three Fulham defenders before teeing up Ayling, whose low, inviting cross is snatched upon by Leno!” (Actual: Goalkeeper Interception, Predicted: Defender Interception)

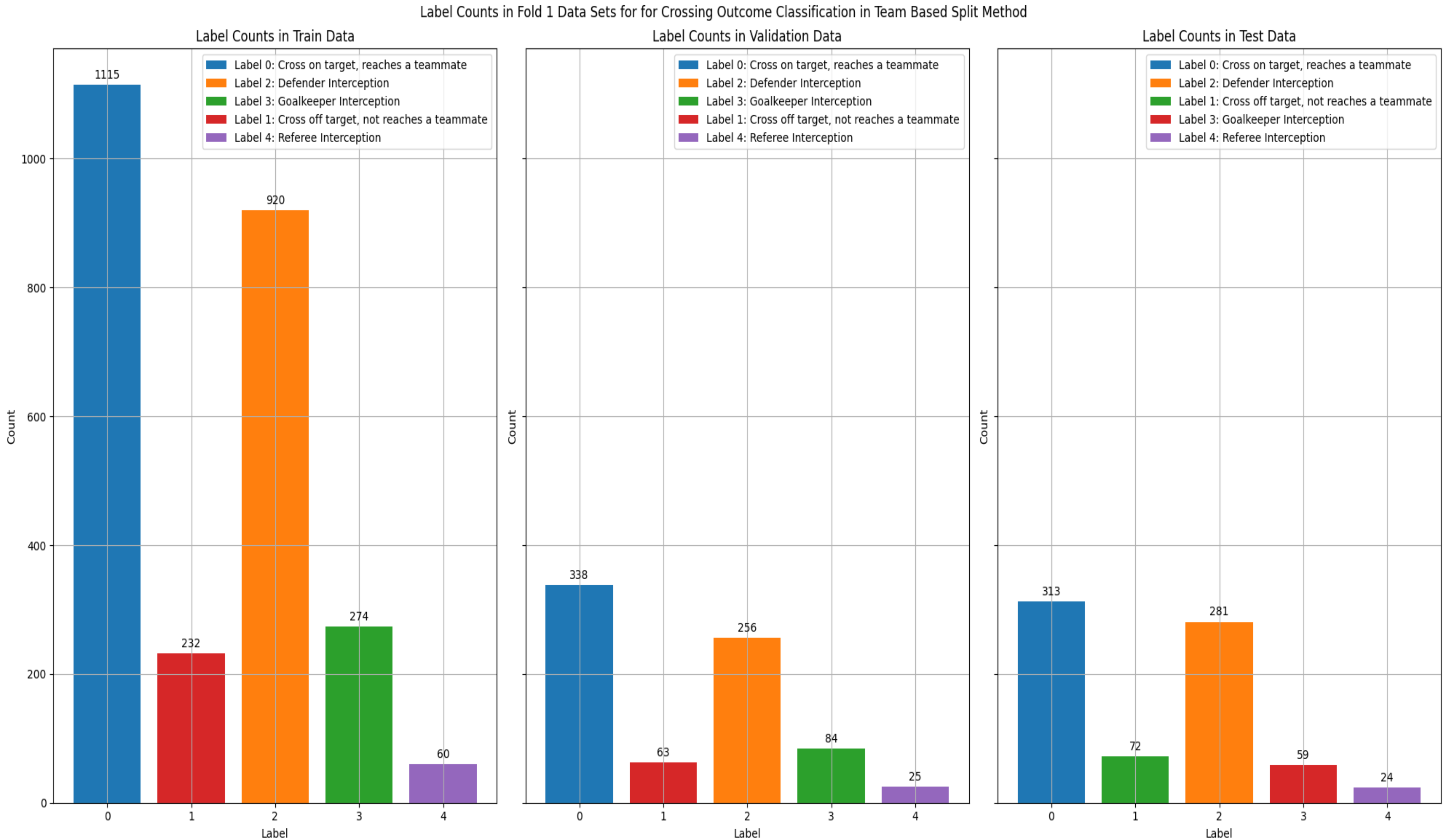


Figure 34 Label distributions of train-validation-test datasets in Fold 1 for Crossing Outcome Classification with Team Based Splitting Strategy

3.3.4. Crossing Quality Classification with finetuned LLM Models:

The cross validation is applied on the dataset that contains only crossing actions to determine the crossing outcome for the classification task. In preliminary findings, it is observed that the model can select one of the qualities for the commentaries that have more than one crossing action. Therefore these 247 commentaries containing more than one crossing action are eliminated from the dataset before training for different folds. The final dataset containing crossing actions becomes 4124 instances.

Optuna library is used to tune the hyperparameters. As the authors of the BERT model suggested, the following hyperparameters are considered as the input:

- Batch Size: 8, 16, 32, 64
- Learning Rate: 1e-5, 3e-5, 5e-5

The dataset split is done with match split method by applying 20 epochs to find the best model. The hyperparameters that resulted with least test loss with 1.19 are founded as:

- batch_size: 8,
- learning_rate: 5e-05

After finding the best hyperparameters, the cross-validation procedure is conducted with the best hyperparameters. 20 epochs are considered for the training process of all strategies. Each epoch consists of 310 batches with shape [8,128].

As 0.6-0.2-0.2 ratios for train-validation-test datasets are considered, k=5-fold cross-validation is applied for random split and match based split strategies. The dataset is randomly divided 5 subsets (or folds) of approximately equal size for all strategies. To get 0.6-0.2-0.2 ratios for train-validation-test datasets, k=10-fold cross-validation is applied for team based split method.
The model is then trained 5 times for random and match based splitting strategies and 10 times for team based strategies, each time using a different fold as the test set and the remaining folds as the training and validation set and the performance metrics accuracy and macro F1 score are calculated for each iteration. Also, loss characteristics for each iteration is depicted for all strategies. Only the first fold's loss graphs and confusion matrices are shared for each strategy below to observe the performance of the fine-tuned model for crossing quality classification.

3.3.4.1. Crossing Quality Classification with Random Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for random splitting method. In other words, 2474 commentaries are used for training, 825 commentaries are used for validation and 825 commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing quality classification with random splitting strategy are shared in Figure 35, Figure 36 and Figure 37 respectively.

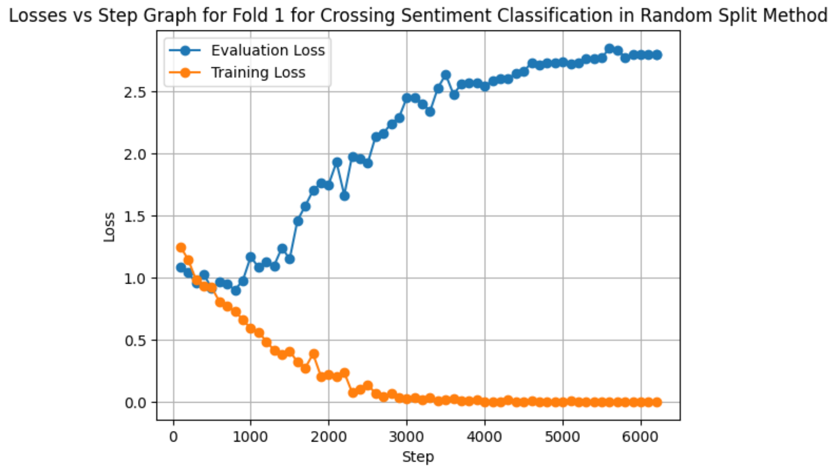


Figure 35 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Random Splitting Strategy

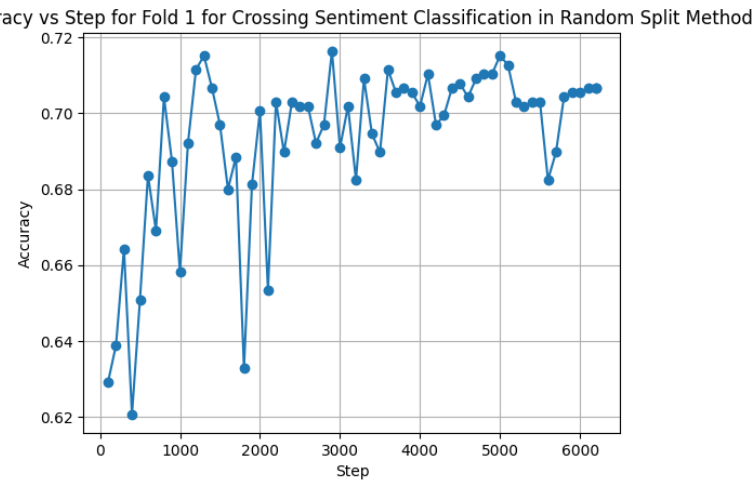


Figure 36 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Random Splitting Strategy

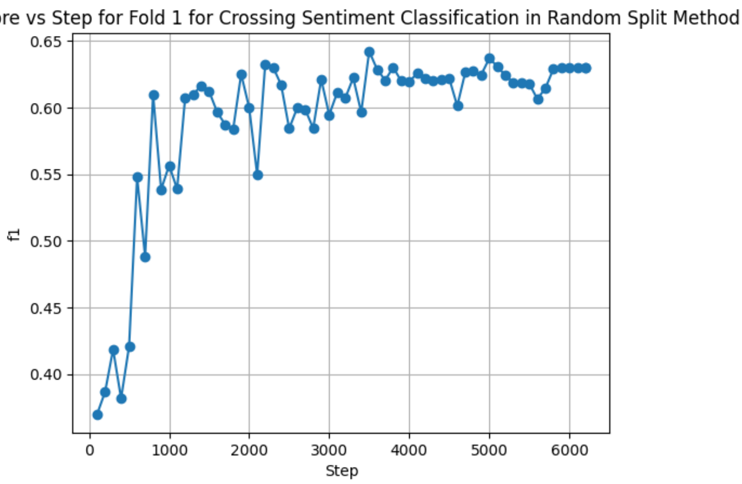


Figure 37 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Random Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing outcome classification with Random Splitting Strategy is given in Figure 38. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing quality classification with Random Splitting Strategy is given in Figure 34 in which the rows represent the actual labels and columns represents the predicted labels.

Table 37 Confusion Matrix of test dataset for Fold 1 in Crossing Quality Classification with Random Splitting Strategy

	Very Bad	**Bad**	**Neutral**	**Good**	**Very Good**
Very Bad	5	12	7	0	0
Bad	4	50	42	2	0
Neutral	3	29	335	41	4
Good	2	3	31	135	21
Very Good	0	1	2	19	77

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 38. The mean values and standard deviation of these metrics are also shown in Table 38. The average accuracy for all folds is calculated as 77.8% and the average macro F1 score is calculated as 0.705.

Table 38 Performance metrics of fine-tuned model for each fold in Crossing Quality Classification with Random Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.730	0.610	0.625	0.603
2	0.715	0.620	0.671	0.601
3	0.741	0.624	0.646	0.609
4	0.721	0.633	0.675	0.609
5	0.737	0.641	0.650	0.638
Mean	**0.729**	**0.626**	**0.653**	**0.612**
Std Dev	**0.011**	**0.012**	**0.020**	**0.015**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as “Neutral”. Using this label as the as the classifier, the performance metrics are founded as in Table 39. It can be seen that, the average accuracy is increased by 23.9% and macro F1 score is increased by 0.494 using the fine-tuned model.

Table 39 Performance metrics of baseline model for each fold in Crossing Quality Classification with Random Splitting Strategy

fold	accuracy	Macro F1 score	precision	recall
1	0.499	0.133	0.100	0.200
2	0.476	0.129	0.095	0.200
3	0.493	0.132	0.099	0.200
4	0.486	0.131	0.097	0.200
5	0.494	0.132	0.099	0.200
Mean	**0.490**	**0.132**	**0.098**	**0.200**
Std Dev	**0.009**	**0.002**	**0.002**	**0.000**

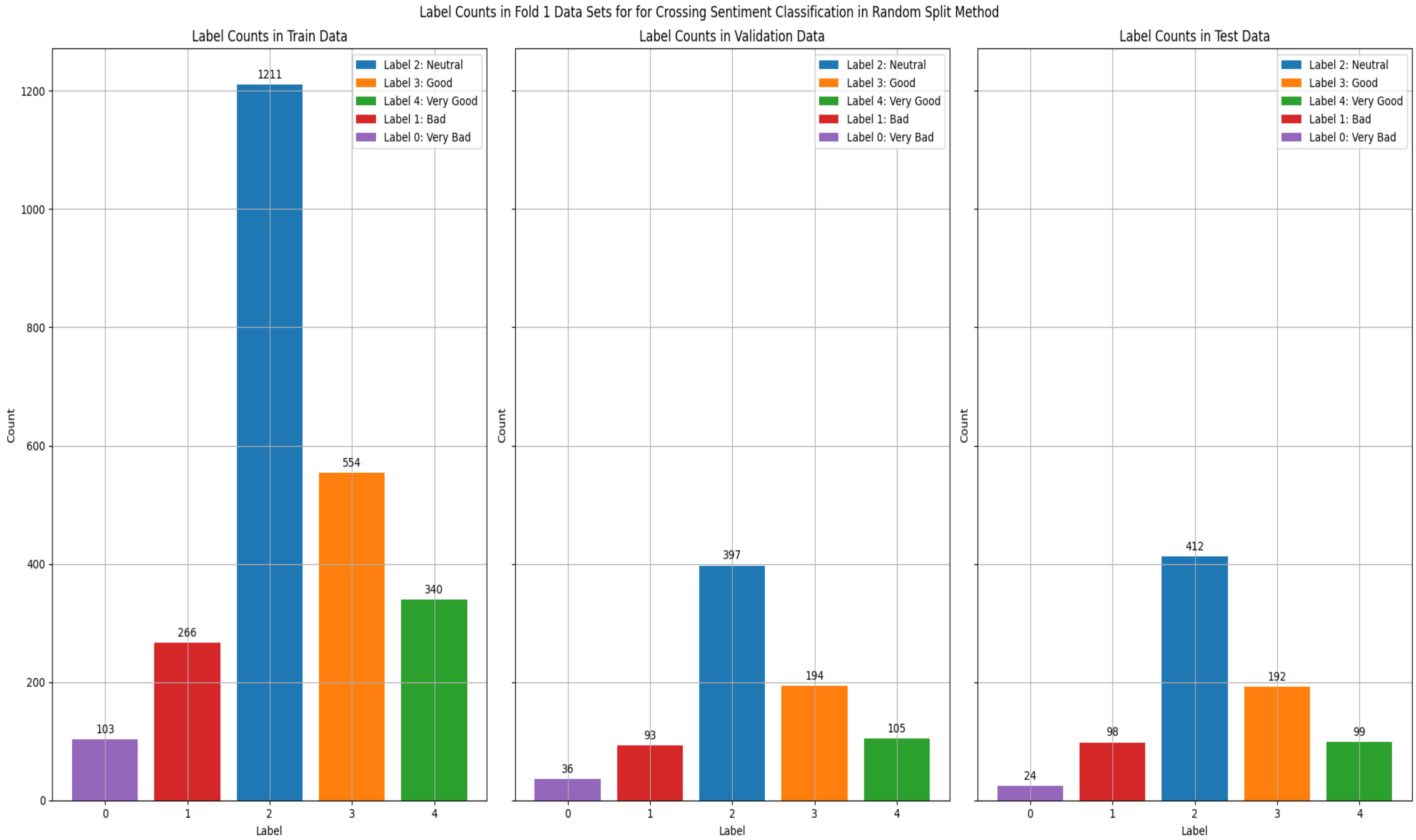


Figure 38 Label distributions of train-validation-test datasets in Fold 1 for Crossing Quality Classification with Random Splitting Strategy

3.3.4.2. Crossing Quality Classification with Match Based Splitting Strategy:

0.6-0.2-0.2 ratios for train-validation-test dataset is considered and k=5-fold cross-validation is applied for match based splitting method. In other words, 0.6 of the 304 matches, 182 matches' commentaries are used for training, 61 matches' commentaries are used for validation and 61 matches' commentaries are used for test datasets. The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing quality classification with match based splitting strategy are shared in Figure 39, Figure 40 and Figure 41 respectively.

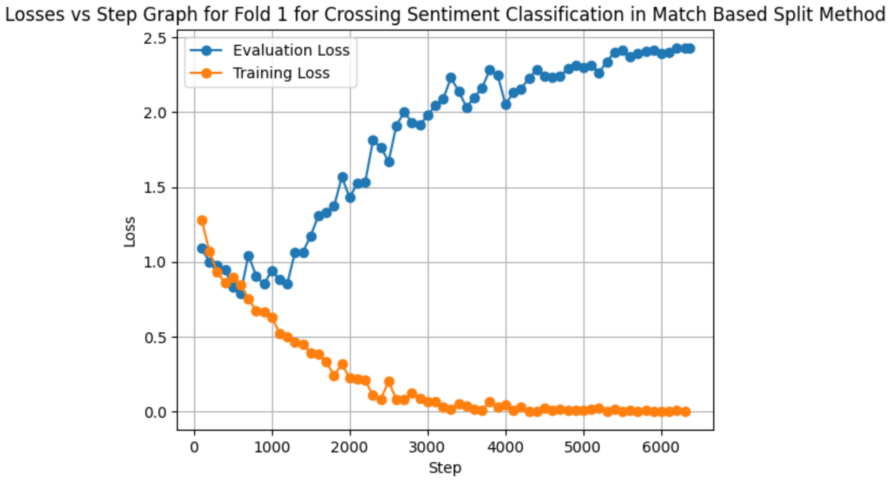


Figure 39 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Match Based Splitting Strategy

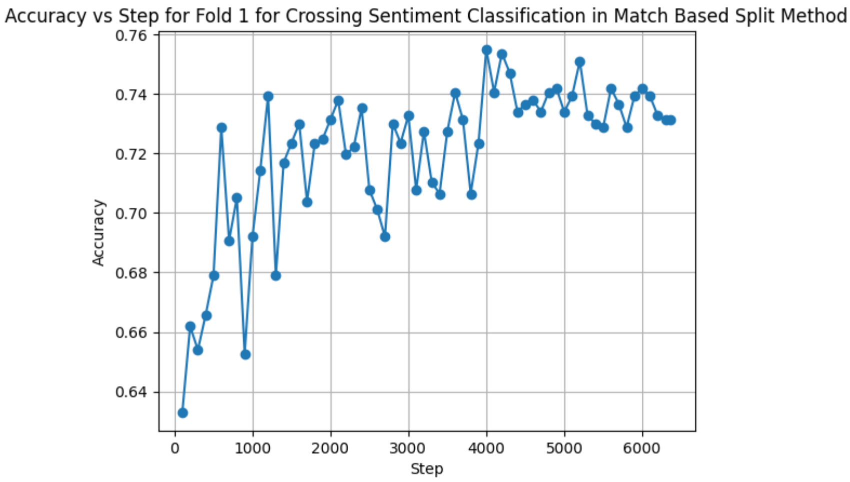


Figure 40 Accuracy vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Match Based Splitting Strategy

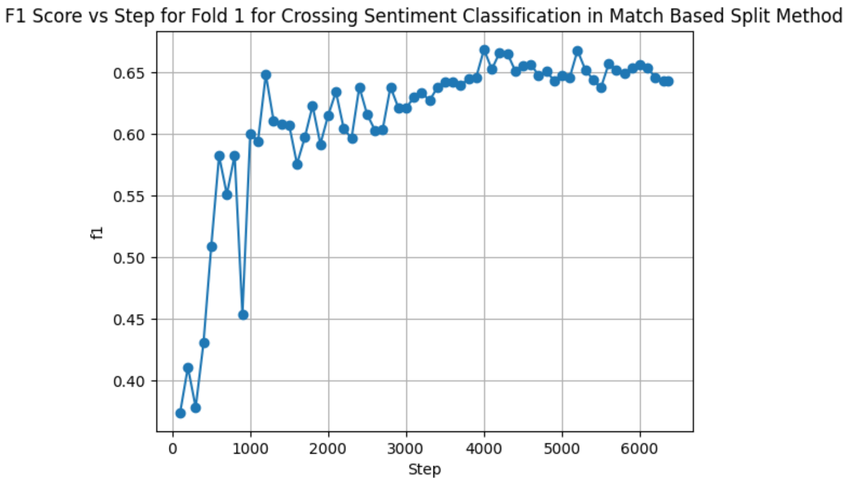


Figure 41 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Match Based Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing quality classification with Match Based Splitting Strategy is given in Figure 42. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing quality classification with Match Based Splitting Strategy is given in Figure 37 in which the rows represent the actual labels and columns represents the predicted labels.

Table 40 Confusion Matrix of test dataset for Fold 1 in Crossing Quality Classification with Match Based Splitting Strategy

	Very Bad	**Bad**	**Neutral**	**Good**	**Very Good**
Very Bad	9	15	9	0	2
Bad	6	45	46	6	0
Neutral	5	25	315	29	0
Good	0	4	45	130	24
Very Good	0	2	4	26	76

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 41. The mean values and standard deviation of these metrics are also shown in Table 41. The average accuracy for all folds is calculated as 72.0% and the average macro F1 score is calculated as 0.623.

Table 41 Performance metrics of fine-tuned model for each fold in Crossing Quality Classification with Match Based Strategy

fold	accuracy	Macro F1 score	precision	recall
1	0.699	0.594	0.624	0.576
2	0.734	0.628	0.650	0.625
3	0.714	0.613	0.632	0.600
4	0.720	0.621	0.658	0.600
5	0.735	0.657	0.671	0.651
Mean	**0.720**	**0.623**	**0.647**	**0.610**
Std Dev	**0.015**	**0.023**	**0.019**	**0.029**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in random splitting strategy for each fold is founded as "Neutral". Using this label as the as the classifier, the performance metrics are founded as in Table 42. It can be seen that, the average accuracy is increased by 23.0% and macro F1 score is increased by 0.492 using the fine-tuned model.

Table 42 Performance metrics of baseline model for each fold in Crossing Quality Classification with Match Based Splitting Strategy

fold	accuracy	Macro F1 score	precision	recall
1	0.454	0.125	0.091	0.200
2	0.521	0.137	0.104	0.200
3	0.475	0.129	0.095	0.200
4	0.509	0.135	0.102	0.200
5	0.490	0.132	0.098	0.200
Mean	**0.490**	**0.131**	**0.098**	**0.200**
Std Dev	**0.026**	**0.005**	**0.005**	**0.000**

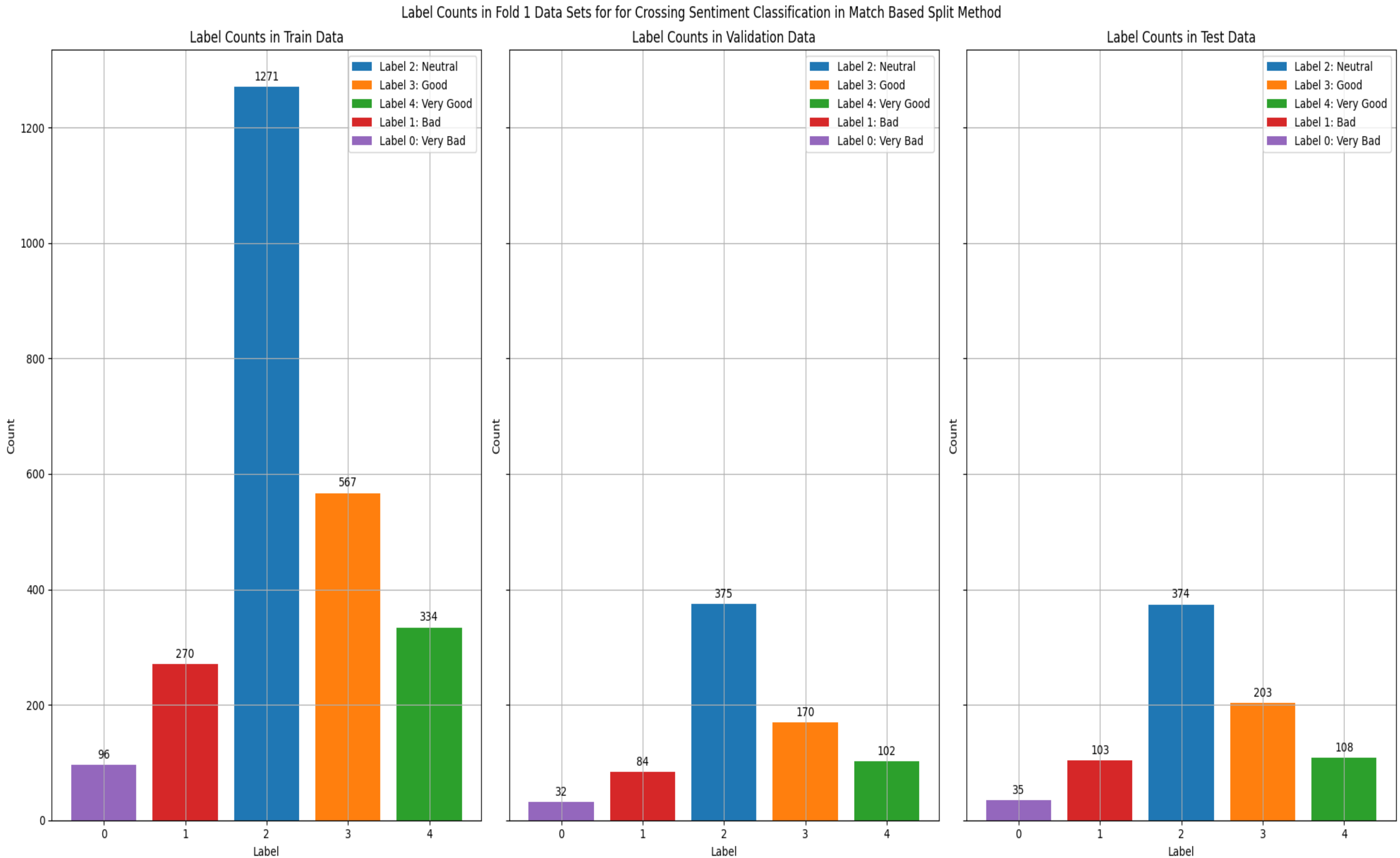


Figure 42 Label distributions of train-validation-test datasets in Fold 1 for Crossing Quality Classification with Match Based Splitting Strategy

3.3.4.3. Crossing Quality Classification with Team Based Splitting Strategy:

There are 20 different teams in the datasets and these teams are split into train-validation-test to get roughly 0.6-0.2-0.2 ratios for them. Thus, 2 teams are selected for test datasets and 2 teams are selected for validation datasets and 16 for training dataset which leads to 10 folds. The teams that are tested in each fold is given as below:

- Fold 1: ['Arsenal', 'Tottenham Hotspur']
- Fold 2: ['Aston Villa', 'Nottingham Forest']
- Fold 3: ['Chelsea', 'Fulham']
- Fold 4: ['Brentford', 'Liverpool']
- Fold 5: ['Southampton', 'West Ham United']
- Fold 6: ['Bournemouth', 'Manchester United']
- Fold 7: ['Leeds United', 'Wolverhampton Wanderers']
- Fold 8: ['Brighton', 'Manchester City']
- Fold 9: ['Everton', 'Leicester City']
- Fold 10: ['Crystal Palace', 'Newcastle United']

The loss, accuracy and macro F1 score vs step graphs in Fold 1 for crossing quality classification with team based splitting strategy are shared in Figure 43, Figure 44 and Figure 45 respectively.

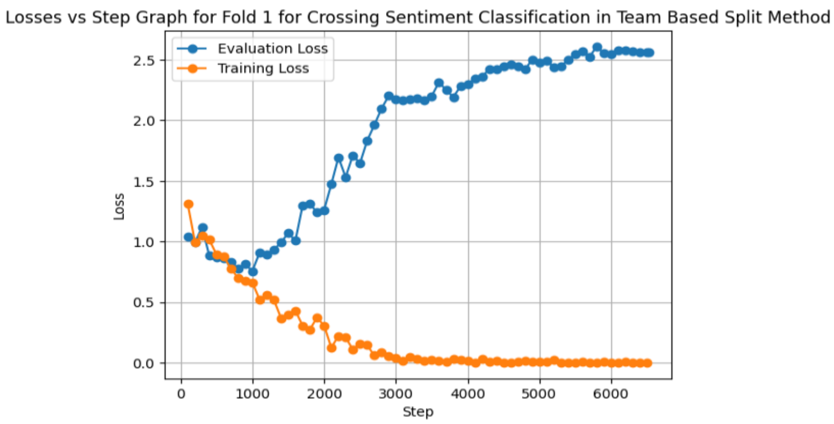


Figure 43 Loss vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Team Based Splitting Strategy

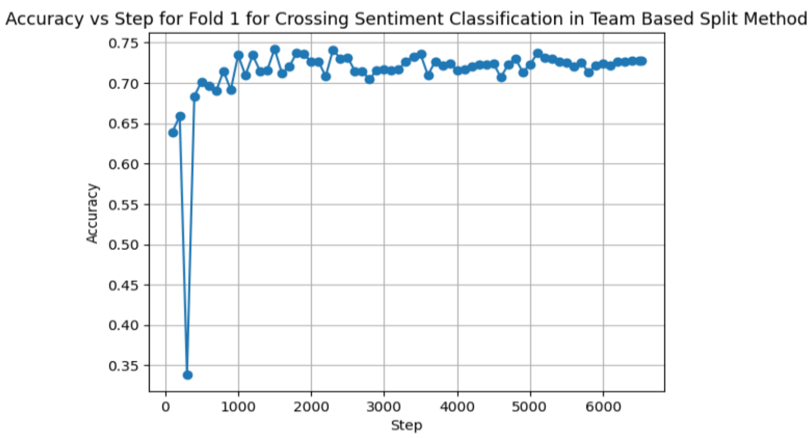


Figure 44 Accuracy vs Step Graph of the fine-tuned model for Fold 3 in Crossing Quality Classification with Team Based Splitting Strategy

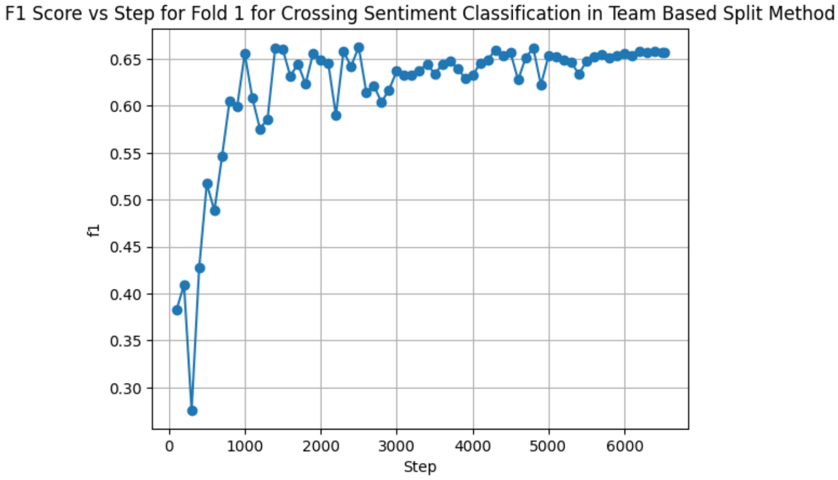


Figure 45 Macro F1 Score vs Step Graph of the fine-tuned model for Fold 1 in Crossing Quality Classification with Team Based Splitting Strategy

The label distributions for train-validation-test datasets in Fold 1 for crossing quality classification with Team Based Splitting Strategy is given in Figure 46. The labels are distributed in stratified manner for train, validation, and test sets.

The confusion matrix formed with test dataset in Fold 1 for crossing quality classification with Team Based Splitting Strategy is given in Table 43 in which the rows represent the actual labels and columns represents the predicted labels.

Table 43 Confusion Matrix of test dataset for Fold 1 in Crossing Quality Classification with Team Based Splitting Strategy

	Very Bad	**Bad**	**Neutral**	**Good**	**Very Good**
Very Bad	15	11	5	1	0
Bad	13	36	35	2	2
Neutral	6	18	306	44	6
Good	1	3	23	113	13
Very Good	0	0	4	19	74

The performance metrics of fine-tuned model calculated for test dataset including accuracy, macro F1 score, precision and recall results are for each fold are given in Table 44. The mean values and standard deviation of these metrics are also shown in Table 44. The average accuracy for all folds is calculated as 72.0% and the average macro F1 score is calculated as 0.623.

Table 44 Performance metrics of fine-tuned model for each fold in Crossing Quality Classification with Team Based Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.725	0.635	0.638	0.637
2	0.726	0.623	0.646	0.611
3	0.690	0.575	0.592	0.565
4	0.765	0.657	0.670	0.649
5	0.720	0.600	0.624	0.590
6	0.749	0.674	0.677	0.671
7	0.737	0.641	0.668	0.625
8	0.748	0.653	0.666	0.644
9	0.702	0.584	0.618	0.571
10	0.731	0.628	0.645	0.617
Mean	**0.729**	**0.627**	**0.644**	**0.618**
Std Dev	**0.022**	**0.032**	**0.027**	**0.034**

The majority class classifier is considered as the baseline model. The majority class in each training dataset considered in team based splitting strategy for each fold is founded as "Neutral". Using this label as the as the classifier, the performance metrics are founded as in Table 45. It can be seen that, the average accuracy is increased by 23.8% and macro F1 score is increased by 0.495 using the fine-tuned model.

Table 45 Performance metrics of baseline model for each fold in Crossing Quality Classification with Team Based Splitting Strategy

fold	**accuracy**	**Macro F1 score**	**precision**	**recall**
1	0.507	0.135	0.101	0.200
2	0.498	0.133	0.100	0.200
3	0.470	0.128	0.094	0.200
4	0.525	0.138	0.105	0.200
5	0.477	0.129	0.095	0.200
6	0.487	0.131	0.097	0.200
7	0.481	0.130	0.096	0.200
8	0.501	0.133	0.100	0.200
9	0.464	0.127	0.093	0.200
10	0.498	0.133	0.100	0.200
Mean	**0.491**	**0.132**	**0.098**	**0.200**
Std Dev	**0.018**	**0.003**	**0.004**	**0.000**

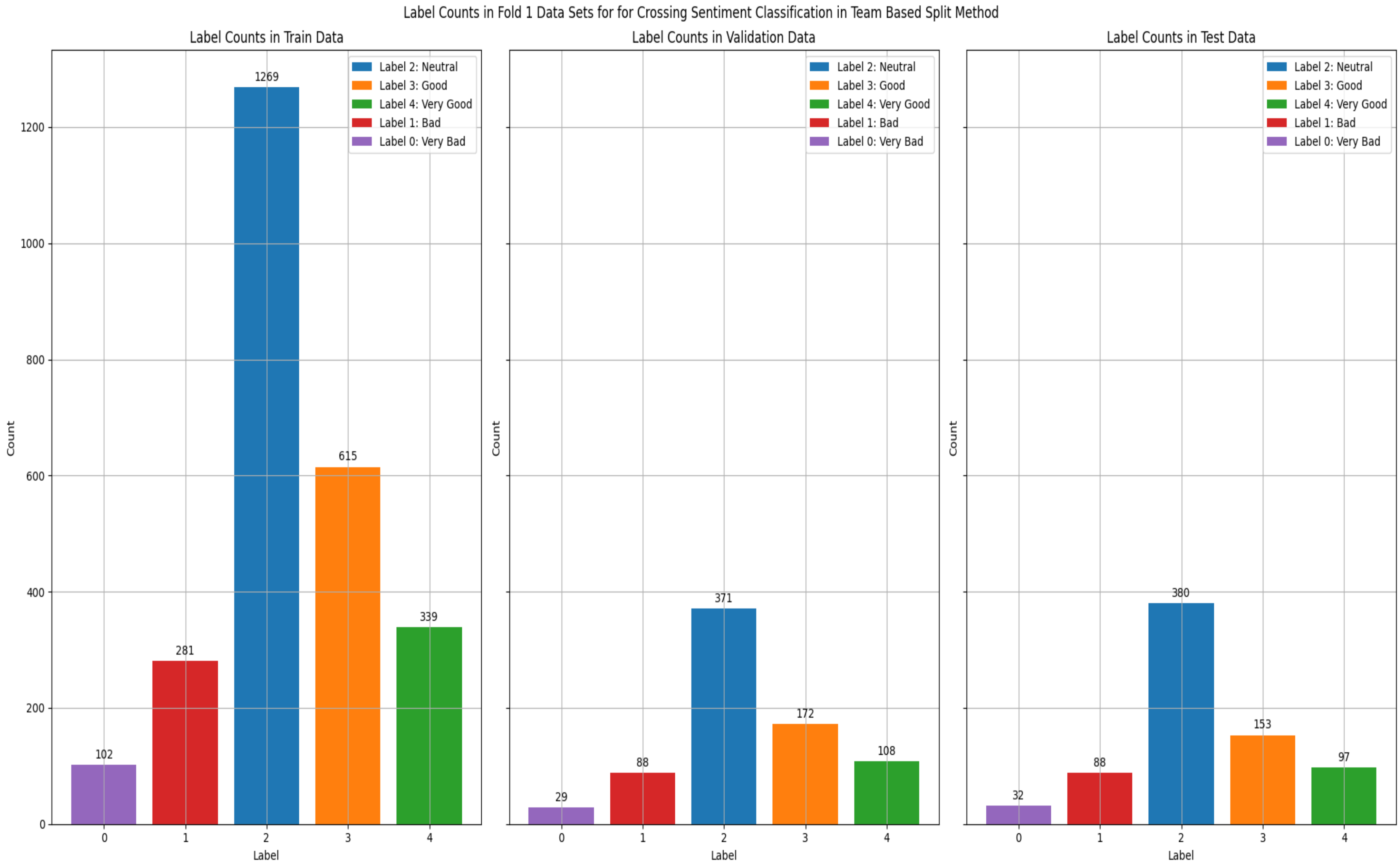


Figure 46 Label distributions of train-validation-test datasets in Fold 1 for Crossing Quality Classification with Team Based Splitting Strategy

3.3.4.4. Crossing Quality Classification Results of the Strategies:

The average metric results after cross validation applied for each strategy on test datasets are given in Table 46. The standard deviations for all metrics are low (<0.01) so it can be said that the average results of each strategy are nearly same. Overall, for all strategies the average accuracy is founded as 77.7% and macro F1 score is founded as 0.701.

Table 46 Average metric results of fine-tuned models for crossing quality classification after cross validation applied for each strategy on test datasets

Strategy	**accuracy**	**Macro F1 score**	**precision**	**recall**
Random Split	0.729	0.626	0.653	0.612
Match Split	0.720	0.623	0.647	0.610
Team Split	0.729	0.627	0.644	0.618
Mean	**0.726**	**0.625**	**0.648**	**0.613**
Std Dev	**0.005**	**0.002**	**0.005**	**0.004**

When test datasets and predictions are investigated, the followings are founded (the ratios and results are from Random Split Strategy Fold 3) :

- There are 214 mislabeled commentaries for crossing qualities. 176 of them are mislabeled by one level (i.e. “Very Bad” is mislabeled as “Bad”, “Neutral” is labeled as “Good” or “Bad” and “Very Good” is mislabeled as “Good” visa versa). In other words, only 38 of 825 records are mislabeled more than level. That indicates, the 95.4% of predicted qualities are within one level.

- On the other hand, the sentiment of 50 mislabels are correct (i.e. “Very Bad” is mislabeled as “Bad” and “Very Good” is mislabeled as “Good” visa versa). That makes 80.1% of the predictions’ sentiment is predicted correctly.

CHAPTER 4

DISCUSSION AND LIMITATIONS

In this section, the findings and implications related to the research questions defined in Section 1.1 are discussed and limitations of the present study are presented.

4.1.Discussions on the Research Questions

Research Question 1: What is the reliability of online football commentaries for gaining insights into key performance indicators, such as crossing action, in football and how can a dataset be constructed through labeling of these actions?

In this thesis, 20255 unique live commentaries sourced from goal.com covering 304 matches in the Premier League 2022-2023 Season are employed to investigate the Key Performance Indicators (KPIs) in football matches and crossing actions in particular. This raw dataset is labeled by 17 human annotators to detect the crossing actions mentioned in commentaries and define the characteristics of these crosses. After the filtering and merging processes, the dataset is ended up with 19686 rows which are fully labeled and constructed as a ground truth dataset.

- *Research Question 1.1: How can crossing actions in a football match be found and annotated using online football commentaries?*

  During the labeling phase, level of agreement on "Is There a Cross" label which indicates the existence of a crossing action for that commentary was identified as "Moderate". The main reasons of these disagreements were determined as the ambiguity of set piece crosses received from corners and free kicks and type of the crosses that were excluded systematically such as pull back and cut back crosses.

  After calculation of Inter Annotator Agreements (IAA) and determining the reasons behind the disagreements, a consensus phase is conducted to decide the final labels. As described in 3.2.4.2, a spread sheet is constructed to review the disagreements by labelers. In particular, the definition and types of crosses are introduced to labelers again before starting the consensus phase to be precise in the final labels. Each disagreed label is checked and after the consensus final label is determined. In case of the disagreements, a third labeler was invited to resolve the conflict and decide on the final level.

4371 commentaries out of 19686 are labeled as "Yes" for "Is there a cross" label after the consensus phase which indicates the existence of a crossing action for that commentary.

The task of detecting crossing actions in football match commentaries carries inherent challenges. The "moderate" level of agreement among labelers underscores the complexity associated with interpreting textual descriptions of crossing actions, especially when considering the nuances of different types of crosses. The identified ambiguities, often inherent in the commentary language, contributed to inconsistencies in labeling, necessitating a consensus phase for resolution.

Despite the challenges encountered, this labeled dataset retrieved from the online football commentary now serves as a valuable resource for further analysis and exploration of crossing actions in football matches.

- *Research Question 1.2: How can players responsible for specific actions in football matches, such as crossing, be found and annotated using online football commentaries?*

  During the labeling phase, level of agreement on "Crossing Player" label which indicates the player who performs the crossing action was identified as "Strong" as described in Section 3.2.4.1.3. IAA was measured only for the commentaries that both annotators initially labeled that a crossing action exists.

  The main reasons of the disagreements were determined as the ambiguity between cross sender and receiver and the players with same surnames were presented in the same team.

  Each disagreed label is checked and after the consensus final label is determined. Furthermore, the commentaries that labelers initially have a disagreement on an existence of a crossing action were also labeled if the final label was determined as "Yes". In conclusion, the crossing player of the 4.371 commentaries containing a crossing action were labeled in this process.

  The research demonstrated the feasibility of identifying specific players responsible for actions, such as crossing, in online football commentaries, even in the presence of disagreements during the labeling phase which is overcome with consensus. This dataset can be a valuable source to analyze the KPIs specific to players to increase their training performance, analyze the team and opponents and for scouting.

- Research Question 1.3: *How can outcomes of the specific actions in football matches, such as crossing, be found and annotated using online football commentaries?*

  During the labeling phase, level of agreement on "Outcome of Cross – Category 1", "Outcome of Cross – Category 2" and "Outcome of Cross – Category 3" labels which indicates the outcome of the crossing action was identified as “Moderate” as described in detail at Sections 3.2.4.1.4, 3.2.4.1.5 and 3.2.4.1.6 respectively. IAA was measured only for the commentaries that both annotators initially labeled that a crossing action exists.

The main reasons of the disagreements were determined as the ambiguity between the roles of the opponents as defender and goalkeeper, the shot made by a receiver and goalkeeper saves and the final position of the ball after opponent’s interceptions. The other detailed disagreements are also mentioned at Sections 3.2.4.1.4, 3.2.4.1.5 and 3.2.4.1.6.

Each disagreed label is checked and after the consensus final label is determined. Furthermore, the commentaries that labelers initially have a disagreement on an existence of a crossing action were also labeled if the final label was determined as “Yes”. In conclusion, the categorized outcome of the 4.371 commentaries containing a crossing action were labeled in this process.

The research demonstrated the feasibility of identifying the outcome for specific actions, such as crossing, in online football commentaries, even in the presence of disagreements during the labeling phase which is overcome with consensus. This dataset can be a valuable source that can inform strategic decisions for analyzing the team and opponents and determining the strategies of the team’s success over the outcome of the crosses.

- *Research Question 1.4: How can quality of the specific actions in football matches, such as crossing, be identified using online football commentaries?*

During the labeling phase, level of agreement on “Quality of Cross” label which indicates the quality and quality of the crossing action was identified as “Moderate” as described in detail at Section 3.2.4.1.7. IAA was measured only for the commentaries that both annotators initially labeled that a crossing action exists.

The main reasons of the disagreement were determined as not defining the successful crosses as “Very Good” unless it is finalized with a goal. As mentioned in Section 2.1, the successful crosses are also the ones that create scoring opportunity that receiver cannot touch or score and ones that defenders make vital challenges to prevent it from scoring besides the ones ended with a goal.

Each disagreed label is checked and after the consensus final label is determined. Furthermore, the commentaries that labelers initially have a disagreement on an existence of a crossing action were also labeled if the final label was determined as “Yes”. In conclusion, the crossing action quality of the 4371 commentaries containing a crossing action were labeled in this process.

The research demonstrated the feasibility of identifying the quality for specific actions, such as crossing, in online football commentaries, even in the presence of disagreements during the labeling phase which is overcome with consensus. This dataset can be a valuable source that can inform strategic decisions for analyzing the team and opponents and determining the strategies of the team's success over the quality of the crosses. Also, by combining with other labels such as "Crossing Player" and "Outcome of Cross", this sentiment for the crossing actions can be benefited for scouting activities and determining key skills of the players.

Research Question 2: How can labels that will be employed in Large Language Models be defined and validated to automate the extraction of key performance indicators?

First, the characteristics of the KPI that will be observed is identified. In this case, for the crossing action: the exitance of the crossing action, the player who performs this action, the outcome of the cross and quality of the cross are identified as the features to be extracted from the commentaries.

To get the ground truth dataset to train the prospective Large Language Models (LLMs) with the features that are defined, a methodology called Human Expert Labeling Process (HELP) has been applied (Aslan et al., 2017). The labels for annotating the commentaries are identified in the planning phase of this process by conducting a literature review.

The literature review was focused on "outcome of the cross" because the other features had predefined categories. For example, "Yes" and "No" labels are defined for "Is There a Cross" and quality of the crosses are defined as a general sentiment analysis with labels "Very Good", "Good", "Neutral", "Bad" and "Very Bad". Furthermore, the "Crossing Team" was chosen from the rivals in the matches and the "Crossing Player" labels were determined from the squad list of each teams 2022-2023 squads determined at the beginning of the season. However, a literature review was neccessasry to define all the possible outcomes of a cross.

In the literature review, many KPIs and parameters were considered to define the outcomes of the cross other than the cross outcome itself. Shot, pass, corner kick and freekick outcomes are all considered. In addition, the notational analysis and performance indicator researches are inestigated to get much more insight about this domain.

A lexicon is crafted based on the findings derived from after the literature review. Subsequently, these outcomes are tested on a sample dataset. Following the examination of the labels over this dataset, the neccesity to categorize the outcomes of the crosses and definition of new outcomes that were not addressed in the literature were introduced.

After these labels were defined, the annotators were selected, trained and their work was measured by the calculation of IAAs. The differences between the annotation of two labels were checked and consensus for the final label was considered and reached for them.

Lastly, some preprocessing on the commentaries that are either coming before or after the goal scoring commentaries. Such commentaries contain a detailed description of the scroing action, therefore in order not to duplicate the actions, these commentaries are merged with the ones that express the scoring action.

By using HELP methodology and the conductance of a detailed literature review, final dataset that contains 19686 labeled commentaries is constructed to be used in prospective LLMs.

After forming the final datasets, the extracted labels are compared with the real stats obtained from the offical English Pemier League website (Premier League Player Stats - Crosses, n.d.) The top 20 players that conducted the most crossing actions in 2022-23 English Pemier League Season and number of occurrence of these labels in the commentaries are given. It can be seen that, these 20 players also belong to the top 28 players that conducted the most of the crosses in the final dataset. Furthermore, in the table the ratio of the crosses that were mentioned in the dataset is also given. On average, 27% of the total crosses are detecet in the commentary. However, the commentary dataset contains only 304 matches, these percentage could increase when all matches' commentaries are considered. In addition, the commentaries describes the key moments in the match that has direct effect to change the score. Therefore, some crossing actions that do not create any chance may not been mentioned in the commentaries.

Table 47 Top 20 Players That Performed the Most Crossing Actions in 2022-23 English Premier League Season

Player Name	**Position in Real Stats**	**Number of Crosses in Real Stats**	**Position in Commentary**	**Number of Crosses in Commentary**	**Crosses Detected in Commentary**
Kieran Trippier	1	393	1	92	23%
Trent Alexander-Arnold	2	252	4	60	24%
Michael Olise	3	242	2	72	30%
James Ward-Prowse	4	241	12	51	21%
Kevin De Bruyne	5	226	3	64	28%
Pascal Groß	6	224	10	52	23%
Andreas Pereira	7	196	11	51	26%
Ivan Perisic	8	196	9	53	27%
Andy Robertson	9	195	7	53	27%
Dwight McNeil	10	194	14	49	25%

Table 47 (cont.)

Player Name	Position in Real Stats	Number of Crosses in Real Stats	Position in Commentary	Number of Crosses in Commentary	Crosses Detected in Commentary
Jack Harrison	11	186	13	50	27%
Demarai Gray	12	173	17	47	27%
Bukayo Saka	13	172	8	53	31%
Morgan Gibbs-White	14	170	6	56	33%
Jarrod Bowen	15	159	19	41	26%
Mathias Jensen	16	158	18	47	30%
Bryan Mbeumo	17	156	5	57	37%
Solly March	18	155	17	47	30%
Bruno Fernandes	19	141	25	31	22%
Antonee Robinson	20	134	28	29	22%

Research Question 3: How effectively can Large Language Models be used to extract key performance indicators in football, such as crossing action, from online football commentaries?

The average results after the application of cross-validation on fine-tuned LLMs for each task is given in Table 48. As discussed in subsections of Section 3.3, the fine-tuned models have conducted a better performance than the baseline models in all tasks with all strategies.

Table 48 Average Metric Results of the Fine-Tuned LLM Classifying Crossing Features

	accuracy	F1 score	precision	recall
Crossing Check	0.961	0.943	0.944	0.943
Crossing Player[1]	0.834	0.668	0.671	0.698
Crossing Outcome	0.777	0.701	0.722	0.691
Crossing Quality	0.726	0.625	0.648	0.613

[1] The average of random and match based splitting strategies are considered.

In sections 3.3.1.4. Crossing Check Classification Results of the Strategies: 3.3.2.4. Crossing Player Classification Results of the Strategies:, 3.3.3.4. Crossing Outcome Classification Results of the Strategies: and 3.3.3.4. Crossing Outcome Classification Results of the Strategies: the cause of the mislabeling done by the model are explained. The main reasons, the models made mistakes for the classification tasks are founded as: misjudgment of the set pieces, not understanding all types of crosses, lack of information about players inside the commentary (their position, team etc.) and ground truth mistakes.

As described in detail in section 3.3.4.4. Crossing Quality Classification Results of the Strategies:, the fine-tuned models (by Random Split Strategy Fold 3 Result) can predict the qualities within a one level of quality with 95.4% and sentiment of the commentary with 80.1%.

With these results, it can be conducted that from the online football commentaries, many KPIs can be extracted and these KPIs can be used for various applications such as scouting, tactical analysis and opponent analysis.

4.2.Limitations

4.2.1. Dataset Limitations

The dataset in this thesis is constructed using the commentaries from 304 2022-2023 English Premier League matches. The sample size for the training of the LLMs seems sufficient with 19686 labeled commentaries. Due to the popularity of the English Premier League and the fanbase that has a higher knowledge about the language, the commentaries made for these matches were high in quality. However, the commentaries for other leagues, which have lower popularity and followers who do not require commentary in English may be shorter and less detailed. This situation may cause the detection of the KPIs harder by using commentaries for such matches and leagues.

4.2.2. Data Quality

The potential quality of dataset is affected by two aspects: the quality of the raw data and the quality of the labels. The quality of the raw data is determined by the commentators' proficiency in articulating match events in textual form. Therefore, the speed of annotating the events, the accurate identification of involved players in events and the comprehensiveness of details in the event descriptions significantly impact the overall quality of the commentaries and, consequently, the raw dataset.

The labels for the characteristic of the crossing actions are determined by the consensus of two labelers for each commentary to get a larger dataset with a practical approach. The consensus that these two raters are more sensitive to individual biases from either rater. Furthermore, the commentaries may not benefit from diverse perspectives and understanding of them may have been lower. Lastly, third raters were introduced in each case that a consensus could not be reached and that caused a cost in time.

CHAPTER 5

5. CONCLUSION AND FUTURE WORK

5.1. Conclusion

The live online commentaries from the 304 matches of the English Premier League 2022-2023 Season are used to extract features of the crosses using fine-tuned Large Language Model (LLM) Model in this research. To do so, these commentaries are labeled by 17 different annotators to determine the existence of a cross and its properties: crossing player, outcome of the cross and quality of the cross. The labeled dataset containing the commentaries is used as the ground truth to fine tune the LLMs and observe their performances on this task.

Initially, raw data is gathered from the live commentaries sourced from goal.com focusing on matches from the Premier League 2022-2023 Season. These commentaries are then labeled by 17 labelers by following Human Expert Labeling Process (HELP). The cross characteristics that labelers will consider are defined under Table 6. To define the labels that will be used in the outcomes of crosses, a systematic literature review is applied. 5 main categories with 23 subcategories of cross outcomes are determined that are given in Table 7. After the labeling phase is completed by the labelers, the final labels were determined with consensus of the labeler and author.

The final dataset that is used as the ground truth for the fine tuning of the large language model called SportsBERT. After that, four crossing features are determined to be analyzed which are crossing check, crossing player, crossing outcome, and crossing quality. The cross-validation technique is used to measure the performance for the fine-tuned models that are aimed to label for these tasks.

After the models are fine-tuned, the performances of the fine-tuned models are calculated as in Table 48. The models' performances are all higher than the baseline majority class models that are described in sections 3.3.1.4. Crossing Check Classification Results of the Strategies:, 3.3.2.4. Crossing Player Classification Results of the Strategies:, 3.3.3.4. Crossing Outcome Classification Results of the Strategies:, and 3.3.4.4. Crossing Quality Classification Results of the Strategies:

5.2. Implications

In this study, a novel football commentary dataset tailored for training state-of-the-art natural language processing (NLP) models. This dataset not only serves as a valuable resource for analyzing future matches but also enables detailed assessments of football player performance. Moreover, comprehensive crossing outcome categories are formed that can be used for future studies of crossing actions in football.

Additionally, this research has yielded the development of four fine-tuned models from SportsBERT model that aim to classify various aspects of crossing plays: the presence of crossing actions, the players that performs the crossing actions, the outcomes of crosses, and the qualitative assessment of crossing execution. These models not only showcase the potential of machine learning in analyzing football gameplay but also provide practical tools for coaches, analysts, and enthusiasts.

Lastly, the findings in thesis underscore the effectiveness of LLMs in extracting key performance indicators from football data. By leveraging the power of LLMs, the capacity to unveil insights into player strategies, team tactics, and overall match dynamics are demonstrated.

5.3. Future Work

This thesis is providing valuable insights into usage of LLM models in football domain and it also reveals new directions for the future research. The following are potential areas for future work:

5.3.1. Other Key Performance Indicators in Football Domain:

In this study, the reliability of online football commentaries for gaining insights into crossing actions in football is considered. However, it is important to note that there are numerous other key performance indicators in football, including shooting, passing, and defending, which can also be observed from football commentaries and extracted along with their respective features.

5.3.2. More labelers

In this study, as mentioned earlier, the labels for the characteristic of the crossing actions are determined by the consensus of two labelers for each commentary to get a larger dataset with a practical approach. The consensus that these two raters are more sensitive to individual biases from either rater. For future work, the number of labelers can be increased to compare the quality of the datasets, or they can be used to form a larger dataset.

5.3.3. Better models

The performance of LLMs is advancing exponentially with each passing day, driven by the introduction of new model architectures and the continual increase in model parameters. The SportsBERT model, fine-tuned in this thesis, is a BERT model that was introduced in 2018, trained from scratch. For future work, consideration can be given to exploring brand new commercial and open-source Large Language Models for the analysis of crossing actions.

5.3.4. Changing the NLP Task

In this study, the prediction of the player who performs the crossing action is conducted by Sequence Classification Task. Each player that has a license in 2022-2023 English Premier League Season has been considered as a unique class. However, this could have been performed instead by conducting Named Entity Recognition (NER) task. To conduct a NER task, the dataset entities, in this case crossing player's position in each commentary, must be initially labeled.

REFERENCES

Ahuja, R., Chug, A., Gupta, S., Ahuja, P., & Kohli, S. (2019). Classification and Clustering Algorithms of Machine Learning with their Applications (pp. 225–248). https://doi.org/10.1007/978-3-030-28553-1_11

Alamar, B. C. (2013). Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers. Columbia University Press. https://doi.org/10.7312/alam16292

Alcock, A. (2010). Analysis of direct free kicks in the women's football World Cup 2007. European Journal of Sport Science, 10(4), 279–284. https://doi.org/10.1080/17461390903515188

Almeida, C., Ferreira, A., & Volossovitch, A. (2014). Effects of Match Location, Match Status and Quality of Opposition on Regaining Possession in UEFA Champions League. Journal of Human Kinetics, 41, 203–214. https://doi.org/10.2478/hukin-2014-0048

Almenara, S. (2021, March 18). Key Performance Indicators (KPIs) for Football Analysis. Retrieved January 4, 2024, from https://www.nacsport.com/blog/en-gb/Tips/key-performance-indicators-football

Arastey, G. M. (2018, June 19). Notational analysis: A synonym of today's performance analysis. Sport Performance Analysis. Retrieved January 3, 2024, from https://www.sportperformanceanalysis.com/article/2018/6/18/notational-analysis-a-synonym-to-performance-analysis

Artstein, R. (2017). Inter-annotator Agreement (N. Ide & J. Pustejovsky, Eds.; pp. 297–313). Springer Netherlands. https://doi.org/10.1007/978-94-024-0881-2_11

Aslan, S., Mete, S., Okur, E., Oktay, E., Alyuz, N., Genc, U., Stanhill, D., & Esme, A. (2017). Human Expert Labeling Process (HELP): Towards a Reliable Higher-Order User State Labeling Process and Tool to Assess Student Engagement. Journal of Educational Technology, 57(1), 53–59.

Bastian, M. (2023). OpenAI's GPT-4 is a safer and more useful ChatGPT that understands images. THE DECODER. Retrieved January 5, 2024, from https://the-decoder.com/open-ai-gpt-4-announcement/

Beare, H., & Stone, J. A. (2019). Analysis of attacking corner kick strategies in the FA women's super league 2017/2018. International Journal of Performance Analysis in Sport, 19(6), 893–903. https://doi.org/10.1080/24748668.2019.1677329

Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The Long-Document Transformer (arXiv:2004.05150). arXiv. https://doi.org/10.48550/arXiv.2004.05150

Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications*, *34*(2), 825–832. https://doi.org/10.1016/j.eswa.2006.10.022

Bera, A., Joshi, S., & Nair, Akhil. M. (2023). Football Player Substitution Analysis using NLP and Survival Analysis. *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 753–757. https://doi.org/10.1109/ICSSIT55814.2023.10061118

Braylan, A., Alonso, O., & Lease, M. (2022). Measuring Annotator Agreement Generally across Complex Structured, Multi-object, and Free-text Annotation Tasks. Proceedings of the ACM Web Conference 2022, 1720–1730. https://doi.org/10.1145/3485447.3512242

Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*, *41*(3), 687–699.

Ott, M., Edunov, S., Grangier, D., & Auli, M. (2018). Scaling Neural Machine Translation (arXiv:1806.00187). arXiv. https://doi.org/10.48550/arXiv.1806.00187

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners (arXiv:2005.14165). arXiv. http://arxiv.org/abs/2005.14165

Castelão, D., Garganta, J., Santos, R., & Teoldo, I. (2014). Comparison of tactical behaviour and performance of youth soccer players in 3v3 and 5v5 small-sided games. International Journal of Performance Analysis in Sport, 14(3), 801–813. https://doi.org/10.1080/24748668.2014.11868759

Carmichael, F., Thomas, D., & Ward, R. (2000). Team Performance: The Case of English Premiership Football. Managerial and Decision Economics, 21(1), 31–45.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Cordón-Carmona, A., García-Aliaga, A., Marquina, M., Calvo, J. L., Mon-López, D., & Refoyo Roman, I. (2020). What Is the Relevance in the Passing Action between the Passer and the Receiver in Soccer? Study of Elite Soccer in La Liga. International Journal of Environmental Research and Public Health, 17(24), Article 24. https://doi.org/10.3390/ijerph17249396

Dai, W., Yoshigoe, K., & Parsley, W. (2018). Improving Data Quality through Deep Learning and Statistical Models (Vol. 558, pp. 515–522). https://doi.org/10.1007/978-3-319-54978-1_66

De Baranda, P. S., & Lopez-Riquelme, D. (2012). Analysis of corner kicks in relation to match status in the 2006 World Cup. European Journal of Sport Science, 12(2), 121–129. https://doi.org/10.1080/17461391.2010.551418

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals: Valuing Player Actions in Soccer. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1851–1861. https://doi.org/10.1145/3292500.3330758

Delgado-Bordonau, J. L., Domenech-Monforte, C., Francisco Guzman, J., & Mendez-Villanueva, A. (2013). Offensive and defensive team performance: Relation to successful and unsuccessful participation in the 2010 Soccer World Cup. Journal of Human Sport And Exercise, 8(4), 894–904. https://doi.org/10.4100/jhse.2013.84.02

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

FIFA (2021) Cross type. Retrieved January 4, 2024, from https://www.fifatrainingcentre.com/en/resources-tools/football-language/in-possession/distributing-the-ball/cross/cross-type/index.php

FIFA (2021) The football landscape – The Vision 2020-2023. FIFA Publications. Retrieved January 4, 2024, from https://publications.fifa.com/en/vision-report-2021/the-football-landscape/

FIFA (2023). Use of wide areas to create goal-scoring opportunities. FIFA Training Centre. Retrieved January 2, 2024, from www.fifatrainingcentre.com/en/fwc2022/technical-and-tactical-analysis/use-of-wide-areas.php

GenAI, META, Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., … Scialom, T. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models (arXiv:2307.09288). arXiv. http://arxiv.org/abs/2307.09288

Ghar, S., Patil, S., & Arunachalam, V. (2021). Data Driven football scouting assistance with simulated player performance extrapolation. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 1160–1167. https://doi.org/10.1109/ICMLA52953.2021.00189S

Gómez, M.-Á., Mitrotasios, M., Armatas, V., & Lago-Peñas, C. (2018). Analysis of playing styles according to team quality and match location in Greek professional soccer. International Journal of Performance Analysis in Sport, 18(6), 986–997. https://doi.org/10.1080/24748668.2018.1539382

GPT-3 powers the next generation of apps. (2021). Retrieved January 5, 2024, from https://openai.com/blog/gpt-3-apps

Hayes, A., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures, 1, 77–89. https://doi.org/10.1080/19312450709336664

Herold, M., Kempe, M., Bauer, P., & Meyer, T. (2021). Attacking Key Performance Indicators in Soccer: Current Practice and Perceptions from the Elite to Youth Academy Level. Journal of Sports Science & Medicine, 20(1), 158–169. https://doi.org/10.52082/jssm.2021.158

Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., & Wang, H. (2023). Large Language Models for Software Engineering: A Systematic Literature Review (arXiv:2308.10620). arXiv. https://doi.org/10.48550/arXiv.2308.10620

Hughes, M., & Franks, I. M. (2004). Notational analysis of sport: Systems for better coaching and performance in Sport. Routledge & CRC Press.

Hughes, M., & Lovell, T. (2019). Transition to attack in elite soccer. Journal of Human Sport and Exercise, 14. https://doi.org/10.14198/jhse.2019.141.20

Hughes, M., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer—An analysis of the key performance indicators of elite male soccer players by position. Journal of Human Sport and Exercise, 7. https://doi.org/10.4100/jhse.2012.72.06

International Centre for Sports Studies (CIES) (2018). The Socio-Economic Impact of Football. Retrieved January 4, 2024, from https://www.wto.org/english/res_e/webcas_e/podcast_s5e3_note1_e.pdf

James, N. (2006). Notational analysis in soccer: Past, present and future. International Journal of Performance Analysis in Sport, 6(2), 67–81. https://doi.org/10.1080/24748668.2006.11868373

Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (1st ed.). Prentice Hall PTR.

Kim, J., James, N., Parmar, N., Ali, B., & Vučković, G. (2019). Determining unstable game states to aid the identification of perturbations in football. International Journal of Performance Analysis in Sport, 19(3), 302–312. https://doi.org/10.1080/24748668.2019.1602439

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Kubayi, A., & Larkin, P. (2019). Analysis of teams' corner kicks defensive strategies at the FIFA World Cup 2018. International Journal of Performance Analysis in Sport, 19(5), 809–819. https://doi.org/10.1080/24748668.2019.1660547

Kvesić, I., Jozak, R., & Bašić, D. (2017). Analysis of Ball Clearance in the First Croatian Football League. https://doi.org/10.13140/RG.2.2.15363.96805

Lee, J., & Mills, S. (2021). Analysis of corner kicks at the FIFA Women's World Cup 2019 in relation to match status and team quality. International Journal of Performance Analysis in Sport, 21, 1–21. https://doi.org/10.1080/24748668.2021.1936408

Lee, M., Kim, D. Y., Williams, A., & Pedersen, P. (2016). Investigating the Role of Sports Commentary: An Analysis of Media-Consumption Behavior and Programmatic Quality and Satisfaction. Journal of Sports Media, 11, 145–167. https://doi.org/10.1353/jsm.2016.0001

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension (arXiv:1910.13461). arXiv. https://doi.org/10.48550/arXiv.1910.13461

Link, D. (2018). Data Analytics in Professional Soccer. Springer Fachmedien. https://doi.org/10.1007/978-3-658-21177-6

Link, D., Kolbinger, O., Weber, H., & Stöckl, M. (2016). A topography of free kicks in soccer. Journal of Sports Sciences, 34(24), 2312–2320. https://doi.org/10.1080/02640414.2016.1232487

Liu, H., Gomez, M. A., & Lago-Penas, C. (2015). Match Performance Profiles of Goalkeepers of Elite Football Teams. International Journal of Sports Science & Coaching, 10(4), 669–682. https://doi.org/10.1260/1747-9541.10.4.669

Liu, H., Hopkins, W. G., & Gómez, M.-A. (2016). Modelling relationships between match events and match outcome in elite football. European Journal of Sport Science, 16(5), 516–525. https://doi.org/10.1080/17461391.2015.1042527

Liu, H., Hopkins, W., Gómez, A. M., & Molinuevo, S. J. (2013). Inter-operator reliability of live football match statistics from OPTA Sportsdata. International Journal of Performance Analysis in Sport, 13(3), 803–821. https://doi.org/10.1080/24748668.2013.11868690

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. Meta-Radiology, 1(2), 100017. https://doi.org/10.1016/j.metrad.2023.100017

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Lorains, M., Ball, K., & MacMahon, C. (2013). Performance analysis for decision making in team sports. International Journal of Performance Analysis in Sport, 13(1), 110–119. https://doi.org/10.1080/24748668.2013.11868635

Mahalle, P. N., Shinde, G. R., Ingle, Y. S., & Wasatkar, N. N. (2023). Data Centric Artificial Intelligence: A Beginner's Guide. Springer Nature. https://doi.org/10.1007/978-981-99-6353-9

Mara, J. K., Wheeler, K. W., & Lyons, K. (2012). Attacking Strategies That Lead to Goal Scoring Opportunities in High Level Women's Football. International Journal of Sports Science & Coaching, 7(3), 565-577. https://doi.org/10.1260/1747-9541.7.3.565

McHugh, M. (2012). Interrater reliability: The kappa statistic. Biochemia Medica : Časopis Hrvatskoga Društva Medicinskih Biokemičara / HDMB, 22, 276–282. https://doi.org/10.11613/BM.2012.031

Meta AI, Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models (arXiv:2302.13971). arXiv. http://arxiv.org/abs/2302.13971

Mitrotasios, M., Casal, C., Armatas, V., Losada, J., & Dios, R. (2021). Analysis of Corner Kick Success in Laliga Santander 2019/2020. 47, 8–22. https://doi.org/10.21134/eurjhm.2021.47.2

Mitrotasios, M., Kubayi, A., Armatas, V., & Larkin, P. (2021). Analysis of Crossing Opportunities at the 2018 FIFA World Cup. Montenegrin Journal of Sports Science and Medicine, 11. https://doi.org/10.26773/mjssm.220305

Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. International Journal of Data Science and Analytics, 5(4), 213–222. https://doi.org/10.1007/s41060-017-0093-7

Nagarhalli, T., Vaze, V., & Rana, N. (2021). Impact of Machine Learning in Natural Language Processing: A Review (p. 1534). https://doi.org/10.1109/ICICV50876.2021.9388380

Nguyen, Q.-M., Cao, T.-D., & Nguyen, T.-T. (2014). Automatic creation of semantic data about football transfer in sport news. *Proceedings of the 16th International Conference on Information Integration and Web-Based Applications & Services*, 356–364. https://doi.org/10.1145/2684200.2684289

O'Donoghue, P., Papadimitriou, K., Gourgoulis, V., & Haralambis, K. (2012). Statistical methods in performance analysis: An example from international soccer. International Journal of Performance Analysis in Sport, 12(1), 144–155. https://doi.org/10.1080/24748668.2012.11868590

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Zoph, B. (2023). GPT-4 Technical Report. arXiv.Org. https://arxiv.org/abs/2303.08774v4

OpenAI, Introducing ChatGPT. (2022). Retrieved January 5, 2024, from https://openai.com/blog/chatgpt

Opta Event Definitions. (n.d.). Stats Perform. Retrieved January 3, 2024, from https://www.statsperform.com/opta-event-definitions/

Partridge, D. and Franks, I. M. (1989) A detailed analysis of crossing opportunities from the 1986 World Cup (Part I), Soccer Journal, 34(2), 47–50.

Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., Pannier, B., Almazrouei, E., & Launay, J. (2023). The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only (arXiv:2306.01116). arXiv. http://arxiv.org/abs/2306.01116

Peterson, S. W., & Bruton, M. R. (2020). A review of the interaction between the striker and the goalkeeper at the individual tactical level in football. INTERNATIONAL JOURNAL OF SPORTS SCIENCE & COACHING, 15(3), 452–464. https://doi.org/10.1177/1747954120915193

Piłka, T., Grzelak, B., Sadurska, A., Górecki, T., & Dyczkowski, K. (2023). Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors. Sensors, 23(3), Article 3. https://doi.org/10.330/s23031227

Pollard, R., & Reep, C. (1997). Measuring the Effectiveness of Playing Strategies at Soccer. Journal of the Royal Statistical Society. Series D (The Statistician), 46(4), 541–550.

Premier League Player Stats—Crosses. (n.d.). Retrieved April 3, 2024, from https://www.premierleague.com/stats/top/players/total_cross

Prithvishankar Srinivasan. (n.d.). *Microsoft/SportsBERT · Hugging Face*. Retrieved April 1, 2024, from https://huggingface.co/microsoft/SportsBERT

Pulling, C. (2015). Long corner kicks in the English premier league: Deliveries into the goal area and critical area. Kinesiology, 47(2), 193–201.

Pulling, C., & Newton, J. (2017). Defending corner kicks in the English Premier League: Near-post guard systems. International Journal of Performance Analysis in Sport, 17(3), 283–292. https://doi.org/10.1080/24748668.2017.1331577

Pulling, C., Eldridge, D., Ringshall, E., & Robins, M. T. (2018). Analysis of crossing at the 2014 FIFA World Cup. International Journal of Performance Analysis in Sport, 18(4), 657–677. https://doi.org/10.1080/24748668.2018.1509255

Pulling, C., Robins, M., & Rixon, T. (2013). Defending Corner Kicks: Analysis from the English Premier League. International Journal of Performance Analysis in Sport, 13. https://doi.org/10.1080/24748668.2013.11868637

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2023). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (arXiv:1910.10683). arXiv. https://doi.org/10.48550/arXiv.1910.10683

Reilly, C. C., A. Mark Williams, Thomas. (2005). Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance. Routledge. https://doi.org/10.4324/9780203448625

Ruiz-Vanoye, J., Díaz-Parra, O., Fuentes-Penna, A., Díaz, D., García Munguía, M., Ruiz-Díaz, J., & Ruiz-Díaz, F. (2017). Motivation Index to Improve the Soccer Performance. International Journal of Combinatorial Optimization Problems and Informatics, 8(3), 45–57. Retrieved from https://ijcopi.org/ojs/article/view/18

Sabater, C., Peña-Martínez, G., Turney, E., & Montero-Fleta, B. (2008). A Spoken Genre Gets Written: Online Football Commentaries in English, French and Spanish. Written Communication, 25. https://doi.org/10.1177/0741088307313174

Sainz de Baranda, P., Adán, L., García-Angulo, A., Gómez-López, M., Nikolic, B., & Ortega-Toro, E. (2019). Differences in the Offensive and Defensive Actions of the Goalkeepers at Women's FIFA World Cup 2011. Frontiers in Psychology, 10. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00223

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter (arXiv:1910.01108). arXiv. https://doi.org/10.48550/arXiv.1910.01108

Sarkar, S. (2018). Paradox of crosses in association football (soccer)—A game-theoretic explanation. Journal of Quantitative Analysis in Sports, 14, 25–36. https://doi.org/10.1515/jqas-2017-0073

Sarmento, H., Figueiredo, A., Lago-Peñas, C., Milanovic, Z., Barbosa, A., Tadeu, P., & Bradley, P. S. (2018). Influence of Tactical and Situational Variables on Offensive Sequences During Elite Football Matches. The Journal of Strength & Conditioning Research, 32(8), 2331. https://doi.org/10.1519/JSC.0000000000002147

Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J. W., Kreps, S., McCain, M., Newhouse, A., Blazakis, J., McGuffie, K., & Wang, J. (2019). Release Strategies and the Social Impacts of Language Models (arXiv:1908.09203). arXiv. https://doi.org/10.48550/arXiv.1908.09203

Stats Perform (2017). The art of crossing. Stats Perform. Retrieved January 2, 2024, from https://www.statsperform.com/resource/the-art-of-crossing/

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., Kosub, S., Schreck, T., Keim, D. A., & Grossniklaus, M. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. Data, 2(1), Article 1. https://doi.org/10.3390/data2010002

Strafford, B. W., Smith, A., North, J. S., & Stone, J. A. (2019). Comparative analysis of the top six and bottom six teams' corner kick strategies in the 2015/2016 English Premier League. International Journal of Performance Analysis in Sport, 19(6), 904–918. https://doi.org/10.1080/24748668.2019.1677379

Tang, K., Bao, Y., Zhao, Z., Zhu, L., Lin, Y., & Peng, Y. (2018). AutoHighlight: Automatic Highlights Detection and Segmentation in Soccer Matches. *2018 IEEE International Conference on Big Data (Big Data)*, 4619–4624. https://doi.org/10.1109/BigData.2018.8621906

Taniguchi, Y., Feng, Y., Takamura, H., & Okumura, M. (2019). Generating Live Soccer-Match Commentary from Play Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), Article 01. https://doi.org/10.1609/aaai.v33i01.33017096

Tecimer, K., Tüzün, E., Dibeklioğlu, H., & Erdogmus, H. (2021). Detection and Elimination of Systematic Labeling Bias in Code Reviewer Recommendation Systems. https://doi.org/10.1145/3463274.3463336

Tenga, A., Holme, I., Ronglan, L., & Bahr, R. (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. Journal of Sports Sciences, 28, 245–255. https://doi.org/10.1080/02640410903502766

Tienza-Valverde, A., Hernández-Beltrán, V., Espada, M. C., Bravo-Sánchez, A., Santos, F. J., & Gamonales, J. M. (2023). Analysis of individual performance indicators of football goalkeeper. Apunts Sports Medicine, 58(219). https://doi.org/10.1016/j.apunsm.2023.10042

Tijssen, L. (2018). Analyzing Offensive Player- and Team Performance in Soccer Using Position Data.

Trouvain, J. (2011). Between Excitement and Triumph—Live Football Commentaries in Radio vs. TV. International Congress of Phonetic Sciences. https://www.semanticscholar.org/paper/Between-Excitement-and-Triumph-Live-Football-in-vs.-Trouvain/9e41329068ca039a3fd19c9319dc9131d863fa14

Turner, B. J., & Sayers, M. G. L. (2010). The influence of transition speed on event outcomes in a high performance football team. International Journal of Performance Analysis in Sport, 10(3), 207–220. https://doi.org/10.1080/24748668.2010.11868516

van Maarseveen, M. J., Oudejans, R. R., & Savelsbergh, G. J. (2017). System for notational analysis in small-sided soccer games. International Journal of Sports Science & Coaching, 12(2), 194–206. https://doi.org/10.1177/1747954117694922

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems, 6000–6010.

Vecer, J. (2014). Crossing in Soccer has a Strong Negative Impact on Scoring: Evidence from the English Premier League the German Bundesliga and the World Cup 2014 (SSRN Scholarly Paper 2225728). https://doi.org/10.2139/ssrn.2225728

Warrens, M. (2011). Cohen's Linearly Weighted Kappa is a Weighted Average of 2×2 Kappas. *Psychometrika*, *76*, 471–486. https://doi.org/10.1007/s11336-011-9210-z

Worville, T. (2021). We need to talk about crossing. The Athletic. Retrieved January 2, 2024, from https://theathletic.com/2521493/2021/04/17/we-need-to-talk-about-crossing/

Wright, C., Polman, R., Jones, B., & Sargeson, L. (2011). Factors Associated with Goals and Goal Scoring Opportunities in Professional Soccer. International Journal of Performance Analysis in Sport, 11, 438–449. https://doi.org/10.1080/24748668.2011.11868563

Wu, L., Danielson, A., Hu, X., & Swartz, T. (2020). A contextual analysis of crossing the ball in soccer. Journal of Quantitative Analysis in Sports, 1. https://doi.org/10.1515/jqas-2020-0060

Yamada, H., & Hayashi, Y. (2015). Characteristics of Goal-scoring Crosses in International Soccer Tournaments. Football Science, 12, 24–32. https://doi.org/10.57547/jssfenfs.12.1_24

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). XLNet: Generalized Autoregressive Pretraining for Language Understanding (arXiv:1906.08237). arXiv. https://doi.org/10.48550/arXiv.1906.08237

Yi, Q., Gómez-Ruano, M.-Á., Liu, H., Zhang, S., Gao, B., Wunderlich, F., & Memmert, D. (2020). Evaluation of the Technical Performance of Football Players in the UEFA Champions League. International Journal of Environmental Research and Public Health, 17(2), Article 2. https://doi.org/10.3390/ijerph17020604

Zhang, Y., Hofleitner, A., Furnas, H., Chung, P., Lee, W., Chen, X., & Fifield, B. (2022, August 16). Introducing the Ground Truth Maturity Framework for assessing and improving ground truth data quality. Meta Research. https://research.facebook.com/blog/2022/8/-introducing-the-ground-truth-maturity-framework-for-assessing-and-improving-ground-truth-data-quality/

Zileli, R., & Söyler, M. (2020). Analysis of corner kicks in FIFA 2018 World Cup. https://doi.org/10.14198/jhse.2022.171.15

APPENDICES

APPENDIX A

Football Commentary Labeling

User Manual

Football Commentary Labeling Kullanım Kılavuzu

Anıl ERKUL

Sütun Tanımları:

Kategorilendirme işlemini yapacağımız Comments Sayfasının sütunları şu şekilde tanımlanmıştır:

Kolon İsmi	**Kolon Tanımı**	**Kullanıcı mı Dolduracak?**	**Seçenekler**
1) Match_Name	Belirtilen commentin geçtiği maçtaki takımların isimleri.	Hayır	
2) Team_1	Ev sahibi takım.	Hayır	
3) Team_2	Konuk takım.	Hayır	
4) Min	Commentin gerçekleştiği dakika	Hayır	
5) Is there a Cross?	Belirtilen comment'te orta açılma/crossing mevcut mu?	Evet	No, Yes, Not Sure
6) Crossing Team	Orta açan takım	Evet	Team_1 or Team_2
7) Crossing Player	Ortayı açan oyuncu	Evet	Player from Team_1 or Team_2
8) Outcome of Cross - 1	Ortanın nasıl sonuçlandığı – 1. Seviye	Evet	Cross_Outcome_Dictionary
9) Outcome of Cross - 2	Ortanın nasıl sonuçlandığı – 2. Seviye	Evet	Cross_Outcome_Dictionary
10) Outcome of Cross - 3	Ortanın nasıl sonuçlandığı – 3. Seviye	Evet	Cross_Outcome_Dictionary
11) Quality of Cross	Ortanın kalitesi	Evet	Very Good, Good, Neutral, Bad, Very Bad, Not Sure
12) Definition of Cross	Girilen outcome of crossların açıklaması.	Hayır	
13) Is there a second cross?	Belirtilen commentte birden fazla orta açma aksiyonu var mı seçilir.	Evet	No, Yes, Not Sure
14-18) 6-9 sütunlarıyla aynı şekilde 2. orta için doldurulur.		Evet	

Nasıl Kullanılır?

Comments Sayfasının Açılışı

Öncelikle aşağıda gösterildiği şekilde Comments sayfasına geçiş yapılır

Football Comment Labeling

Bu dokümanda 2022-2023 Premier Lig sezonuna ait goal.com'dan alınmış maç anlatımları/commentleri bulunuyor. Eğer yapılacak işlemlere hakimseniz doğrudan Comments sayfasına geçerek işlemleri yapmaya başlayabilirsiniz. Hızlıca başlayabilmek ve/veya bir comment nasıl doldurulur için de bir örneği yine PDF'ten adım adım görebilirsiniz.

Amaç

Amacımız bu commentler içerisinde orta açma (crossing) aksiyonunun yapıldığı anlatıları (commentleri) ayırt etmek ve bu ortaların özelliklerini belirleyebilmek. Her bir comment (anlatının) için orta aksiyonu içerip içermediği (Is there a cross), ortanın nasıl sonuçlandığı (outcome of corss), ortanın kalitesi (quality of cross), orta aksiyonunu gerçekleştiren takım (crossing team) ve orta aksiyonunu gerçekleştiren oyuncunun ismi (crossing player) tespit edilmeye çalışılıyor. Bunun için de bu anketi dolduran kişilerden bu bilgileri dokümana işlemelerini istiyoruz.

Excel Dokümanın Yapısı

1) Main (bu sayfa) sayfasında bu dokümanın amacı, içeriği ve yapılacaklar açıklanmakta.
2) Comments sayfasında dokümanı paylaşma amacımız olan, her bir commenti içeren ve kategorilendirme yapacağımız bölüm yer alıyor.
3) Cross_Outcome_Dictionary sayfasında, orta açıldığında ortanın sonucunun ne olacağını belirlediğimiz kategoriler ve bu kategorilerin tanımlamarı yer alıyor.Bir nevi sözlük olarak düşünebilirsiniz.
4) Match_Links sayfasında her maçın tüm halini izleyebileceğiniz linkler yer almakta. Kafanızı karıştıran veya anlamakta zorlandığınız bir comment gördüğünüzde maçlarda belirtilen dakikalara bakarak kategorilendirme yapabilirsiniz.
5) Slangs sayfasında İngilizce'de futbol terimi olarak sıklıkla kullanılan argo, resmi olmayan kelimeler tanımlanmıştır. Commentlerde anlamadığınız kelimelerde buradan bilgi alabilirsiniz.
6) Sonraki sayfaların tamamında ise Premier Lig'de 2022-23 sezonunda bulunan takımların o sezonki kadroları bulunuyor.

Comments Sayfası İçeriği

Kategorilendirme işlemini yapacağımız Comments Sayfasının sütunları şu şekilde tanımlanmıştır:
1) Match_Name: Belirtilen commentin geçtiği maçtaki takımların isimleri. Üzerine tıklayarak belirtilen maçın linkine de ulaşılabilir
2) Team_1: Ev sahibi takım. Üzerine tıklar
3) Team_1: Konuk takım. Üzerine tıklanar
4) Min: Commentin gerçekleştiği dakika
Buradan sonraki kolonlar kullanıcı tarafından doldurulacaktır.

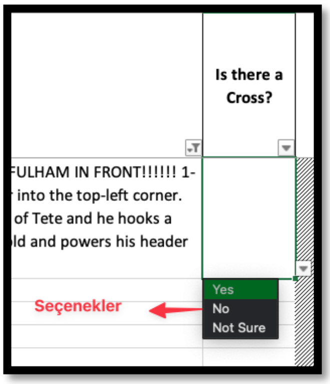
5) Is there a Cross? : Belirtilen comment'te orta açılma/crossing mevcut mu seçilir. Seçenekler: Yes, No, Not Sure
6) Outcome of Cross: Ortanın nasıl sonuçlandığı. Cross_Outcome_Dictionary sayfasında belirtilen outcomelardan biri seçilir. Outcomeların tanımları

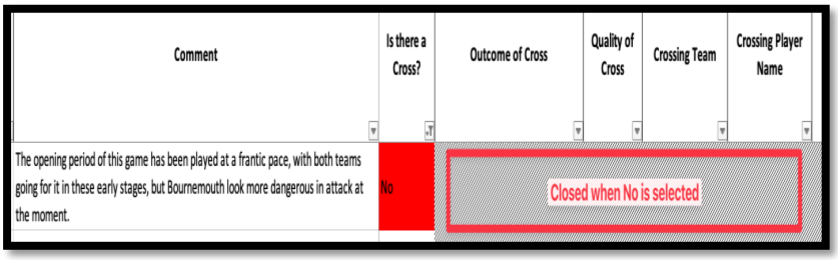
Comment Kolonları

Aşağıda Fulham Liverpool maçının 32. Dakikasına ait bir maç yorumu bulunmaktadır. Commenti okuduktan sonra “Is there a Cross” sütunundaki hücreden bu yorumda orta aksiyonu olup olmadığına karar verebiliriz. Bunun için drop-down olarak gelen alandan şekilde belirtildiği gibi “Yes”, “No” veya “Not Sure” seçeneklerinden bir tanesi seçilir. “No” seçeneği seçildiği taktirde sonraki sütunlar açılmaz. Bir sonraki satırdaki commente geçilebilir.

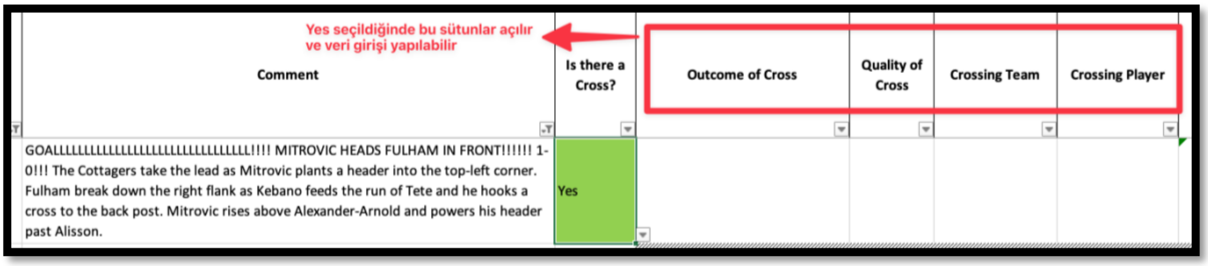
Match_Name	Team_1	Team_2	Min	Comment	Is there a Cross?
Fulham vs Liverpool	Fulham	Liverpool	32'	GOALLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL!!!! MITROVIC HEADS FULHAM IN FRONT!!!!!! 1-0!!! The Cottagers take the lead as Mitrovic plants a header into the top-left corner. Fulham break down the right flank as Kebano feeds the run of Tete and he hooks a cross to the back post. Mitrovic rises above Alexander-Arnold and powers his header past Alisson.	

Is there a Cross Seçimi



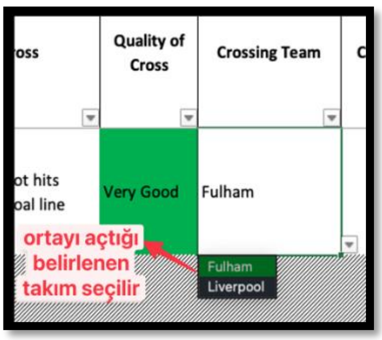


Buradaki örnekte comment'te orta açma aksiyonunun gerçekleştiği gözüküyor. Bu durumda "Yes" seçeneğini seçtiğimizde yanda taralı olan sütunlar açılarak şekilde gösterildiği gibi veri girişi yapmamıza imkân veriliyor.



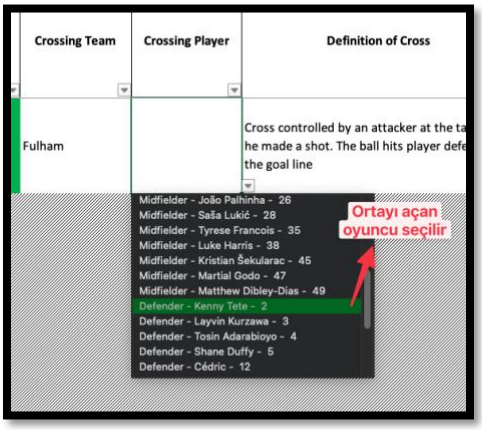
Crossing Team Seçimi

Cross quality seçildikten sonra ortayı açan takım (crossing team) seçilir. Şekilde gösterildiği şekilde bu örnekte ortayı açan takım Fulham olarak belilenmiş ve seçilmiştir. Eğer oyuncuların hangi takımdan olduğu konusunda karmaşa yaşanıyorsa takımların üzerine tıklanarak tüm kadroları görülebilir.



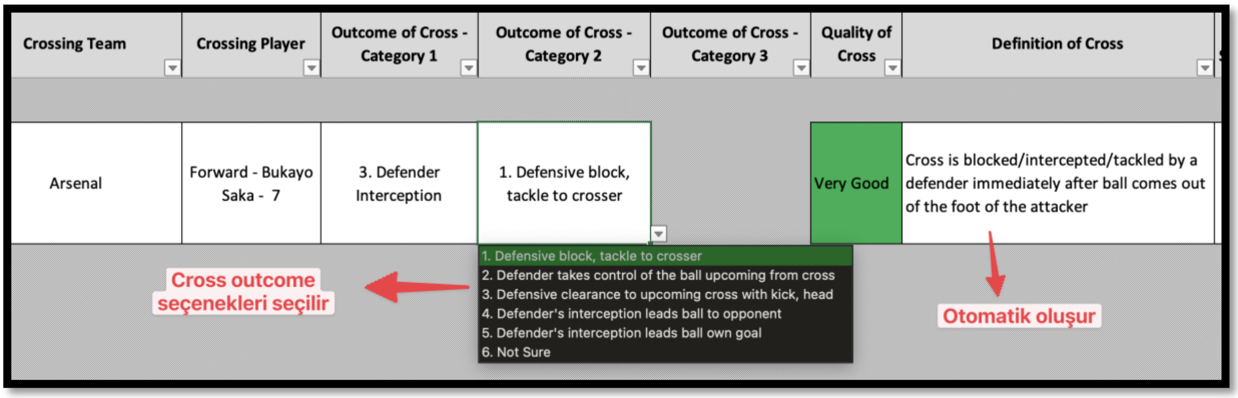
Crossing Player Seçimi

Takım seçildikten sonra ortayı açan oyuncu (Crossing Player) seçilir. Burada bir önceki sütunda ortayı açan takım seçildiğinden, sadece o takıma ait oyuncular listelenir. Bu comment'te ortayı açan oyuncu Tete olduğundan şekilde seçim yapılır.



Outcome of Crossların Seçimi

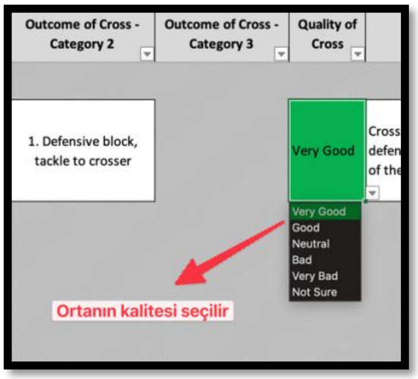
Outcome of cross sütunu, tabloda gösterildiği şekildeki seçeneklerden biriyle doldurulur. Burada açılan ortaya vurulan kafayla gol olduğunu görüyoruz. Bu sebeple aşağıdaki şekilde gösterildiği şekilde "The cross receiver scores, goal" seçilir. Definition of Cross sütunun otomatik geldiği görülür, buradan seçtiğimiz seçeneği teyit edebiliriz.



1. Cross on target, reaches a teammate	2. Cross off target, not reaches a teammate	3. Defender Interception	4. Goalkeeper Interception	5. Referee Interception
1. Cross Receiver Shoots/Heads	1. Team retains possession, ball is recycled	1. Defensive block, tackle to crosser	1. Goalkeeper catches/ gathers the ball upcoming from cross	1. Penalty
1. Cross Receiver Scores, Goal	2. Cross goes out of pitch/play	2. Defender takes control of the ball upcoming from cross	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	2. Offside
2. Cross receiver's shot hits goalpost, woodwork	3. Cross has no contact in the box from both teams	3. Defensive clearance to upcoming cross with kick, head etc.	1. Direction is not mentioned	3. Freekick for attacking team (direct/indirect)
3. Cross receiver's shot blocked by defender		1. Direction is not mentioned	2. To corner	4. Freekick for defending team (direct/indirect)
4. Cross receiver's shot saved by goalkeeper		2. To corner	3. To throw in	
5. Cross receiver's shot goes out, attempt off target		3. To throw in	4. Out of box	
6. Cross receiver's shot hits teammate		4. Out of box	5. To Opponent	
2. Cross Receiver Passes/Crosses to a teammate		5. To Opponent	6. Inside penalty area	
1. Passed teammate shoots and Scores, Goal		6. Inside penalty area	7. To a teammate	
2. Passed teammate shoots and misses/goalkeeper saves		7. To a teammate	8. Not Sure	
3. Passed teammate passes to another teammate		4. Defender's interception leads ball to opponent	3. Goalkeeper's interception leads ball to opponent	
4. Passed teammate cannot control the ball		5. Defender's interception leads ball own goal	4. Goalkeeper's interception leads ball own goal	
5. Pass is intercepted by defender/goalkeeper				
6. Pass to opponent, mispass				
3. Cross Receiver Dribbles				
4. Cross Receiver cannot control or touch the ball, losts possesion				

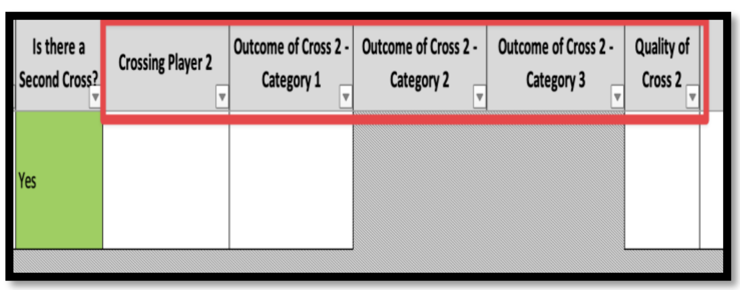
Cross Quality Seçimi

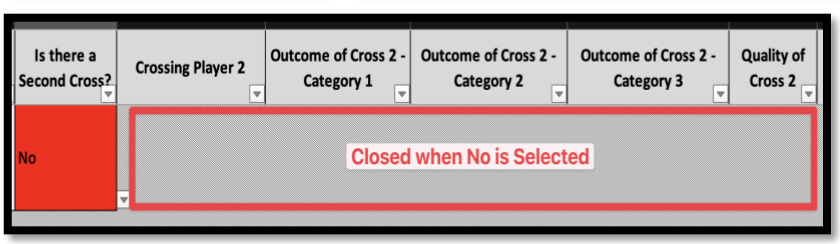
Outcome seçildikten sonra Cross quality'i seçeneği seçilir. Şekilde gösterildiği gibi 5 seçenekten bir tanesi seçilir. Burada golle sonuçlandığı için Very Good Seçilebilir.



Is There a Second Cross Seçimi

Ortayı açan oyuncuyu seçtikten sonra son olarak ikini orta var mı durumu sorgulanan "Is there a second cross" sütunu vardır. Çok sık rastlanmasa da bazı commentlerde iki ortadan veya uzun toptan bahsedildiği görülür. Burada da şekilde gösterildiği şekilde "Yes","No" veya "Not Sure" seçeneklerinden biri seçilir. Burada "Yes" veya "Not Sure" seçilirse sonraki sütunlar açılır, aksi halde doldurulmasına gerek olmadığından sonraki sütunlar açılmaz. Açıldığı taktirde aynı bir önceki işlemlerde yaptığımız gibi Crossing Player, Outcome of Cross – 1, Outcome of Cross – 2, Outcome of Cross – 3 ve Quality of Cross seçimi yapılır. "No" seçilirse sonraki sütunlar açılmaz.





APPENDIX B

Tables for Inter Annotator Agreement (IAA)

Table 49 IAA Metrics for "Is There a Cross" Label

#	Labeler	Labeled Data	Raw Agreement	Kappa	Level of Agreement
1	Labeler 1	1100	0.890	0.773	Moderate
2	Labeler 2	1106	0.812	0.546	Weak
3	Labeler 3	1096	0.821	0.550	Weak
4	Labeler 4	45	0.800	0.595	Moderate
5	Labeler 5	1095	0.907	0.769	Moderate
6	Labeler 6	523	0.876	0.714	Moderate
7	Labeler 7	370	0.886	0.717	Moderate
8	Labeler 8	572	0.883	0.718	Moderate
9	Labeler 9	1089	0.848	0.643	Moderate
10	Labeler 10	337	0.872	0.679	Moderate
11	Labeler 11	1513	0.912	0.795	Strong
12	Labeler 12	1394	0.885	0.747	Moderate
13	Labeler 13	338	0.929	0.840	Strong
14	Labeler 14	386	0.873	0.700	Moderate
15	Labeler 15	1089	0.908	0.777	Moderate
16	Labeler 16	1102	0.855	0.642	Moderate

Table 50 IAA Metrics for "Crossing Team" Label per match

#	**Labeler**	**Match No**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Level of Agreement**
1	Labeler 1	3	5	1.000	1.000	Almost Perfect
2	Labeler 1	23	5	1.000	1.000	Almost Perfect
3	Labeler 1	41	18	1.000	1.000	Almost Perfect
4	Labeler 1	61	16	1.000	1.000	Almost Perfect
5	Labeler 1	64	11	1.000	1.000	Almost Perfect
6	Labeler 1	82	15	1.000	1.000	Almost Perfect
7	Labeler 1	86	21	1.000	1.000	Almost Perfect
8	Labeler 1	102	26	1.000	1.000	Almost Perfect
9	Labeler 1	118	15	1.000	1.000	Almost Perfect
10	Labeler 1	122	18	1.000	1.000	Almost Perfect
11	Labeler 1	131	16	0.938	0.875	Strong
12	Labeler 1	138	18	1.000	1.000	Almost Perfect
13	Labeler 1	146	14	0.929	0.851	Strong
14	Labeler 1	162	27	1.000	1.000	Almost Perfect
15	Labeler 1	178	15	0.933	0.867	Strong
16	Labeler 1	204	13	1.000	1.000	Almost Perfect
17	Labeler 1	210	21	1.000	1.000	Almost Perfect
18	Labeler 1	211	10	1.000	1.000	Almost Perfect
19	Labeler 1	213	16	1.000	1.000	Almost Perfect
20	Labeler 1	219	6	1.000	1.000	Almost Perfect
21	Labeler 1	226	14	1.000	1.000	Almost Perfect
22	Labeler 1	234	19	1.000	1.000	Almost Perfect
23	Labeler 1	240	4	1.000	1.000	Almost Perfect
24	Labeler 1	249	23	1.000	1.000	Almost Perfect
25	Labeler 1	257	8	1.000	1.000	Almost Perfect
26	Labeler 1	281	5	1.000	1.000	Almost Perfect
27	Labeler 1	304	12	1.000	1.000	Almost Perfect
28	Labeler 2	7	2	1.000	1.000	Almost Perfect
29	Labeler 2	9	13	1.000	1.000	Almost Perfect
30	Labeler 2	21	8	1.000	1.000	Almost Perfect
31	Labeler 2	81	6	1.000	1.000	Almost Perfect
32	Labeler 2	89	5	1.000	1.000	Almost Perfect
33	Labeler 2	93	6	1.000	1.000	Almost Perfect
34	Labeler 2	110	13	1.000	1.000	Almost Perfect
35	Labeler 2	115	7	1.000	1.000	Almost Perfect

Table 50 (cont.)

36	Labeler 2	135	10	1.000	1.000	Almost Perfect
37	Labeler 2	145	5	1.000	1.000	Almost Perfect
38	Labeler 2	151	12	1.000	1.000	Almost Perfect
39	Labeler 2	153	8	1.000	1.000	Almost Perfect
40	Labeler 2	170	3	1.000	1.000	Almost Perfect
41	Labeler 2	187	2	1.000	1.000	Almost Perfect
42	Labeler 2	188	7	1.000	1.000	Almost Perfect
43	Labeler 2	190	9	1.000	1.000	Almost Perfect
44	Labeler 2	205	8	1.000	1.000	Almost Perfect
45	Labeler 2	225	11	1.000	1.000	Almost Perfect
46	Labeler 2	229	8	1.000	1.000	Almost Perfect
47	Labeler 2	230	5	1.000	1.000	Almost Perfect
48	Labeler 2	239	16	1.000	1.000	Almost Perfect
49	Labeler 2	263	9	1.000	1.000	Almost Perfect
50	Labeler 2	280	11	1.000	1.000	Almost Perfect
51	Labeler 2	283	14	1.000	1.000	Almost Perfect
52	Labeler 2	300	5	0.800	0.545	Weak
53	Labeler 2	303	10	0.900	0.800	Strong
54	Labeler 3	10	8	1.000	1.000	Almost Perfect
55	Labeler 3	11	4	0.750	0.500	Weak
56	Labeler 3	26	10	1.000	1.000	Almost Perfect
57	Labeler 3	35	5	0.800	0.545	Weak
58	Labeler 3	43	9	1.000	1.000	Almost Perfect
59	Labeler 3	55	11	0.909	0.792	Strong
60	Labeler 3	65	12	0.833	0.667	Moderate
61	Labeler 3	70	9	1.000	1.000	Almost Perfect
62	Labeler 3	79	10	1.000	1.000	Almost Perfect
63	Labeler 3	96	2	0.500	0.000	None
64	Labeler 3	100	4	1.000	1.000	Almost Perfect
65	Labeler 3	119	3	0.667	0.400	Weak
66	Labeler 3	124	9	1.000	1.000	Almost Perfect
67	Labeler 3	132	9	1.000	1.000	Almost Perfect
68	Labeler 3	147	5	0.400	0.000	None
69	Labeler 3	150	8	0.875	0.750	Moderate
70	Labeler 3	168	2	1.000	1.000	Almost Perfect
71	Labeler 3	198	4	1.000	1.000	Almost Perfect
72	Labeler 3	207	10	1.000	1.000	Almost Perfect
73	Labeler 3	223	10	1.000	1.000	Almost Perfect
74	Labeler 3	231	3	1.000	1.000	Almost Perfect
75	Labeler 3	237	11	1.000	1.000	Almost Perfect

Table 50 (cont.)

76	Labeler 3	252	11	1.000	1.000	Almost Perfect
77	Labeler 3	256	6	1.000	1.000	Almost Perfect
78	Labeler 3	292	15	1.000	1.000	Almost Perfect
79	Labeler 3	297	5	1.000	1.000	Almost Perfect
80	Labeler 4	98	15	1.000	1.000	Almost Perfect
81	Labeler 5	8	7	1.000	1.000	Almost Perfect
82	Labeler 5	18	11	0.909	0.744	Moderate
83	Labeler 5	33	13	0.923	0.629	Moderate
84	Labeler 5	44	9	1.000	1.000	Almost Perfect
85	Labeler 5	63	8	1.000	1.000	Almost Perfect
86	Labeler 5	68	3	1.000	1.000	Almost Perfect
87	Labeler 5	75	16	1.000	1.000	Almost Perfect
88	Labeler 5	83	15	0.867	0.727	Moderate
89	Labeler 5	88	7	0.857	0.588	Weak
90	Labeler 5	101	20	1.000	1.000	Almost Perfect
91	Labeler 5	108	10	1.000	1.000	Almost Perfect
92	Labeler 5	130	6	1.000	1.000	Almost Perfect
93	Labeler 5	137	15	0.933	0.857	Strong
94	Labeler 5	163	5	1.000	1.000	Almost Perfect
95	Labeler 5	177	8	1.000	1.000	Almost Perfect
96	Labeler 5	181	13	1.000	1.000	Almost Perfect
97	Labeler 5	196	6	1.000	1.000	Almost Perfect
98	Labeler 5	228	10	1.000	1.000	Almost Perfect
99	Labeler 5	248	9	1.000	1.000	Almost Perfect
100	Labeler 5	265	15	1.000	1.000	Almost Perfect
101	Labeler 5	266	16	0.938	0.875	Strong
102	Labeler 5	268	6	1.000	1.000	Almost Perfect
103	Labeler 5	273	13	1.000	1.000	Almost Perfect
104	Labeler 5	289	3	1.000	1.000	Almost Perfect
105	Labeler 5	302	9	1.000	1.000	Almost Perfect
106	Labeler 6	17	12	0.917	0.833	Strong
107	Labeler 6	37	18	0.944	0.880	Strong
108	Labeler 6	50	13	0.923	0.843	Strong
109	Labeler 6	56	17	0.882	0.767	Moderate
110	Labeler 6	73	12	1.000	1.000	Almost Perfect
111	Labeler 6	113	13	1.000	1.000	Almost Perfect
112	Labeler 6	154	9	0.778	0.357	Minimal
113	Labeler 6	172	5	0.800	0.615	Moderate
114	Labeler 6	202	4	1.000	1.000	Almost Perfect
115	Labeler 6	221	15	1.000	1.000	Almost Perfect

Table 50 (cont.)

116	Labeler 6	254	7	0.857	0.588	Weak
117	Labeler 6	277	9	1.000	1.000	Almost Perfect
118	Labeler 7	14	8	0.875	0.600	Moderate
119	Labeler 7	15	7	0.857	0.696	Moderate
120	Labeler 7	34	15	1.000	1.000	Almost Perfect
121	Labeler 7	36	7	0.857	0.720	Moderate
122	Labeler 7	42	9	1.000	1.000	Almost Perfect
123	Labeler 7	52	10	1.000	1.000	Almost Perfect
124	Labeler 7	57	6	1.000	1.000	Almost Perfect
125	Labeler 7	77	13	0.846	0.649	Moderate
126	Labeler 7	94	5	1.000	1.000	Almost Perfect
127	Labeler 8	5	17	1.000	1.000	Almost Perfect
128	Labeler 8	27	15	1.000	1.000	Almost Perfect
129	Labeler 8	49	13	0.923	0.806	Strong
130	Labeler 8	58	8	1.000	1.000	Almost Perfect
131	Labeler 8	106	8	1.000	1.000	Almost Perfect
132	Labeler 8	139	9	1.000	1.000	Almost Perfect
133	Labeler 8	157	15	1.000	1.000	Almost Perfect
134	Labeler 8	171	6	1.000	1.000	Almost Perfect
135	Labeler 8	208	17	1.000	1.000	Almost Perfect
136	Labeler 8	250	7	1.000	1.000	Almost Perfect
137	Labeler 8	264	10	1.000	1.000	Almost Perfect
138	Labeler 8	286	3	1.000	1.000	Almost Perfect
139	Labeler 8	295	6	1.000	1.000	Almost Perfect
140	Labeler 9	4	16	0.938	0.871	Strong
141	Labeler 9	22	10	1.000	1.000	Almost Perfect
142	Labeler 9	39	15	0.933	0.865	Strong
143	Labeler 9	67	16	0.938	0.862	Strong
144	Labeler 9	76	17	1.000	1.000	Almost Perfect
145	Labeler 9	90	17	0.882	0.742	Moderate
146	Labeler 9	91	3	0.667	0.400	Weak
147	Labeler 9	97	15	1.000	1.000	Almost Perfect
148	Labeler 9	107	9	1.000	1.000	Almost Perfect
149	Labeler 9	111	11	1.000	1.000	Almost Perfect
150	Labeler 9	136	1	1.000	1.000	Almost Perfect
151	Labeler 9	165	5	1.000	1.000	Almost Perfect
152	Labeler 9	166	10	0.900	0.783	Moderate
153	Labeler 9	176	13	1.000	1.000	Almost Perfect
154	Labeler 9	189	8	1.000	1.000	Almost Perfect
155	Labeler 9	191	7	1.000	1.000	Almost Perfect

Table 50 (cont.)

156	Labeler 9	227	13	0.923	0.755	Moderate
157	Labeler 9	235	5	1.000	1.000	Almost Perfect
158	Labeler 9	245	10	0.900	0.783	Moderate
159	Labeler 9	253	4	1.000	1.000	Almost Perfect
160	Labeler 9	262	3	1.000	1.000	Almost Perfect
161	Labeler 9	275	7	1.000	1.000	Almost Perfect
162	Labeler 9	278	9	0.778	0.571	Weak
163	Labeler 9	301	14	0.857	0.576	Weak
164	Labeler 10	112	4	1.000	1.000	Almost Perfect
165	Labeler 10	121	12	1.000	1.000	Almost Perfect
166	Labeler 10	160	12	1.000	1.000	Almost Perfect
167	Labeler 10	180	10	1.000	1.000	Almost Perfect
168	Labeler 10	197	2	1.000	1.000	Almost Perfect
169	Labeler 10	243	7	1.000	1.000	Almost Perfect
170	Labeler 10	269	10	1.000	1.000	Almost Perfect
171	Labeler 10	270	13	1.000	1.000	Almost Perfect
172	Labeler 11	12	8	1.000	1.000	Almost Perfect
173	Labeler 11	32	17	1.000	1.000	Almost Perfect
174	Labeler 11	48	7	1.000	1.000	Almost Perfect
175	Labeler 11	109	9	0.889	0.609	Moderate
176	Labeler 11	133	16	1.000	1.000	Almost Perfect
177	Labeler 11	134	14	1.000	1.000	Almost Perfect
178	Labeler 11	144	9	0.889	0.780	Moderate
179	Labeler 11	149	4	1.000	1.000	Almost Perfect
180	Labeler 11	174	12	1.000	1.000	Almost Perfect
181	Labeler 11	182	12	1.000	1.000	Almost Perfect
182	Labeler 11	186	13	1.000	1.000	Almost Perfect
183	Labeler 11	195	5	1.000	1.000	Almost Perfect
184	Labeler 11	200	17	1.000	1.000	Almost Perfect
185	Labeler 11	215	22	1.000	1.000	Almost Perfect
186	Labeler 11	216	12	0.917	0.800	Strong
187	Labeler 11	218	14	1.000	1.000	Almost Perfect
188	Labeler 11	222	14	0.929	0.632	Moderate
189	Labeler 11	238	13	1.000	1.000	Almost Perfect
190	Labeler 11	255	9	1.000	1.000	Almost Perfect
191	Labeler 11	258	15	1.000	1.000	Almost Perfect
192	Labeler 11	274	21	0.952	0.904	Almost Perfect
193	Labeler 11	279	14	0.929	0.857	Strong
194	Labeler 11	284	8	1.000	1.000	Almost Perfect
195	Labeler 11	288	7	1.000	1.000	Almost Perfect

Table 50 (cont.)

196	Labeler 11	291	12	1.000	1.000	Almost Perfect
197	Labeler 11	104	9	1.000	1.000	Almost Perfect
198	Labeler 11	125	11	1.000	1.000	Almost Perfect
199	Labeler 11	142	13	1.000	1.000	Almost Perfect
200	Labeler 11	161	17	1.000	1.000	Almost Perfect
201	Labeler 11	194	8	1.000	1.000	Almost Perfect
202	Labeler 11	206	8	1.000	1.000	Almost Perfect
203	Labeler 11	233	3	1.000	1.000	Almost Perfect
204	Labeler 11	267	15	1.000	1.000	Almost Perfect
205	Labeler 11	294	11	1.000	1.000	Almost Perfect
206	Labeler 11	296	5	1.000	1.000	Almost Perfect
207	Labeler 12	1	17	1.000	1.000	Almost Perfect
208	Labeler 12	16	16	0.938	0.862	Strong
209	Labeler 12	24	15	1.000	1.000	Almost Perfect
210	Labeler 12	29	7	1.000	1.000	Almost Perfect
211	Labeler 12	45	12	1.000	1.000	Almost Perfect
212	Labeler 12	54	5	1.000	1.000	Almost Perfect
213	Labeler 12	60	8	1.000	1.000	Almost Perfect
214	Labeler 12	66	12	1.000	1.000	Almost Perfect
215	Labeler 12	72	12	1.000	1.000	Almost Perfect
216	Labeler 12	80	9	1.000	1.000	Almost Perfect
217	Labeler 12	95	15	0.867	0.667	Moderate
218	Labeler 12	116	22	0.864	0.703	Moderate
219	Labeler 12	117	11	0.909	0.744	Moderate
220	Labeler 12	126	13	1.000	1.000	Almost Perfect
221	Labeler 12	140	11	0.909	0.814	Strong
222	Labeler 12	158	10	1.000	1.000	Almost Perfect
223	Labeler 12	159	10	1.000	1.000	Almost Perfect
224	Labeler 12	167	11	1.000	1.000	Almost Perfect
225	Labeler 12	179	15	1.000	1.000	Almost Perfect
226	Labeler 12	183	11	0.909	0.792	Strong
227	Labeler 12	192	11	1.000	1.000	Almost Perfect
228	Labeler 12	203	11	1.000	1.000	Almost Perfect
229	Labeler 12	224	14	1.000	1.000	Almost Perfect
230	Labeler 12	236	10	1.000	1.000	Almost Perfect
231	Labeler 12	244	18	1.000	1.000	Almost Perfect
232	Labeler 12	258	1	1.000	1.000	Almost Perfect
233	Labeler 12	92	22	1.000	1.000	Almost Perfect
234	Labeler 12	120	6	1.000	1.000	Almost Perfect
235	Labeler 12	143	11	1.000	1.000	Almost Perfect

Table 50 (cont.)

236	Labeler 12	152	13	1.000	1.000	Almost Perfect
237	Labeler 12	193	18	1.000	1.000	Almost Perfect
238	Labeler 12	217	19	1.000	1.000	Almost Perfect
239	Labeler 12	285	4	0.750	0.500	Weak
240	Labeler 13	114	13	1.000	1.000	Almost Perfect
241	Labeler 13	155	15	0.867	0.706	Moderate
242	Labeler 13	201	6	1.000	1.000	Almost Perfect
243	Labeler 13	214	9	1.000	1.000	Almost Perfect
244	Labeler 13	241	11	1.000	1.000	Almost Perfect
245	Labeler 13	251	15	1.000	1.000	Almost Perfect
246	Labeler 13	272	19	1.000	1.000	Almost Perfect
247	Labeler 13	290	12	0.917	0.824	Strong
248	Labeler 14	2	16	1.000	1.000	Almost Perfect
249	Labeler 14	7	10	1.000	1.000	Almost Perfect
250	Labeler 14	25	4	1.000	1.000	Almost Perfect
251	Labeler 14	28	13	0.846	0.409	Weak
252	Labeler 14	40	15	0.933	0.865	Strong
253	Labeler 14	46	6	1.000	1.000	Almost Perfect
254	Labeler 14	62	5	0.800	0.545	Weak
255	Labeler 14	71	9	0.889	0.769	Moderate
256	Labeler 14	87	14	1.000	1.000	Almost Perfect
257	Labeler 15	6	5	1.000	1.000	Almost Perfect
258	Labeler 15	13	10	0.900	0.800	Strong
259	Labeler 15	19	9	1.000	1.000	Almost Perfect
260	Labeler 15	31	7	1.000	1.000	Almost Perfect
261	Labeler 15	51	15	0.933	0.857	Strong
262	Labeler 15	53	5	1.000	1.000	Almost Perfect
263	Labeler 15	59	18	1.000	1.000	Almost Perfect
264	Labeler 15	78	10	1.000	1.000	Almost Perfect
265	Labeler 15	84	14	1.000	1.000	Almost Perfect
266	Labeler 15	99	12	1.000	1.000	Almost Perfect
267	Labeler 15	123	6	1.000	1.000	Almost Perfect
268	Labeler 15	127	16	1.000	1.000	Almost Perfect
269	Labeler 15	141	13	1.000	1.000	Almost Perfect
270	Labeler 15	148	14	1.000	1.000	Almost Perfect
271	Labeler 15	173	13	1.000	1.000	Almost Perfect
272	Labeler 15	175	12	1.000	1.000	Almost Perfect
273	Labeler 15	185	9	1.000	1.000	Almost Perfect
274	Labeler 15	199	6	1.000	1.000	Almost Perfect
275	Labeler 15	209	3	1.000	1.000	Almost Perfect

Table 50 (cont.)

276	Labeler 15	242	14	1.000	1.000	Almost Perfect
277	Labeler 15	261	4	1.000	1.000	Almost Perfect
278	Labeler 15	271	8	1.000	1.000	Almost Perfect
279	Labeler 15	276	15	0.933	0.815	Strong
280	Labeler 15	287	16	1.000	1.000	Almost Perfect
281	Labeler 15	293	10	1.000	1.000	Almost Perfect
282	Labeler 16	20	4	1.000	1.000	Almost Perfect
283	Labeler 16	30	17	1.000	1.000	Almost Perfect
284	Labeler 16	38	13	0.692	0.435	Weak
285	Labeler 16	47	9	1.000	1.000	Almost Perfect
286	Labeler 16	69	15	0.933	0.857	Strong
287	Labeler 16	74	18	1.000	1.000	Almost Perfect
288	Labeler 16	85	3	1.000	1.000	Almost Perfect
289	Labeler 16	103	5	0.800	0.000	None
290	Labeler 16	105	14	1.000	1.000	Almost Perfect
291	Labeler 16	128	12	1.000	1.000	Almost Perfect
292	Labeler 16	129	6	1.000	1.000	Almost Perfect
293	Labeler 16	156	4	1.000	1.000	Almost Perfect
294	Labeler 16	164	7	1.000	1.000	Almost Perfect
295	Labeler 16	169	8	1.000	1.000	Almost Perfect
296	Labeler 16	184	8	1.000	1.000	Almost Perfect
297	Labeler 16	209	1	1.000	1.000	Almost Perfect
298	Labeler 16	212	9	1.000	1.000	Almost Perfect
299	Labeler 16	220	8	1.000	1.000	Almost Perfect
300	Labeler 16	232	3	1.000	1.000	Almost Perfect
301	Labeler 16	246	7	1.000	1.000	Almost Perfect
302	Labeler 16	247	9	0.889	0.609	Moderate
303	Labeler 16	259	10	1.000	1.000	Almost Perfect
304	Labeler 16	260	7	1.000	1.000	Almost Perfect
305	Labeler 16	282	7	1.000	1.000	Almost Perfect
306	Labeler 16	298	11	1.000	1.000	Almost Perfect
307	Labeler 16	299	6	1.000	1.000	Almost Perfect

Table 51 Weighted Averaged IAA Metrics for “Crossing Team” Label

#	**Labeler**	**Total Labeled**	**Weighted Averaged Raw Agreement**	**Weighted Averaged Kappa**	**Level of Agreement**
1	Labeler 1	391	0.992	0.984	Almost Perfect
2	Labeler 2	213	0.991	0.980	Almost Perfect
3	Labeler 3	195	0.944	0.890	Strong
4	Labeler 4	15	1.000	1.000	Almost Perfect
5	Labeler 5	253	0.972	0.926	Almost Perfect
6	Labeler 6	134	0.933	0.845	Strong
7	Labeler 7	80	0.938	0.852	Strong
8	Labeler 8	134	0.993	0.981	Almost Perfect
9	Labeler 9	238	0.945	0.875	Strong
10	Labeler 10	70	1.000	1.000	Almost Perfect
11	Labeler 11	404	0.985	0.958	Almost Perfect
12	Labeler 12	400	0.975	0.943	Almost Perfect
13	Labeler 13	100	0.970	0.935	Almost Perfect
14	Labeler 14	92	0.946	0.847	Strong
15	Labeler 15	264	0.989	0.974	Almost Perfect
16	Labeler 16	221	0.968	0.918	Almost Perfect

Table 52 IAA Metrics for “Crossing Player” Label per match

#	Labeler	Match No	Labeled Data	Raw Agreement	Kappa	Agreement
1	Labeler 1	3	5	0.800	0.750	Moderate
2	Labeler 1	23	5	1.000	1.000	Almost Perfect
3	Labeler 1	41	18	0.833	0.815	Strong
4	Labeler 1	61	16	0.938	0.922	Almost Perfect
5	Labeler 1	64	11	1.000	1.000	Almost Perfect
6	Labeler 1	82	15	0.933	0.920	Almost Perfect
7	Labeler 1	86	21	1.000	1.000	Almost Perfect
8	Labeler 1	102	26	0.885	0.866	Strong
9	Labeler 1	118	15	0.733	0.695	Moderate
10	Labeler 1	122	18	0.889	0.874	Strong
11	Labeler 1	131	16	0.813	0.789	Moderate
12	Labeler 1	138	18	0.778	0.727	Moderate
13	Labeler 1	146	14	1.000	1.000	Almost Perfect
14	Labeler 1	162	27	0.963	0.956	Almost Perfect
15	Labeler 1	178	15	1.000	1.000	Almost Perfect
16	Labeler 1	204	13	0.846	0.824	Strong
17	Labeler 1	210	21	1.000	1.000	Almost Perfect
18	Labeler 1	211	10	1.000	1.000	Almost Perfect
19	Labeler 1	213	16	0.875	0.852	Strong
20	Labeler 1	219	6	0.667	0.538	Weak
21	Labeler 1	226	14	1.000	1.000	Almost Perfect
22	Labeler 1	234	19	0.895	0.870	Strong
23	Labeler 1	240	4	1.000	1.000	Almost Perfect
24	Labeler 1	249	23	0.870	0.858	Strong
25	Labeler 1	257	8	1.000	1.000	Almost Perfect
26	Labeler 1	281	5	0.800	0.667	Moderate
27	Labeler 1	304	12	0.833	0.800	Strong
28	Labeler 2	7	2	1.000	1.000	Almost Perfect
29	Labeler 2	9	13	0.615	0.575	Weak
30	Labeler 2	21	8	0.875	0.846	Strong
31	Labeler 2	81	6	0.500	0.379	Minimal
32	Labeler 2	89	5	0.600	0.500	Weak
33	Labeler 2	93	6	0.833	0.793	Strong
34	Labeler 2	110	13	0.923	0.916	Almost Perfect
35	Labeler 2	115	7	1.000	1.000	Almost Perfect
36	Labeler 2	135	10	1.000	1.000	Almost Perfect

Table 52 (cont.)

37	Labeler 2	145	5	0.800	0.762	Moderate
38	Labeler 2	151	12	0.667	0.597	Moderate
39	Labeler 2	153	8	0.875	0.784	Moderate
40	Labeler 2	170	3	1.000	1.000	Almost Perfect
41	Labeler 2	187	2	1.000	1.000	Almost Perfect
42	Labeler 2	188	7	0.857	0.821	Strong
43	Labeler 2	190	9	0.889	0.850	Strong
44	Labeler 2	205	8	0.875	0.855	Strong
45	Labeler 2	225	11	0.818	0.780	Moderate
46	Labeler 2	229	8	0.500	0.439	Weak
47	Labeler 2	230	5	1.000	1.000	Almost Perfect
48	Labeler 2	239	16	0.813	0.793	Strong
49	Labeler 2	263	9	0.778	0.753	Moderate
50	Labeler 2	280	11	1.000	1.000	Almost Perfect
51	Labeler 2	283	14	1.000	1.000	Almost Perfect
52	Labeler 2	300	5	0.800	0.750	Moderate
53	Labeler 2	303	10	0.800	0.730	Moderate
54	Labeler 3	10	8	0.750	0.714	Moderate
55	Labeler 3	11	4	0.750	0.667	Moderate
56	Labeler 3	26	10	0.700	0.670	Moderate
57	Labeler 3	35	5	0.800	0.762	Moderate
58	Labeler 3	43	9	1.000	1.000	Almost Perfect
59	Labeler 3	55	11	0.818	0.788	Moderate
60	Labeler 3	65	12	0.833	0.808	Strong
61	Labeler 3	70	9	0.556	0.390	Minimal
62	Labeler 3	79	10	0.800	0.770	Moderate
63	Labeler 3	96	2	0.500	0.333	Minimal
64	Labeler 3	100	4	0.750	0.692	Moderate
65	Labeler 3	119	3	0.667	0.500	Weak
66	Labeler 3	124	9	0.889	0.868	Strong
67	Labeler 3	132	9	0.556	0.486	Weak
68	Labeler 3	147	5	0.800	0.583	Weak
69	Labeler 3	150	8	0.875	0.849	Strong
70	Labeler 3	168	2	1.000	1.000	Almost Perfect
71	Labeler 3	198	4	0.750	0.692	Moderate
72	Labeler 3	207	10	0.900	0.878	Strong
73	Labeler 3	223	10	1.000	1.000	Almost Perfect
74	Labeler 3	231	3	1.000	1.000	Almost Perfect
75	Labeler 3	237	11	0.545	0.500	Weak
76	Labeler 3	252	11	0.909	0.893	Strong

Table 52 (cont.)

77	Labeler 3	256	6	0.833	0.760	Moderate
78	Labeler 3	292	15	0.933	0.923	Almost Perfect
79	Labeler 3	297	5	0.800	0.737	Moderate
80	Labeler 4	98	15	0.667	0.625	Moderate
81	Labeler 5	8	7	1.000	1.000	Almost Perfect
82	Labeler 5	18	11	0.909	0.892	Strong
83	Labeler 5	33	13	0.923	0.911	Almost Perfect
84	Labeler 5	44	9	0.778	0.727	Moderate
85	Labeler 5	63	8	0.750	0.714	Moderate
86	Labeler 5	68	3	1.000	1.000	Almost Perfect
87	Labeler 5	75	16	1.000	1.000	Almost Perfect
88	Labeler 5	83	15	0.867	0.842	Strong
89	Labeler 5	88	7	0.857	0.825	Strong
90	Labeler 5	101	20	0.750	0.729	Moderate
91	Labeler 5	108	10	1.000	1.000	Almost Perfect
92	Labeler 5	130	6	1.000	1.000	Almost Perfect
93	Labeler 5	137	15	0.800	0.763	Moderate
94	Labeler 5	163	5	0.800	0.750	Moderate
95	Labeler 5	177	8	0.875	0.849	Strong
96	Labeler 5	181	13	0.846	0.818	Strong
97	Labeler 5	196	6	1.000	1.000	Almost Perfect
98	Labeler 5	228	10	1.000	1.000	Almost Perfect
99	Labeler 5	248	9	1.000	1.000	Almost Perfect
100	Labeler 5	265	15	0.933	0.920	Almost Perfect
101	Labeler 5	266	16	0.938	0.925	Almost Perfect
102	Labeler 5	268	6	1.000	1.000	Almost Perfect
103	Labeler 5	273	13	1.000	1.000	Almost Perfect
104	Labeler 5	289	3	0.667	0.500	Weak
105	Labeler 5	302	9	1.000	1.000	Almost Perfect
106	Labeler 6	17	12	0.917	0.896	Strong
107	Labeler 6	37	18	0.889	0.874	Strong
108	Labeler 6	50	13	0.846	0.806	Strong
109	Labeler 6	56	17	0.941	0.934	Almost Perfect
110	Labeler 6	73	12	1.000	1.000	Almost Perfect
111	Labeler 6	113	13	0.923	0.912	Almost Perfect
112	Labeler 6	154	9	0.889	0.868	Strong
113	Labeler 6	172	5	1.000	1.000	Almost Perfect
114	Labeler 6	202	4	1.000	1.000	Almost Perfect
115	Labeler 6	221	15	0.800	0.764	Moderate
116	Labeler 6	254	7	0.857	0.806	Strong

Table 52 (cont.)

117	Labeler 6	277	9	1.000	1.000	Almost Perfect
118	Labeler 7	14	8	0.875	0.849	Strong
119	Labeler 7	15	7	0.857	0.825	Strong
120	Labeler 7	34	15	0.800	0.751	Moderate
121	Labeler 7	36	7	1.000	1.000	Almost Perfect
122	Labeler 7	42	9	0.778	0.750	Moderate
123	Labeler 7	52	10	1.000	1.000	Almost Perfect
124	Labeler 7	57	6	0.667	0.586	Weak
125	Labeler 7	77	13	0.769	0.735	Moderate
126	Labeler 7	94	5	0.800	0.750	Moderate
127	Labeler 8	5	17	0.588	0.541	Weak
128	Labeler 8	27	15	0.933	0.921	Almost Perfect
129	Labeler 8	49	13	0.846	0.783	Moderate
130	Labeler 8	58	8	0.875	0.843	Strong
131	Labeler 8	106	8	0.750	0.704	Moderate
132	Labeler 8	139	9	0.889	0.786	Moderate
133	Labeler 8	157	15	0.800	0.773	Moderate
134	Labeler 8	171	6	0.833	0.800	Strong
135	Labeler 8	208	17	0.882	0.863	Strong
136	Labeler 8	250	7	0.857	0.837	Strong
137	Labeler 8	264	10	0.700	0.583	Weak
138	Labeler 8	286	3	1.000	1.000	Almost Perfect
139	Labeler 8	295	6	1.000	1.000	Almost Perfect
140	Labeler 9	4	16	0.938	0.925	Almost Perfect
141	Labeler 9	22	10	0.800	0.773	Moderate
142	Labeler 9	39	15	0.933	0.923	Almost Perfect
143	Labeler 9	67	16	0.875	0.863	Strong
144	Labeler 9	76	17	0.882	0.866	Strong
145	Labeler 9	90	17	0.882	0.867	Strong
146	Labeler 9	91	3	0.333	0.250	Minimal
147	Labeler 9	97	15	0.733	0.701	Moderate
148	Labeler 9	107	9	0.889	0.855	Strong
149	Labeler 9	111	11	0.727	0.697	Moderate
150	Labeler 9	136	1	1.000	1.000	Almost Perfect
151	Labeler 9	165	5	1.000	1.000	Almost Perfect
152	Labeler 9	166	10	0.900	0.880	Strong
153	Labeler 9	176	13	0.923	0.912	Almost Perfect
154	Labeler 9	189	8	0.875	0.837	Strong
155	Labeler 9	191	7	1.000	1.000	Almost Perfect
156	Labeler 9	227	13	0.923	0.912	Almost Perfect

Table 52 (cont.)

157	Labeler 9	235	5	0.800	0.750	Moderate
158	Labeler 9	245	10	0.900	0.881	Strong
159	Labeler 9	253	4	0.750	0.692	Moderate
160	Labeler 9	262	3	1.000	1.000	Almost Perfect
161	Labeler 9	275	7	0.571	0.488	Weak
162	Labeler 9	278	9	0.556	0.507	Weak
163	Labeler 9	301	14	0.714	0.673	Moderate
164	Labeler 10	112	4	1.000	1.000	Almost Perfect
165	Labeler 10	121	12	0.833	0.806	Strong
166	Labeler 10	160	12	1.000	1.000	Almost Perfect
167	Labeler 10	180	10	1.000	1.000	Almost Perfect
168	Labeler 10	197	2	1.000	1.000	Almost Perfect
169	Labeler 10	243	7	0.857	0.821	Strong
170	Labeler 10	269	10	0.900	0.877	Strong
171	Labeler 10	270	13	0.923	0.893	Strong
172	Labeler 11	12	8	1.000	1.000	Almost Perfect
173	Labeler 11	32	17	0.824	0.801	Strong
174	Labeler 11	48	7	1.000	1.000	Almost Perfect
175	Labeler 11	109	9	0.889	0.866	Strong
176	Labeler 11	133	16	0.938	0.925	Almost Perfect
177	Labeler 11	134	14	0.929	0.920	Almost Perfect
178	Labeler 11	144	9	0.889	0.836	Strong
179	Labeler 11	149	4	1.000	1.000	Almost Perfect
180	Labeler 11	174	12	1.000	1.000	Almost Perfect
181	Labeler 11	182	12	0.750	0.723	Moderate
182	Labeler 11	186	13	1.000	1.000	Almost Perfect
183	Labeler 11	195	5	1.000	1.000	Almost Perfect
184	Labeler 11	200	17	1.000	1.000	Almost Perfect
185	Labeler 11	215	22	0.909	0.884	Strong
186	Labeler 11	216	12	0.917	0.894	Strong
187	Labeler 11	218	14	1.000	1.000	Almost Perfect
188	Labeler 11	222	14	0.929	0.918	Almost Perfect
189	Labeler 11	238	13	0.923	0.908	Almost Perfect
190	Labeler 11	255	9	1.000	1.000	Almost Perfect
191	Labeler 11	258	15	0.933	0.922	Almost Perfect
192	Labeler 11	274	21	0.905	0.894	Strong
193	Labeler 11	279	14	0.786	0.745	Moderate
194	Labeler 11	284	8	1.000	1.000	Almost Perfect
195	Labeler 11	288	7	1.000	1.000	Almost Perfect
196	Labeler 11	291	12	1.000	1.000	Almost Perfect

Table 52 (cont.)

197	Labeler 11	104	9	1.000	1.000	Almost Perfect
198	Labeler 11	125	11	0.909	0.892	Strong
199	Labeler 11	142	13	1.000	1.000	Almost Perfect
200	Labeler 11	161	17	1.000	1.000	Almost Perfect
201	Labeler 11	194	8	1.000	1.000	Almost Perfect
202	Labeler 11	206	8	1.000	1.000	Almost Perfect
203	Labeler 11	233	3	1.000	1.000	Almost Perfect
204	Labeler 11	267	15	0.933	0.917	Almost Perfect
205	Labeler 11	294	11	0.909	0.894	Strong
206	Labeler 11	296	5	1.000	1.000	Almost Perfect
207	Labeler 12	1	17	0.941	0.930	Almost Perfect
208	Labeler 12	16	16	0.875	0.857	Strong
209	Labeler 12	24	15	1.000	1.000	Almost Perfect
210	Labeler 12	29	7	0.857	0.811	Strong
211	Labeler 12	45	12	0.917	0.902	Almost Perfect
212	Labeler 12	54	5	0.800	0.722	Moderate
213	Labeler 12	60	8	0.875	0.852	Strong
214	Labeler 12	66	12	1.000	1.000	Almost Perfect
215	Labeler 12	72	12	1.000	1.000	Almost Perfect
216	Labeler 12	80	9	0.889	0.875	Strong
217	Labeler 12	95	15	0.933	0.919	Almost Perfect
218	Labeler 12	116	22	0.955	0.944	Almost Perfect
219	Labeler 12	117	11	1.000	1.000	Almost Perfect
220	Labeler 12	126	13	0.769	0.727	Moderate
221	Labeler 12	140	11	0.818	0.790	Strong
222	Labeler 12	158	10	0.900	0.881	Strong
223	Labeler 12	159	10	0.900	0.877	Strong
224	Labeler 12	167	11	1.000	1.000	Almost Perfect
225	Labeler 12	179	15	0.667	0.629	Moderate
226	Labeler 12	183	11	0.818	0.788	Moderate
227	Labeler 12	192	11	1.000	1.000	Almost Perfect
228	Labeler 12	203	11	1.000	1.000	Almost Perfect
229	Labeler 12	224	14	1.000	1.000	Almost Perfect
230	Labeler 12	236	10	0.900	0.877	Strong
231	Labeler 12	244	18	0.944	0.936	Almost Perfect
232	Labeler 12	258	1	1.000	1.000	Almost Perfect
233	Labeler 12	92	22	0.909	0.894	Strong
234	Labeler 12	120	6	1.000	1.000	Almost Perfect
235	Labeler 12	143	11	0.818	0.788	Moderate
236	Labeler 12	152	13	0.923	0.909	Almost Perfect

Table 52 (cont.)

237	Labeler 12	193	18	0.944	0.935	Almost Perfect
238	Labeler 12	217	19	0.842	0.802	Strong
239	Labeler 12	285	4	1.000	1.000	Almost Perfect
240	Labeler 13	114	13	0.923	0.908	Almost Perfect
241	Labeler 13	155	15	0.800	0.774	Moderate
242	Labeler 13	201	6	1.000	1.000	Almost Perfect
243	Labeler 13	214	9	1.000	1.000	Almost Perfect
244	Labeler 13	241	11	1.000	1.000	Almost Perfect
245	Labeler 13	251	15	1.000	1.000	Almost Perfect
246	Labeler 13	272	19	1.000	1.000	Almost Perfect
247	Labeler 13	290	12	1.000	1.000	Almost Perfect
248	Labeler 14	2	16	0.875	0.848	Strong
249	Labeler 14	7	10	0.800	0.730	Moderate
250	Labeler 14	25	4	0.500	0.385	Minimal
251	Labeler 14	28	13	0.923	0.893	Strong
252	Labeler 14	40	15	0.733	0.704	Moderate
253	Labeler 14	46	6	0.833	0.806	Strong
254	Labeler 14	62	5	0.800	0.737	Moderate
255	Labeler 14	71	9	0.667	0.625	Moderate
256	Labeler 14	87	14	0.786	0.763	Moderate
257	Labeler 15	6	5	1.000	1.000	Almost Perfect
258	Labeler 15	13	10	0.800	0.762	Moderate
259	Labeler 15	19	9	1.000	1.000	Almost Perfect
260	Labeler 15	31	7	1.000	1.000	Almost Perfect
261	Labeler 15	51	15	0.867	0.840	Strong
262	Labeler 15	53	5	1.000	1.000	Almost Perfect
263	Labeler 15	59	18	0.889	0.868	Strong
264	Labeler 15	78	10	1.000	1.000	Almost Perfect
265	Labeler 15	84	14	1.000	1.000	Almost Perfect
266	Labeler 15	99	12	1.000	1.000	Almost Perfect
267	Labeler 15	123	6	1.000	1.000	Almost Perfect
268	Labeler 15	127	16	1.000	1.000	Almost Perfect
269	Labeler 15	141	13	1.000	1.000	Almost Perfect
270	Labeler 15	148	14	1.000	1.000	Almost Perfect
271	Labeler 15	173	13	1.000	1.000	Almost Perfect
272	Labeler 15	175	12	1.000	1.000	Almost Perfect
273	Labeler 15	185	9	1.000	1.000	Almost Perfect
274	Labeler 15	199	6	1.000	1.000	Almost Perfect
275	Labeler 15	209	3	1.000	1.000	Almost Perfect
276	Labeler 15	242	14	0.929	0.907	Almost Perfect

Table 52 (cont.)

277	Labeler 15	261	4	1.000	1.000	Almost Perfect
278	Labeler 15	271	8	1.000	1.000	Almost Perfect
279	Labeler 15	276	15	0.867	0.828	Strong
280	Labeler 15	287	16	1.000	1.000	Almost Perfect
281	Labeler 15	293	10	0.900	0.865	Strong
282	Labeler 16	20	4	1.000	1.000	Almost Perfect
283	Labeler 16	30	17	0.824	0.795	Strong
284	Labeler 16	38	13	1.000	1.000	Almost Perfect
285	Labeler 16	47	9	0.778	0.727	Moderate
286	Labeler 16	69	15	0.867	0.845	Strong
287	Labeler 16	74	18	0.944	0.936	Almost Perfect
288	Labeler 16	85	3	1.000	1.000	Almost Perfect
289	Labeler 16	103	5	1.000	1.000	Almost Perfect
290	Labeler 16	105	14	0.929	0.918	Almost Perfect
291	Labeler 16	128	12	0.583	0.508	Weak
292	Labeler 16	129	6	1.000	1.000	Almost Perfect
293	Labeler 16	156	4	1.000	1.000	Almost Perfect
294	Labeler 16	164	7	0.714	0.659	Moderate
295	Labeler 16	169	8	1.000	1.000	Almost Perfect
296	Labeler 16	184	8	0.875	0.860	Strong
297	Labeler 16	209	1	1.000	1.000	Almost Perfect
298	Labeler 16	212	9	0.889	0.859	Strong
299	Labeler 16	220	8	0.750	0.704	Moderate
300	Labeler 16	232	3	0.333	0.000	None
301	Labeler 16	246	7	1.000	1.000	Almost Perfect
302	Labeler 16	247	9	0.778	0.723	Moderate
303	Labeler 16	259	10	0.900	0.873	Strong
304	Labeler 16	260	7	0.714	0.674	Moderate
305	Labeler 16	282	7	1.000	1.000	Almost Perfect
306	Labeler 16	298	11	0.909	0.887	Strong
307	Labeler 16	299	6	1.000	1.000	Almost Perfect

Table 53 Weighted Averaged IAA Metrics for "Crossing Player" Label

#	**Labeler**	**Total Labeled**	**Weighted Averaged Raw Agreement**	**Weighted Averaged Kappa**	**Agreement**
1	Labeler 1	391	0.905	0.887	Strong
2	Labeler 2	213	0.831	0.797	Strong
3	Labeler 3	195	0.805	0.760	Moderate
4	Labeler 4	15	0.667	0.625	Moderate
5	Labeler 5	253	0.905	0.888	Strong
6	Labeler 6	134	0.910	0.893	Strong
7	Labeler 7	80	0.838	0.805	Strong
8	Labeler 8	134	0.821	0.779	Moderate
9	Labeler 9	238	0.840	0.816	Strong
10	Labeler 10	70	0.929	0.911	Almost Perfect
11	Labeler 11	404	0.941	0.930	Almost Perfect
12	Labeler 12	400	0.913	0.896	Strong
13	Labeler 13	100	0.960	0.954	Almost Perfect
14	Labeler 14	92	0.793	0.754	Moderate
15	Labeler 15	264	0.962	0.953	Almost Perfect
16	Labeler 16	221	0.873	0.847	Strong

Table 54 IAA Metrics for "Outcome of Cross - Category 1" Label

#	**Labeler**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Agreement**
1	Labeler 1	391	0.824	0.734	Moderate
2	Labeler 2	213	0.714	0.541	Weak
3	Labeler 3	195	0.595	0.420	Weak
4	Labeler 4	15	0.667	0.525	Weak
5	Labeler 5	253	0.854	0.785	Moderate
6	Labeler 6	134	0.784	0.702	Moderate
7	Labeler 7	80	0.813	0.731	Moderate
8	Labeler 8	134	0.866	0.793	Strong
9	Labeler 9	238	0.773	0.663	Moderate
10	Labeler 10	70	0.886	0.834	Strong
11	Labeler 11	404	0.876	0.820	Strong
12	Labeler 12	400	0.840	0.766	Moderate
13	Labeler 13	100	0.850	0.776	Moderate
14	Labeler 14	92	0.641	0.463	Weak
15	Labeler 15	264	0.852	0.785	Moderate
16	Labeler 16	221	0.710	0.600	Moderate

Table 55 IAA Metrics for "Outcome of Cross - Category 2" Label per "Outcome of Cross - Category 1"

#	Labeler	Labeled Data	Outcome 1	Raw Agreement	Kappa	Kappa Agreement
1	Labeler 1	138	1. Cross on target, reaches a teammate	0.862	0.678	Moderate
2	Labeler 2	78	1. Cross on target, reaches a teammate	0.821	-0.034	None
3	Labeler 3	55	1. Cross on target, reaches a teammate	0.709	0.227	Minimal
4	Labeler 4	6	1. Cross on target, reaches a teammate	0.833	0.571	Weak
5	Labeler 5	96	1. Cross on target, reaches a teammate	0.896	0.741	Moderate
6	Labeler 6	45	1. Cross on target, reaches a teammate	0.844	0.682	Moderate
7	Labeler 7	25	1. Cross on target, reaches a teammate	0.840	0.438	Weak
8	Labeler 8	54	1. Cross on target, reaches a teammate	0.815	0.484	Weak
9	Labeler 9	80	1. Cross on target, reaches a teammate	0.788	0.440	Weak
10	Labeler 10	29	1. Cross on target, reaches a teammate	0.931	0.758	Moderate
11	Labeler 11	145	1. Cross on target, reaches a teammate	0.910	0.759	Moderate
12	Labeler 12	136	1. Cross on target, reaches a teammate	0.882	0.669	Moderate
13	Labeler 13	32	1. Cross on target, reaches a teammate	0.875	0.667	Moderate
14	Labeler 14	32	1. Cross on target, reaches a teammate	0.750	0.291	Minimal
15	Labeler 15	98	1. Cross on target, reaches a teammate	0.888	0.649	Moderate
16	Labeler 16	66	1. Cross on target, reaches a teammate	0.848	0.387	Minimal
17	Labeler 1	16	2. Cross off target, not reaches a teammate	0.563	0.253	Minimal
18	Labeler 2	7	2. Cross off target, not reaches a teammate	0.286	0.000	None
19	Labeler 3	7	2. Cross off target, not reaches a teammate	0.714	0.533	Weak

Table 55 (cont.)

20	Labeler 4	2	2. Cross off target, not reaches a teammate	0.500	0.000	None
21	Labeler 5	20	2. Cross off target, not reaches a teammate	0.500	0.194	None
22	Labeler 6	12	2. Cross off target, not reaches a teammate	0.583	0.286	Minimal
23	Labeler 7	10	2. Cross off target, not reaches a teammate	0.300	0.054	None
24	Labeler 8	7	2. Cross off target, not reaches a teammate	0.571	0.276	Minimal
25	Labeler 9	15	2. Cross off target, not reaches a teammate	0.733	0.623	Moderate
26	Labeler 10	4	2. Cross off target, not reaches a teammate	0.250	0.143	None
27	Labeler 11	36	2. Cross off target, not reaches a teammate	0.722	0.556	Weak
28	Labeler 12	37	2. Cross off target, not reaches a teammate	0.865	0.724	Moderate
29	Labeler 13	10	2. Cross off target, not reaches a teammate	0.600	0.444	Weak
30	Labeler 14	4	2. Cross off target, not reaches a teammate	1.000	1.000	Almost Perfect
31	Labeler 15	16	2. Cross off target, not reaches a teammate	0.688	0.527	Weak
32	Labeler 16	16	2. Cross off target, not reaches a teammate	0.688	0.532	Weak
33	Labeler 1	128	3. Defender Interception	0.750	0.420	Weak
34	Labeler 2	59	3. Defender Interception	0.424	0.027	None
35	Labeler 3	45	3. Defender Interception	0.244	0.114	None
36	Labeler 4	0	3. Defender Interception	0.000	0.000	No Record
37	Labeler 5	76	3. Defender Interception	0.724	0.420	Weak
38	Labeler 6	30	3. Defender Interception	0.600	0.271	Minimal
39	Labeler 7	25	3. Defender Interception	0.640	0.202	Minimal
40	Labeler 8	41	3. Defender Interception	0.439	-0.045	None
41	Labeler 9	70	3. Defender Interception	0.486	0.235	Minimal

Table 55 (cont.)

42	Labeler 10	18	3. Defender Interception	0.778	-0.043	None
43	Labeler 11	126	3. Defender Interception	0.722	0.537	Weak
44	Labeler 12	130	3. Defender Interception	0.669	0.429	Weak
45	Labeler 13	35	3. Defender Interception	0.771	0.572	Weak
46	Labeler 14	19	3. Defender Interception	0.579	0.101	None
47	Labeler 15	79	3. Defender Interception	0.684	0.349	Minimal
48	Labeler 16	50	3. Defender Interception	0.340	0.190	None
49	Labeler 1	32	4. Goalkeeper Interception	0.844	0.624	Moderate
50	Labeler 2	7	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
51	Labeler 3	8	4. Goalkeeper Interception	0.500	-0.143	None
52	Labeler 4	2	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
53	Labeler 5	24	4. Goalkeeper Interception	0.917	0.795	Strong
54	Labeler 6	17	4. Goalkeeper Interception	0.941	0.821	Strong
55	Labeler 7	5	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
56	Labeler 8	14	4. Goalkeeper Interception	0.857	0.692	Moderate
57	Labeler 9	16	4. Goalkeeper Interception	0.875	0.704	Moderate
58	Labeler 10	8	4. Goalkeeper Interception	0.750	0.429	Weak
59	Labeler 11	40	4. Goalkeeper Interception	0.975	0.946	Almost Perfect
60	Labeler 12	27	4. Goalkeeper Interception	0.926	0.842	Strong
61	Labeler 13	8	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
62	Labeler 14	4	4. Goalkeeper Interception	0.750	0.556	Weak
63	Labeler 15	24	4. Goalkeeper Interception	0.958	0.922	Almost Perfect

Table 55 (cont.)

64	Labeler 16	23	4. Goalkeeper Interception	0.783	0.350	Minimal
65	Labeler 1	6	5. Referee Interception	0.667	0.538	Weak
66	Labeler 2	1	5. Referee Interception	1.000	1.000	Almost Perfect
67	Labeler 3	1	5. Referee Interception	1.000	1.000	Almost Perfect
68	Labeler 4	0	5. Referee Interception	0.000	0.000	No Record
69	Labeler 5	0	5. Referee Interception	0.000	0.000	No Record
70	Labeler 6	1	5. Referee Interception	1.000	1.000	Almost Perfect
71	Labeler 7	0	5. Referee Interception	0.000	0.000	No Record
72	Labeler 8	0	5. Referee Interception	0.000	0.000	No Record
73	Labeler 9	3	5. Referee Interception	1.000	1.000	Almost Perfect
74	Labeler 10	3	5. Referee Interception	0.667	0.500	Weak
75	Labeler 11	7	5. Referee Interception	1.000	1.000	Almost Perfect
76	Labeler 12	6	5. Referee Interception	0.833	0.769	Moderate
77	Labeler 13	0	5. Referee Interception	0.000	0.000	No Record
78	Labeler 14	0	5. Referee Interception	0.000	0.000	No Record
79	Labeler 15	7	5. Referee Interception	1.000	1.000	Almost Perfect
80	Labeler 16	2	5. Referee Interception	1.000	1.000	Almost Perfect

Table 56 IAA Metrics for “Outcome of Cross - Category 3” Label per “Outcome of Cross - Category 2”

#	Labeler	Labeled Data	Outcome 2	Raw Agree ment	Kappa	Kappa Agreement
1	Labeler 1	94	1. Cross Receiver Shoots/Heads	0.904	0.864	Strong
2	Labeler 2	64	1. Cross Receiver Shoots/Heads	0.875	0.817	Strong
3	Labeler 3	36	1. Cross Receiver Shoots/Heads	0.611	0.514	Weak
4	Labeler 4	4	1. Cross Receiver Shoots/Heads	0.750	0.667	Moderate
5	Labeler 5	68	1. Cross Receiver Shoots/Heads	0.897	0.857	Strong
6	Labeler 6	27	1. Cross Receiver Shoots/Heads	0.815	0.746	Moderate
7	Labeler 7	19	1. Cross Receiver Shoots/Heads	0.947	0.928	Almost Perfect
8	Labeler 8	38	1. Cross Receiver Shoots/Heads	0.737	0.666	Moderate
9	Labeler 9	54	1. Cross Receiver Shoots/Heads	0.870	0.818	Strong
10	Labeler 10	23	1. Cross Receiver Shoots/Heads	0.913	0.884	Strong
11	Labeler 11	107	1. Cross Receiver Shoots/Heads	0.944	0.922	Almost Perfect
12	Labeler 12	99	1. Cross Receiver Shoots/Heads	0.960	0.943	Almost Perfect
13	Labeler 13	23	1. Cross Receiver Shoots/Heads	0.957	0.933	Almost Perfect
14	Labeler 14	22	1. Cross Receiver Shoots/Heads	0.818	0.747	Moderate
15	Labeler 15	74	1. Cross Receiver Shoots/Heads	0.973	0.961	Almost Perfect
16	Labeler 16	52	1. Cross Receiver Shoots/Heads	0.904	0.868	Strong
17	Labeler 1	15	2. Cross Receiver Passes/ Crosses to a teammate	0.867	0.795	Strong
18	Labeler 2	0	2. Cross Receiver Passes/ Crosses to a teammate	0.000	0.000	No Record
19	Labeler 3	3	2. Cross Receiver Passes/ Crosses to a teammate	0.333	0.143	None
20	Labeler 4	1	2. Cross Receiver Passes/ Crosses to a teammate	0.000	0.000	None
21	Labeler 5	9	2. Cross Receiver Passes/ Crosses to a teammate	0.667	0.578	Weak
22	Labeler 6	7	2. Cross Receiver Passes/ Crosses to a teammate	0.571	0.447	Weak
23	Labeler 7	2	2. Cross Receiver Passes/ Crosses to a teammate	0.500	0.333	Minimal
24	Labeler 8	4	2. Cross Receiver Passes/ Crosses to a teammate	0.500	0.333	Minimal

Table 56 (cont.)

25	Labeler 9	7	2. Cross Receiver Passes/ Crosses to a teammate	1.000	1.000	Almost Perfect
26	Labeler 10	4	2. Cross Receiver Passes/ Crosses to a teammate	1.000	1.000	Almost Perfect
27	Labeler 11	18	2. Cross Receiver Passes/ Crosses to a teammate	0.889	0.842	Strong
28	Labeler 12	16	2. Cross Receiver Passes/ Crosses to a teammate	0.563	0.467	Weak
29	Labeler 13	5	2. Cross Receiver Passes/ Crosses to a teammate	0.600	0.333	Minimal
30	Labeler 14	1	2. Cross Receiver Passes/ Crosses to a teammate	1.000	1.000	Almost Perfect
31	Labeler 15	11	2. Cross Receiver Passes/ Crosses to a teammate	0.727	0.560	Weak
32	Labeler 16	3	2. Cross Receiver Passes/ Crosses to a teammate	1.000	1.000	Almost Perfect
33	Labeler 1	4	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	None
34	Labeler 2	1	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	1.000	1.000	Almost Perfect
35	Labeler 3	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	No Record
36	Labeler 4	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	No Record
37	Labeler 5	5	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.600	0.444	Weak
38	Labeler 6	3	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	1.000	1.000	Almost Perfect
39	Labeler 7	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	No Record
40	Labeler 8	3	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.333	0.143	None
41	Labeler 9	3	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	-0.125	None

Table 56 (cont.)

42	Labeler 10	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	No Record
43	Labeler 11	11	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.636	0.551	Weak
44	Labeler 12	7	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.429	0.282	Minimal
45	Labeler 13	3	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.333	0.250	Minimal
46	Labeler 14	1	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	None
47	Labeler 15	7	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.571	0.382	Minimal
48	Labeler 16	2	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	1.000	1.000	Almost Perfect
49	Labeler 1	80	3. Defensive clearance to upcoming cross with kick, head	0.475	0.349	Minimal
50	Labeler 2	23	3. Defensive clearance to upcoming cross with kick, head	0.478	0.338	Minimal
51	Labeler 3	5	3. Defensive clearance to upcoming cross with kick, head	0.200	0.167	None
52	Labeler 4	0	3. Defensive clearance to upcoming cross with kick, head	0.000	0.000	No Record
53	Labeler 5	45	3. Defensive clearance to upcoming cross with kick, head	0.378	0.229	Minimal
54	Labeler 6	14	3. Defensive clearance to upcoming cross with kick, head	0.357	0.171	None
55	Labeler 7	15	3. Defensive clearance to upcoming cross with kick, head	0.667	0.519	Weak
56	Labeler 8	17	3. Defensive clearance to upcoming cross with kick, head	0.588	0.498	Weak
57	Labeler 9	18	3. Defensive clearance to upcoming cross with kick, head	0.889	0.786	Moderate
58	Labeler 10	14	3. Defensive clearance to upcoming cross with kick, head	0.429	0.200	None
59	Labeler 11	60	3. Defensive clearance to upcoming cross with kick, head	0.650	0.504	Weak

Table 56 (cont.)

60	Labeler 12	61	3. Defensive clearance to upcoming cross with kick, head	0.525	0.393	Weak
61	Labeler 13	19	3. Defensive clearance to upcoming cross with kick, head	0.632	0.502	Weak
62	Labeler 14	10	3. Defensive clearance to upcoming cross with kick, head	0.100	0.000	None
63	Labeler 15	47	3. Defensive clearance to upcoming cross with kick, head	0.532	0.410	Weak
64	Labeler 16	4	3. Defensive clearance to upcoming cross with kick, head	0.500	0.333	Minimal

Table 57 IAA Metrics for “Quality of Cross” Label

#	**Labeler**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Kappa Agreement**
0	Labeler 1	387	0.530	0.626	Moderate
1	Labeler 2	211	0.403	0.251	Minimal
2	Labeler 3	168	0.399	0.213	Minimal
3	Labeler 4	15	0.400	0.545	Weak
4	Labeler 5	244	0.557	0.539	Weak
5	Labeler 6	122	0.566	0.718	Moderate
6	Labeler 7	71	0.592	0.732	Moderate
7	Labeler 8	132	0.341	0.484	Weak
8	Labeler 9	237	0.595	0.611	Moderate
9	Labeler 10	70	0.343	0.610	Moderate
10	Labeler 11	402	0.575	0.749	Moderate
11	Labeler 12	398	0.598	0.554	Weak
12	Labeler 13	100	0.540	0.688	Moderate
13	Labeler 14	92	0.587	0.462	Weak
14	Labeler 15	263	0.639	0.662	Moderate
15	Labeler 16	218	0.450	0.580	Weak

Table 58 IAA Metrics for “Is There a Second Cross” Label

#	**Labeler**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Kappa Agreement**
1	Labeler 1	390	0.956	0.668	Moderate
2	Labeler 2	213	0.925	0.345	Minimal
3	Labeler 3	195	0.944	0.244	Minimal
4	Labeler 4	15	1.000	N/A	N/A
5	Labeler 5	253	0.960	0.623	Moderate
6	Labeler 6	134	0.948	-0.026	None
7	Labeler 7	74	0.986	0.793	Strong
8	Labeler 8	134	0.948	0.560	Weak
9	Labeler 9	238	0.924	0.460	Weak
10	Labeler 10	70	0.971	0.653	Moderate
11	Labeler 11	404	0.963	0.686	Moderate
12	Labeler 12	399	0.962	0.553	Weak
13	Labeler 13	100	0.960	0.693	Moderate
14	Labeler 14	92	0.924	0.340	Minimal
15	Labeler 15	264	0.981	0.773	Moderate
16	Labeler 16	218	0.963	0.413	Weak

Table 59 IAA Metrics for “Second Cross - Crossing Player” Label per match

#	**Labeler**	**Match No**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Agreement**
1	Labeler 1	61	1	1.000	1.000	Almost Perfect
2	Labeler 1	82	2	1.000	1.000	Almost Perfect
3	Labeler 1	86	1	1.000	1.000	Almost Perfect
4	Labeler 1	102	3	1.000	1.000	Almost Perfect
5	Labeler 1	131	1	1.000	1.000	Almost Perfect
6	Labeler 1	178	2	0.500	0.000	None
7	Labeler 1	204	1	1.000	1.000	Almost Perfect
8	Labeler 1	213	2	1.000	1.000	Almost Perfect
9	Labeler 1	234	3	1.000	1.000	Almost Perfect
10	Labeler 1	249	3	1.000	1.000	Almost Perfect
11	Labeler 2	21	1	1.000	1.000	Almost Perfect
12	Labeler 2	151	1	0.000	0.000	None
13	Labeler 2	280	1	1.000	1.000	Almost Perfect
14	Labeler 2	283	1	1.000	1.000	Almost Perfect

Table 59 (cont.)

15	Labeler 2	303	1	1.000	1.000	Almost Perfect
16	Labeler 3	55	1	1.000	1.000	Almost Perfect
17	Labeler 3	231	1	1.000	1.000	Almost Perfect
18	Labeler 5	18	1	1.000	1.000	Almost Perfect
19	Labeler 5	44	1	1.000	1.000	Almost Perfect
20	Labeler 5	101	1	1.000	1.000	Almost Perfect
21	Labeler 5	137	1	1.000	1.000	Almost Perfect
22	Labeler 5	196	1	1.000	1.000	Almost Perfect
23	Labeler 5	265	1	1.000	1.000	Almost Perfect
24	Labeler 5	266	2	0.500	0.333	Minimal
25	Labeler 5	268	1	1.000	1.000	Almost Perfect
26	Labeler 7	94	2	1.000	1.000	Almost Perfect
27	Labeler 8	27	2	0.500	0.333	Minimal
28	Labeler 8	139	1	1.000	1.000	Almost Perfect
29	Labeler 8	208	1	0.000	0.000	None
30	Labeler 8	264	1	1.000	1.000	Almost Perfect
31	Labeler 9	4	1	1.000	1.000	Almost Perfect
32	Labeler 9	39	1	1.000	1.000	Almost Perfect
33	Labeler 9	67	1	0.000	0.000	None
34	Labeler 9	90	1	0.000	0.000	None
35	Labeler 9	97	1	1.000	1.000	Almost Perfect
36	Labeler 9	166	1	1.000	1.000	Almost Perfect
37	Labeler 9	227	1	0.000	0.000	None
38	Labeler 9	245	1	0.000	0.000	None
39	Labeler 9	301	1	1.000	1.000	Almost Perfect
40	Labeler 10	180	1	1.000	1.000	Almost Perfect
41	Labeler 10	270	1	1.000	1.000	Almost Perfect
42	Labeler 11	174	1	1.000	1.000	Almost Perfect
43	Labeler 11	186	1	1.000	1.000	Almost Perfect
44	Labeler 11	215	1	0.000	0.000	None
45	Labeler 11	222	2	0.500	0.333	Minimal
46	Labeler 11	238	1	1.000	1.000	Almost Perfect
47	Labeler 11	258	4	0.750	0.636	Moderate
48	Labeler 11	274	3	1.000	1.000	Almost Perfect
49	Labeler 11	291	1	1.000	1.000	Almost Perfect
50	Labeler 11	125	1	1.000	1.000	Almost Perfect
51	Labeler 11	142	1	1.000	1.000	Almost Perfect
52	Labeler 11	194	1	1.000	1.000	Almost Perfect
53	Labeler 11	206	1	1.000	1.000	Almost Perfect
54	Labeler 12	1	2	1.000	1.000	Almost Perfect

Table 59 (cont.)

55	Labeler 12	16	1	1.000	1.000	Almost Perfect
56	Labeler 12	80	1	0.000	0.000	None
57	Labeler 12	95	1	1.000	1.000	Almost Perfect
58	Labeler 12	116	1	1.000	1.000	Almost Perfect
59	Labeler 12	126	1	1.000	1.000	Almost Perfect
60	Labeler 12	193	1	1.000	1.000	Almost Perfect
61	Labeler 12	217	2	1.000	1.000	Almost Perfect
62	Labeler 13	214	1	0.000	0.000	None
63	Labeler 13	241	1	1.000	1.000	Almost Perfect
64	Labeler 13	251	2	1.000	1.000	Almost Perfect
65	Labeler 13	272	1	1.000	1.000	Almost Perfect
66	Labeler 14	40	1	1.000	1.000	Almost Perfect
67	Labeler 14	87	1	1.000	1.000	Almost Perfect
68	Labeler 15	6	1	1.000	1.000	Almost Perfect
69	Labeler 15	13	1	1.000	1.000	Almost Perfect
70	Labeler 15	51	2	1.000	1.000	Almost Perfect
71	Labeler 15	53	2	1.000	1.000	Almost Perfect
72	Labeler 15	141	1	1.000	1.000	Almost Perfect
73	Labeler 15	173	1	1.000	1.000	Almost Perfect
74	Labeler 15	276	1	1.000	1.000	Almost Perfect
75	Labeler 16	38	1	1.000	1.000	Almost Perfect
76	Labeler 16	85	1	1.000	1.000	Almost Perfect
77	Labeler 16	298	1	1.000	1.000	Almost Perfect

Table 60 Weighted Averaged IAA Metrics for “Second Cross - Crossing Player” Label

#	Labeler	Total Labeled	Raw Agreement	Weighted Kappa	Agreement
1	Labeler 1	19	0.947	0.895	Strong
2	Labeler 2	5	0.800	0.800	Strong
3	Labeler 3	2	1.000	1.000	Almost Perfect
4	Labeler 4	0	No record	No record	No record
5	Labeler 5	9	0.889	0.852	Strong
6	Labeler 6	0	No record	No record	No record
7	Labeler 7	2	1.000	1.000	Almost Perfect
8	Labeler 8	5	0.600	0.533	Almost Perfect
9	Labeler 9	9	0.556	0.556	Almost Perfect
10	Labeler 10	2	1.000	1.000	Almost Perfect
11	Labeler 11	18	0.833	0.790	Moderate
12	Labeler 12	10	0.900	0.900	Strong
13	Labeler 13	5	0.800	0.800	Strong
14	Labeler 14	2	1.000	1.000	Almost Perfect
15	Labeler 15	9	1.000	1.000	Almost Perfect
16	Labeler 16	3	1.000	1.000	Almost Perfect

Table 61 IAA Metrics for “Outcome of Second Cross - Category 1” Label

#	Labeler	Labeled Data	Raw Agreement	Kappa	Agreement
1	Labeler 1	19	0.895	0.866	Strong
2	Labeler 2	5	0.600	0.333	Minimal
3	Labeler 3	2	0.500	0.333	Minimal
4	Labeler 4	0	No record	No record	No record
5	Labeler 5	9	0.667	0.509	Weak
6	Labeler 6	0	No record	No record	No record
7	Labeler 7	2	1.000	1.000	Almost Perfect
8	Labeler 8	5	1.000	1.000	Almost Perfect
9	Labeler 9	9	0.778	0.571	Weak
10	Labeler 10	2	0.000	-1.000	None
11	Labeler 11	18	0.944	0.918	Almost Perfect
12	Labeler 12	10	0.800	0.710	Moderate
13	Labeler 13	5	0.800	0.643	Moderate
14	Labeler 14	2	0.000	-0.333	None
15	Labeler 15	9	0.889	0.830	Strong
16	Labeler 16	3	1.000	1.000	Almost Perfect

Table 62 IAA Metrics for "Outcome of Second Cross - Category 2" Label per "Outcome of Second Cross - Category 1"

#	**Labeler**	**Labeled Data**	**Outcome 1**	**Raw Agreement**	**Kappa**	**Kappa Agreement**
1	Labeler 1	5	1. Cross on target, reaches a teammate	0.800	0.583	Weak
2	Labeler 2	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
3	Labeler 3	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
4	Labeler 4	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
5	Labeler 5	1	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
6	Labeler 6	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
7	Labeler 7	1	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
8	Labeler 8	1	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
9	Labeler 9	5	1. Cross on target, reaches a teammate	0.800	0.000	None
10	Labeler 10	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
11	Labeler 11	6	1. Cross on target, reaches a teammate	0.833	0.000	None
12	Labeler 12	2	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
13	Labeler 13	1	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
14	Labeler 14	0	1. Cross on target, reaches a teammate	0.000	0.000	No Record
15	Labeler 15	5	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
16	Labeler 16	2	1. Cross on target, reaches a teammate	1.000	1.000	Almost Perfect
17	Labeler 1	2	2. Cross off target, not reaches a teammate	1.000	1.000	Almost Perfect
18	Labeler 2	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
19	Labeler 3	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record

Table 62 (cont.)

20	Labeler 4	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
21	Labeler 5	2	2. Cross off target, not reaches a teammate	0.500	0.333	Minimal
22	Labeler 6	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
23	Labeler 7	1	2. Cross off target, not reaches a teammate	0.000	0.000	None
24	Labeler 8	1	2. Cross off target, not reaches a teammate	1.000	1.000	Almost Perfect
25	Labeler 9	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
26	Labeler 10	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
27	Labeler 11	2	2. Cross off target, not reaches a teammate	0.000	0.000	None
28	Labeler 12	1	2. Cross off target, not reaches a teammate	1.000	1.000	Almost Perfect
29	Labeler 13	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
30	Labeler 14	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
31	Labeler 15	1	2. Cross off target, not reaches a teammate	0.000	0.000	None
32	Labeler 16	0	2. Cross off target, not reaches a teammate	0.000	0.000	No Record
33	Labeler 1	5	3. Defender Interception	0.800	0.688	Moderate
34	Labeler 2	2	3. Defender Interception	0.000	-0.333	None
35	Labeler 3	0	3. Defender Interception	0.000	0.000	No Record
36	Labeler 4	0	3. Defender Interception	0.000	0.000	No Record
37	Labeler 5	3	3. Defender Interception	0.667	0.500	Weak
38	Labeler 6	0	3. Defender Interception	0.000	0.000	No Record
39	Labeler 7	0	3. Defender Interception	0.000	0.000	No Record
40	Labeler 8	0	3. Defender Interception	0.000	0.000	No Record
41	Labeler 9	1	3. Defender Interception	1.000	1.000	Almost Perfect
42	Labeler 10	0	3. Defender Interception	0.000	0.000	No Record
43	Labeler 11	7	3. Defender Interception	0.571	0.125	None
44	Labeler 12	4	3. Defender Interception	0.500	0.200	None
45	Labeler 13	3	3. Defender Interception	0.333	0.000	None
46	Labeler 14	0	3. Defender Interception	0.000	0.000	No Record
47	Labeler 15	0	3. Defender Interception	0.000	0.000	No Record

Table 62 (cont.)

48	Labeler 16	1	3. Defender Interception	0.000	0.000	None
49	Labeler 1	3	4. Goalkeeper Interception	0.667	0.400	Weak
50	Labeler 2	1	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
51	Labeler 3	1	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
52	Labeler 4	0	4. Goalkeeper Interception	0.000	0.000	No Record
53	Labeler 5	0	4. Goalkeeper Interception	0.000	0.000	No Record
54	Labeler 6	0	4. Goalkeeper Interception	0.000	0.000	No Record
55	Labeler 7	0	4. Goalkeeper Interception	0.000	0.000	No Record
56	Labeler 8	3	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
57	Labeler 9	1	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
58	Labeler 10	0	4. Goalkeeper Interception	0.000	0.000	No Record
59	Labeler 11	1	4. Goalkeeper Interception	1.000	1.000	Almost Perfect
60	Labeler 12	0	4. Goalkeeper Interception	0.000	0.000	No Record
61	Labeler 13	0	4. Goalkeeper Interception	0.000	0.000	No Record
62	Labeler 14	0	4. Goalkeeper Interception	0.000	0.000	No Record
63	Labeler 15	0	4. Goalkeeper Interception	0.000	0.000	No Record
64	Labeler 16	0	4. Goalkeeper Interception	0.000	0.000	No Record
65	Labeler 1	1	5. Referee Interception	1.000	1.000	Almost Perfect
66	Labeler 2	0	5. Referee Interception	0.000	0.000	No Record
67	Labeler 3	0	5. Referee Interception	0.000	0.000	No Record
68	Labeler 4	0	5. Referee Interception	0.000	0.000	No Record
69	Labeler 5	0	5. Referee Interception	0.000	0.000	No Record
70	Labeler 6	0	5. Referee Interception	0.000	0.000	No Record
71	Labeler 7	0	5. Referee Interception	0.000	0.000	No Record
72	Labeler 8	0	5. Referee Interception	0.000	0.000	No Record
73	Labeler 9	0	5. Referee Interception	0.000	0.000	No Record
74	Labeler 10	0	5. Referee Interception	0.000	0.000	No Record
75	Labeler 11	1	5. Referee Interception	1.000	1.000	Almost Perfect
76	Labeler 12	1	5. Referee Interception	1.000	1.000	Almost Perfect
77	Labeler 13	0	5. Referee Interception	0.000	0.000	No Record
78	Labeler 14	0	5. Referee Interception	0.000	0.000	No Record
79	Labeler 15	1	5. Referee Interception	1.000	1.000	Almost Perfect

Table 62 (cont.)

80	Labeler 16	0	5. Referee Interception	0.000	0.000	No Record

Table 63 IAA Metrics for “Outcome of Second Cross - Category 3” Label per “Outcome of Second Cross - Category 2”

#	**Labeler**	**Labeled Data**	**Outcome 2**	**Raw Agreement**	**Kappa**	**Kappa Agreement**
1	Labeler 1	3	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
2	Labeler 2	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
3	Labeler 3	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
4	Labeler 4	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
5	Labeler 5	1	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
6	Labeler 6	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
7	Labeler 7	1	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
8	Labeler 8	1	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
9	Labeler 9	4	1. Cross Receiver Shoots/Heads	0.750	0.600	Moderate
10	Labeler 10	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
11	Labeler 11	5	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
12	Labeler 12	2	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
13	Labeler 13	1	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
14	Labeler 14	0	1. Cross Receiver Shoots/Heads	No record	No record	No record
15	Labeler 15	4	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
16	Labeler 16	2	1. Cross Receiver Shoots/Heads	1.000	1.000	Almost Perfect
17	Labeler 1	1	2. Cross Receiver Passes/Crosses to a teammate	1.000	1.000	Almost Perfect

Table 63 (cont.)

18	Labeler 2	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
19	Labeler 3	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
20	Labeler 4	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
21	Labeler 5	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
22	Labeler 6	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
23	Labeler 7	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
24	Labeler 8	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
25	Labeler 9	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
26	Labeler 10	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
27	Labeler 11	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
28	Labeler 12	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
29	Labeler 13	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
30	Labeler 14	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No record
31	Labeler 15	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
32	Labeler 16	0	2. Cross Receiver Passes/Crosses to a teammate	No record	No record	No Record
33	Labeler 1	2	3. Defensive clearance to upcoming cross with kick, head	0.500	0.000	None
34	Labeler 2	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
35	Labeler 3	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
36	Labeler 4	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
37	Labeler 5	1	3. Defensive clearance to upcoming cross with kick, head	1.000	1.000	Almost Perfect
38	Labeler 6	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
39	Labeler 7	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record

Table 63 (cont.)

40	Labeler 8	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
41	Labeler 9	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
42	Labeler 10	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
43	Labeler 11	4	3. Defensive clearance to upcoming cross with kick, head	0.750	0.556	Weak
44	Labeler 12	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
45	Labeler 13	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
46	Labeler 14	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No record
47	Labeler 15	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
48	Labeler 16	0	3. Defensive clearance to upcoming cross with kick, head	No record	No record	No Record
49	Labeler 1	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
50	Labeler 2	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
51	Labeler 3	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
52	Labeler 4	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
53	Labeler 5	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
54	Labeler 6	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
55	Labeler 7	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
56	Labeler 8	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record

Table 63 (cont.)

57	Labeler 9	1	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	0.000	0.000	None
58	Labeler 10	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
59	Labeler 11	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
60	Labeler 12	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
61	Labeler 13	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
62	Labeler 14	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No record
63	Labeler 15	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record
64	Labeler 16	0	2. Goalkeeper clearance to the ball from cross with punch, fist, slap, touch	No record	No record	No Record

Table 64 IAA Metrics for “Quality of Second Cross” Label

#	**Labeler**	**Labeled Data**	**Raw Agreement**	**Kappa**	**Kappa Agreement**
1	Labeler 1	18	0.444	0.485	Weak
2	Labeler 2	5	0.600	0.000	None
3	Labeler 3	2	0.000	0.000	None
4	Labeler 4	0	No record	No record	No record
5	Labeler 5	8	0.750	0.500	Weak
6	Labeler 6	0	No record	No record	No record
7	Labeler 7	2	0.500	0.667	Moderate
8	Labeler 8	5	0.400	0.000	None
9	Labeler 9	9	0.556	0.479	Weak
10	Labeler 10	2	0.000	-0.667	None
11	Labeler 11	18	0.611	0.807	Strong
12	Labeler 12	10	0.600	0.714	Moderate
13	Labeler 13	5	0.800	0.000	None
14	Labeler 14	2	0.500	0.000	None
15	Labeler 15	9	0.444	0.429	Weak
16	Labeler 16	3	0.667	0.000	None