NAMED ENTITY RECOGNITION AND EXPLAINABILITY ANALYSIS ON TURKISH SPORTS NEWS TEXTS

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

BY

YÜKSEL PELIN KILIÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

DECEMBER 2023

Approval of the thesis:

NAMED ENTITY RECOGNITION AND EXPLAINABILITY ANALYSIS ON TURKISH SPORTS NEWS TEXTS

submitted by **YÜKSEL PELIN KILIÇ** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

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

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _____

Prof. Dr. Pınar Karagöz
Supervisor, **Computer Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Fehime Nihan Çiçekli
Computer Engineering, METU _____

Prof. Dr. Pınar Karagöz
Computer Engineering, METU _____

Assoc. Prof. Dr. Hacer Yalım Keleş
Computer Engineering, Hacettepe University _____

Date:06.12.2023

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

Name, Surname: Yüksel Pelin Kılıç

Signature :

ABSTRACT

NAMED ENTITY RECOGNITION AND EXPLAINABILITY ANALYSIS ON TURKISH SPORTS NEWS TEXTS

Kılıç, Yüksel Pelin
M.S., Department of Computer Engineering
Supervisor: Prof. Dr. Pınar Karagöz

December 2023, 229 pages

In Natural Language Processing (NLP) and Information Extraction, Named Entity Recognition (NER) is a significant challenge. NER involves identifying entities like Person, Location, and Organization from text. While NER is well-researched in English and Chinese, Turkish NER lags, especially in domain-specific areas like sports.

The sports industry has seen a remarkable transformation with the convergence of sports and technology, impacting performance enhancement, fan engagement, and management. There is an untapped potential in extracting qualitative insights from textual data, offering a deeper understanding of the dynamics between athletes, teams, and supporters.

One key area needing further exploration is applying deep learning techniques to Turkish NER, particularly in comparison with traditional methods. Additionally, there is a lack of research on the interpretability and explainability of transformer-based models in this context.

This study introduces domain-specific Turkish NER data sets, mainly those relevant to sports, to evaluate the effectiveness of transformer-based models in Turkish NER.

A significant aspect of this research is comparing these models and analyzing how different annotation formats impact the results. The effects of named entity distribution on model performance are investigated through cross-validation techniques.

Another crucial component of this study is focusing on interpretability. By employing interpretability methods, we aim to uncover the rationale and mechanisms behind the model predictions. This aspect is crucial in understanding how these models function and make decisions, a relatively under-explored area in Turkish NER.

This research contributes to NLP and Information Extraction and has implications for enriching sports research and management, providing new insights into the interaction between sports and technology.

Keywords: Named Entity Recognition, Sports Domain, Google BERT, LIME, SHAP, Sequence Tagging, Token Classification

ÖZ

TÜRKÇE SPOR HABERLERİ METİNLERİNDE VARLIK İSMİ TANIMA VE YORUMLANABİLİRLİK ANALİZİ

Kılıç, Yüksel Pelin
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Pınar Karagöz

Aralık 2023 , 229 sayfa

Varlık İsmi Tanıma (VİT) metinlerden varlık isimlerini saptamak ve belirlenmiş etiketlere göre sınıflandırmayı amaçlayan doğal dil işleme ve bilgi çıkarımı problemidir. Özellikle spor gibi alana özel VİT çalışmaları, İngilizce, Çince gibi yaygın kullanılan dillere kıyasla Türkçe metinlerde oldukça kısıtlı sayıdadır.

Spor dünyasında teknolojinin kullanımının yaygınlaşmasıyla beraber spor yönetimi, sporcu performansı artırılması, sporcu-taraftar arasındaki ilişkinin güçlendirilmesi gibi alanlarda çalışmalar yapılmaya başlanmıştır. Spora yapılan yatırımların artmasıyla birlikte, sporda finans, pazarlama ve sporcu psikolojisi gibi alanlardaki çalışmalar artmıştır. Sporcu, takım ve taraftar arasındaki ilişkiyi daha iyi anlayabilmek için niteliksel verilerden bilgi çıkarma alanında bir boşluk vardır.

Geleneksel yöntemlerle karşılaştırıldığında derin öğrenme modellerinin Türkçe VİT alanında uygulanmasında daha fazla araştırmaya ihtiyaç duyulmaktadır. Ayrıca, bu modellerin yorumlanabilirliği ve açıklanabilirliği üzerine daha az çalışma yapılmıştır.

Bu çalışmada spor metinlerinden oluşturulmuş yeni veri kümeleri sunulmuş ve deney-

lerde kullanılmıştır. Bu çalışmada farklı derin öğrenme modellerinin performansının karşılaştırılması ve farklı etiketleme şemalarının etkisi incelenmiştir. Ayrıca, varlık isimlerinin dağılımının performansa etkisi de çapraz doğrulama yöntemiyle incelenmiştir.

Çalışmanın diğer odaklarından birisi de modellerin yorumlanabilirliğidir. Yorumlanabilirlik yöntemleri uygulanarak, modellerin tahminlerinin arkasındaki mantığı anlayabilmek amaçlanmıştır. Modellerin nasıl karar verdiğinin anlaşılması Türkçe VİT alanında henüz çok çalışılmamıştır.

Bu araştırma, sadece doğal dil işleme ve bilgi çıkarma alanına katkıda bulunmakla kalmayıp, spor araştırmaları ve yönetim uygulamalarını zenginleştirme ve spor ile teknoloji arasındaki etkileşime yeni içgörüler sağlama potansiyeline de sahiptir.

Anahtar Kelimeler: Varlık İsmi Tanıma, Google BERT, LIME, SHAP, Sıra Etiketleme, Belirteç Sınıflandırılması

To my beloved family

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Pınar Karagöz. Her expertise, invaluable support, and patience greatly enriched my journey through graduate school.

I would also like to thank my thesis committee members, Prof. Dr. Fehime Nihan Çiçekli and Assoc. Prof. Dr. Hacer Yalım Keleş, for their insightful comments and suggestions, which were immensely helpful in enhancing the quality of my thesis.

A special word of thanks goes to my late father, Önder Kılıç, whose memory continues to inspire me, and to my mother, Aysun Kılıç, and my lovely sister, Gizem Kılıç, for their unconditional love, support, and encouragement. They have always been my strength and my safe haven.

I am also thankful to Duygu Dinç and Fırat Çekinel for their unwavering support, invaluable assistance, and moral boosting during the most challenging phases of my research.

Furthermore, I express my gratitude to my teammates and friends, Nihan Öztürk, Elif Kübra Çontar, Aylin Aygüneş, Gözde Türkan, Rabia Güney, Şebnem Doğru, Emre Demiray, for their endless support and diligent work in annotating the data set, which played a crucial role in the success of my research. I would also like to thank my dearest friends Veysel Alperen Özkahriman, Ayşegül Doğan, and Seda Civelek for their endless support and motivation.

Moreover, I would like to acknowledge METU Computer Engineering faculty members for their expertise and patience.

Finally, I would like to give special thanks to Yıldırım Keskinpala, Erol Öztaş, and Sibel Keskinpala for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

NER	Named Entity Recognition
NLP	Natural Language Processing
IE	Information Extraction
BERT	Bidirectional Encoder Representations from Transformers
CoNLL	Conferences on Natural Language Learning
MUC	Message Understanding Conferences
NERC	Named Entity Recognition and Classification
DBLSTM	Deep Bidirectional Long Short-Term Memory
POS	Part-of-Speech
FSC	Finite State Cascades
CPU	Central Processing Unit
RAM	Random Access Memory

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Named Entity Recognition (NER) constitutes a widely acknowledged challenge within Natural Language Processing (NLP) and Information Extraction (IE), with the primary objective of autonomously identifying *Named Entities (NEs)* embedded within a given text. These named entities encompass distinctive categories of nouns and noun phrases that pertain to particular entities, encompassing designations like *Person*, *Location*, *Organization*, *Date-time* expressions, and more. Within existing literature, NER is conceptualized as a task involving sequence labeling, wherein a text is presented as a sequence of words, with tokens assigned corresponding labels indicative of their named entity type or categorized as *Other* to denote instances *Not a Named Entity*.

Research within Named Entity Recognition (NER) predominantly centers around languages like English and Chinese. Nevertheless, the scope of NER investigations for the Turkish language remains notably limited, particularly in the domain-specific context. Therefore, Turkish NER data sets are limited.

A significant portion of the landscape of Named Entity Recognition (NER) research is predominantly occupied by studies centered around the fintech and health domains. Nonetheless, in recent years, the intersection of sports and technology has ushered in a new era of innovation, influencing various aspects of sports management, performance enhancement, and fan engagement. As the world witnesses an upsurge in investments in sports, the impact of these developments transcends the boundaries of stadiums, extending into the realms of finance, marketing, and even psychol-

ogy. While quantitative analyses have traditionally dominated the landscape of sports statistics, the emergence of qualitative factors, drawn from textual data, holds the promise of enriching our understanding of the intricate interplay between athletes, teams, and supporters. For instance, in an article, Beşiktaş (a well-known sports club in Turkey) broke the decibel record in a match. In that match and the following matches, the players are positively affected. Conversely, instances, where articles spotlighted delays in salary disbursements, were accompanied by discernible downturns in player morale and performance. Through an interdisciplinary lens encompassing sports management, psychology, marketing, and finance, this research endeavors to harness the power of textual data to forge new pathways in sports research and management, ultimately shaping the future landscape of athletic endeavors.

Furthermore, the emergence of deep learning methods has led to a noticeable enhancement in the performance of NLP tasks, including the realm of Named Entity Recognition (NER). This phenomenon is particularly evident in research conducted primarily in English. However, due to the recent surge in the popularity of deep learning techniques, the utilization of these methods in addressing the Turkish NER problem remains comparatively limited in contrast to the traditional approaches.

Moreover, there exists a notable scarcity of research endeavors focusing on the interpretability and explainability aspects of models based on transformer architectures. This prevailing gap highlights a promising avenue for future investigations.

In this thesis work, we aim to propose novel domain-specific Turkish NER data sets, and measured the performance of popular transformers-based models on the constructed data sets. Also, we conduct the interpretability of the models.

1.2 Proposed Methods and Models

In this thesis, our primary objectives encompass three key aspects. Firstly, we create domain-specific Turkish data sets on the sports domain. These data sets were annotated by using both parent and child entities and two distinct tagging formats, such as BIO and NonBIO. Secondly, we conducted a comprehensive evaluation of five different pre-trained deep-learning models through fine-tuning on these data sets. The

main focus of this evaluation was to assess the models' performance in the context of named entity recognition for Turkish texts within the sports domain. Also, to assess the importance of data set balance in our study, we applied the 5-fold cross-validation method. Through this approach, we gained insights into how the distribution of data instances impacts the model's performance. Lastly, we explicated the rationale behind the precisions made by these deep learning models while using two well-known explainability methods. Through the fulfillment of these objectives, this thesis not only makes notable contributions to the field of named entity recognition in the sports domain in Turkish but also advances our understanding of the underlying mechanisms of deep learning models, thus fostering greater clarity and comprehensibility in their practical usage.

1.3 Contributions and Novelties

Our contributions are as follows:

- Introducing a novel annotated data set for the sports domain. These data sets contain two level of named entities such as parent and child, and are annotated by using both BIO and NonBIO formats, offering the potential for experimentation in future research with new models.
- Comparing and analyzing various language models' results on different annotated Turkish data sets from the sports domain which were constructed in the scope of this thesis.
- Evaluating the model's performance by comparing the impact of different annotation formats (BIO and NonBIO) on the same data set while keeping all other variables the same.
- Analyzing the influence of named entity distribution across different splits of data sets based on the results obtained from k-fold cross-validation experiments, mainly when dealing with imbalanced data sets.
- Using LIME and SHAP, explain why and how the models employed in the study, which are based on well-known methods, made their predictions.

1.4 The Outline of the Thesis

The rest of the thesis is organized as follows. Chapter 2 represents a comprehensive literature survey and an overview of related studies. Deep learning models employed in the study, explainability methods, and the data set including preparation, preprocessing, and a detailed explanation about the named entities are represented in Chapter 3. The experiments and their results are discussed in Chapter 4. Lastly, Chapter 5 provides a conclusion of the study.

CHAPTER 2

LITERATURE SURVEY

2.1 Named Entity Recognition

Named Entity Recognition (NER), also known as entity extraction or entity identification, is a subtask of information extraction. It aims to automatically identify pre-defined named entities such as person, location, and organization. It has practical applications in other Natural Language Processing (NLP) tasks such as summarization and question answering.

The term named entity was initially introduced within the context of the Message Understanding Conferences (MUC), particularly in MUC-6 [4]. Initially, the task was denoted as "Named Entity Recognition and Classification" (NERC), but subsequently, the term "Named Entity Recognition" (NER) has been universally accepted to refer to the same task. Apart from MUC, NER has been a central topic in various international conferences, including the Conferences on Natural Language Learning (CoNLL) [5].

In the initial stages of NER tasks, rule-based algorithms were commonly employed. After that, supervised and semi-supervised machine learning techniques are used. In supervised learning, a considerable quantity of labeled data is required, leading to potential expenses and time constraints associated with data collection. In response to this issue, alternative approaches such as semi-supervised and unsupervised learning have been introduced, exploiting unlabeled. Also, novel deep-learning models are introduced, such as BERT. These days, Conditional Random Field (CRF) and deep-learning methods are used together.

2.2 Named Entity Recognition for Turkish

Cucerzan and Yarowsky's work [6] is known as one of the first studies on named entity recognition in Turkish. They investigated *"language-independent named entity recognition by combining morphological and contextual evidence"*. The key feature of their algorithm is its independence from language-specific information, making it applicable to various languages.

In 2003, Tür et. al. [7] introduced the pioneering study in Turkish language processing. Their approach involved the proposition of statistical language processing methodologies, encompassing the tasks of sentence segmentation, topic segmentation, and name-tagging. Considering the issue of data sparsity in Turkish texts, the authors addressed this concern by using both lexical and morphological information. Specifically, they introduced a Hidden Markov Model-based approach.

In the publication by Küçük and Yazıcı [8], a rule-based approach for Turkish named entity recognition was presented in 2009. The methodology utilized lexical resources and pattern bases to achieve its objectives. The experiments were conducted on a Turkish news data set, and the study focused on identifying seven distinct types of named entities.

In 2010, Küçük and Yazıcı [9] published a study that proposed a hybrid named entity recognizer that is the improved version of the rule-based method. They indicate that the hybrid recognizer is more successful compared to the rule-based approach.

In 2011, Tatar and Çiçekli [10] introduced an automated rule learning system that relies on morphological, contextual, and orthographic characteristics of Turkish words. Their data sets encompass diverse Turkish newspapers as the source of information for their research.

Yeniterzi [11] investigated the impact of morphological features on a named entity recognition system designed for Turkish. The study focused on addressing the challenges arising from Turkish being an agglutinative language with intricate morphological structures. The experiments started with word-level representations and progressively incorporated morpheme-level features, including syntactic and contextual

properties. By observing the results achieved through the incremental addition of features, the author concluded that the morpheme-level model exhibited a significant enhancement in performance.

Özkaya and Diri [12] introduced a Conditional Random Field (CRF) approach for extracting named entities from emails from academic institutions, corporations, and individuals. Their study focused on data sets comprising solely person, location, and organization entities.

In [13], Şeker and Eryiğit constructed extensive gazetteers, denoted as base and generator gazetteers. Additionally, the authors presented the evaluation results of the Conditional Random Field method when combined with utilizing these gazetteers. Their gazetteers specifically encompassed ENAMEX types exclusively. After the work on ENAMEX types, Şeker and Eryiğit published a study [14] with an extended version of their previous work on TIMEX and NUMEX types.

Gunes and Tantug conducted a research work [15] focusing on named entity recognition using deep learning techniques for Turkish. The authors introduced a Deep Bidirectional Long Short-Term Memory (DBLSTM) model applied to news articles in their study. They reported a remarkable recognition performance with an achieved f1-score of 93.69%.

Çekinel et. al. [16] developed a system based on the utilization of Conditional Random Fields. They made improvements to the model by integrating morphological features of the words. According to their findings, the addition of word features had a positive impact on the overall performance of named entity recognition.

Kılıç et. al. [17] discussed the performance of the neural model, Google BERT, on Turkish texts. Furthermore, the authors have provided a distinct perspective and also interpreted the performance of different training levels from pre-trained to full trained by freezing the layers.

Akkaya and Can [18] investigated using deep neural networks with different word representation techniques for NER on Turkish noisy text. They utilized character-level, character n-gram-level, morpheme-level, and orthographic character-level word representations. To handle infrequent entity types, the authors introduced a transfer

learning model. The proposed model incorporated an additional CRF layer.

In their work [19], Çarık and Yeniterzi (2022) presented a novel data set comprised of Twitter data. The data set was labeled by multiple annotators, exhibiting a high agreement score. It encompasses various entities, including person, organization, location, time, money, product, and tv-show. For conducting the experiments, the authors utilized pre-trained models like BertTurk[20].

In [21], Çetindağ et. al. constructed a novel corpus for the Turkish legal domain. They further proposed a Named Entity Recognition model, combining CRFs and BiLSTMs. Additionally, the authors explored diverse word embedding techniques, including GloVe, Morph2Vec, and character feature extraction through neural networks, using either BiLSTM or convolutional neural networks.

2.3 Studies of Domain Specific Named Entity Recognition in Sports and Related Domains

In 2003, Yao et. al. [22] proposed an approach to address word segmentation and Part-of-Speech (POS) tagging errors. They devised Finite State Cascades (FSC) as a shallow parser. The data set employed in their study consisted of various entity types, including personal name (PN), date or time (DT), location name (LN), team name (TN), competition title (CT), and personal identity (PI) entities. According to the authors' findings, the system achieved an average recall of 83% and precision of 85%.

In [23], Küçük investigated the impact of named entities on the task of stance detection in tweets. The data set utilized in the study was initially presented in [24] and consisted of tweets related to the two major Turkish sports clubs, Fenerbahçe and Galatasaray. In [23], this existing data set was expanded by adding annotations for named entities, including person, location, and organization. Two distinct methods were employed for annotating the named entities: extraction using a NER tool (developed in [25]) and manual annotation.

Compared to the previous study [24], the experiments for stance detection were con-

ducted by using NE in addition. The author observed that named entities obtained through a high-performance NER system positively impacted the stance detection performance in tweets.

In their work [26], Küçük and Can provided a valuable contribution by releasing a publicly available data set of Turkish tweets that have been annotated for both named entities and stance information.

In 2020, Seti et. al. [27] presented an approach for named entity recognition in Chinese sports text. Their method utilized a character graph convolutional neural network (Char GCN) with a self-attention mechanism model. The authors designed a graph wherein Chinese characters serve as nodes, and the presence of edges is determined by both the character position and the character feature of the named entity. Furthermore, compared to the traditional methods, the proposed method is more successful in recognizing named entities in the sports domain.

In [27], they utilized the SportsNER data set, which is not publicly available. This data set comprises ten different entities related to sports, including sports competition, sports team, sports location, sports person, sports job name, sports level, sports time, sports media, and sports organization. Unfortunately, no further information about the data set was provided in [27].

In 2023, Wijesinghe and Tissera [28] presented a novel data set in the Sinhala language, specifically targeting the sports domain. The data set was carefully collected from e-news articles, including annotated named entities, such as ground, tournament, and school.

As part of their research in [29], Yang et. al. constructed a comprehensive data set encompassing diverse domains, including sports and e-sports. This data set was collected from Chinese Wikipedia and annotated with hierarchical categories. Within the sports domain, one parent named entity, i.e., team was identified, which, in turn, had two associated child entities: sports team and e-sport team.

Although this thesis focused on named entity recognition in the sports domain, several studies exist in the e-sports domain.

In their work [30], Liu et. al. introduced a Named Entity Recognition (NER) system explicitly designed for e-sports news. The data set used in their study was curated from articles within the e-sports domain, gathered from e-sports forums. This data set encompasses five distinct named entities: GAME (game), TOURN (tournament), ORG (organization), PLAYER (player), and AVATAR. The authors employed Conditional Random Field (CRF) and BERT-based methods to evaluate the recognition performance on this data set. Additionally, the authors constructed an *"end-to-end e-sports entity Liquipedia system which recognizes e-sports players, teams, and tournaments from texts"*.

2.4 Explainable Named Entity Recognition

Kokalj et. al. [31] introduced TransSHAP, a novel adaptation of SHAP for transformer models like BERT. While transformer-based neural networks excel in classification, they lack interpretability. Existing explanation methods like SHAP are unsuitable for transformers and overlook the sequential nature of text. TransSHAP overcomes these limitations, offering sequential explanations that rival top solutions according to human evaluators.

Hossain et. al. [32] introduced "garNER". The approach involves knowledge augmentation through Wikipedia API queries to enrich input sentences with entity summaries. The entities are sourced from labeled data or entity taggers. Ensembling techniques are applied to multiple models for final predictions. Also, the authors analyzed outputs with LIME to understand how augmented input affects NER classification. LIME explains the adjacent tokens within an input sentence in this work. Contextual augmentation is found effective when relevant to target entities but counterproductive when irrelevant.

CHAPTER 3

DATA SET CONSTRUCTION AND EXPLAINABLE NAMED ENTITY RECOGNITION APPROACHES

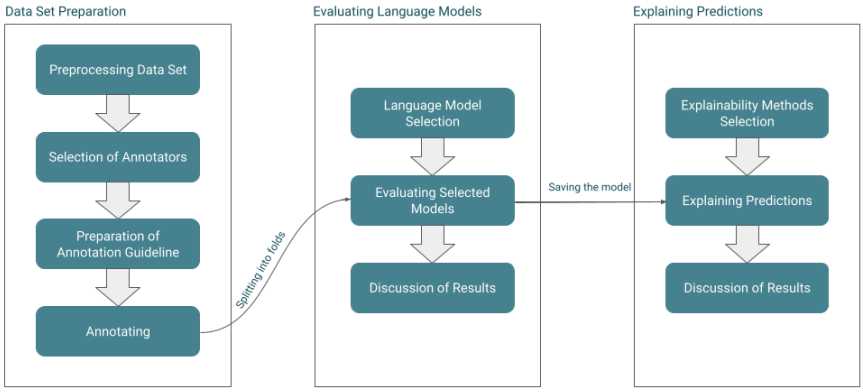


Figure 3.1: Method Overview

This thesis comprises three primary phases encompassing data set preparation, language model evaluation, and prediction interpretation. The overarching architecture of the study is depicted in Figure 3.1. The inception of the study involved a meticulous data set preparation process, which sequentially encompassed the pre-processing of the data set, the selection of annotators possessing domain expertise, the formulation of precise annotation guidelines, and the subsequent annotation of the data set. Following the meticulous annotation process, the data set was partitioned into distinct folds, facilitating its utilization as input for the subsequent modeling phase.

Concurrently, a systematic approach was undertaken to select appropriate language models, a pivotal aspect of the study. The selected models underwent rigorous evaluation using the annotated data set, ensuring an informed understanding of their per-

formance within the context of sports-related named entity recognition. In parallel to these evaluation endeavors, a judicious selection of explainability methods was executed, a critical undertaking to unveil the rationale underpinning the models' predictions.

Subsequently, the chosen models were subjected to the selected explainability methods, thus shedding light on the intricate decision-making processes within the models. The comprehensive architectural representation provided in Figure 3.1 underscores the methodical nature of the study, delineating its holistic structure across the domains of data set curation, model evaluation, and interpretability enhancement.

3.1 Data Set

In the scope of the thesis, mainly four different data sets are used, and both of them are newly constructed for this study.

3.1.1 Preparation of TurkishSportsNER Data Set

TurkishSportsNER data set was constructed by preprocessing and annotating the articles from a Turkish newspaper internet page Milliyet.com for the period of 1997 to 2019. The data set is publicly available on GitHub.

TurkishSportsNER data set contains 1570 news, 571822 words, and 130165 tagged words. TurkishSportsNER has four different versions:

- TurkishSportsNER-Parent-NonBIO: This version of the data set contains only parent entities with Non-BIO format. For this data set, 18 unique annotations were used. For example, *taç atışı* was annotated as *SPORT_TERM SPORT_TERM*.

- TurkishSportsNER-Parent-BIO: The data set was constructed by annotating only parent entities with BIO format. There are 35 unique labels in this version of the data set. For instance, *B-SPORT_TERM I-SPORT_TERM* was used for *taç atışı*.

- TurkishSportsNER-Child-NonBIO: The data set contains child entities with Non-BIO format. There are 28 labels in this version of the data set. For example, *taç atışı* was labelled as *FOOTBALL_TERM FOOTBALL_TERM*.

- TurkishSportsNER-Child-BIO: This version of the data set is the most special version. It contains all the child entities with BIO format. 55 unique annotations represents in this version. For instance, *B-FOOTBALL_TERM I-FOOTBALL_TERM* was used for *taç atışı* as an annotation.

The list of labels for each TurkishSportsNER data set can be found in Table 3.1. Also, an example sentence with labels according to constructed data sets can be found in Table 3.2. Detailed explanations about the format of tagging and named entities can be found in subsections of this section. Detailed statistics of the TurkishSportsNER data set were explained in Section 4.1.

Table 3.1: The List of Labels for each TurkishSportsNER Data Set

TurkishSportsNER-Parent-NonBIO	TurkishSportsNER-Parent-BIO		TurkishSportsNER-Child-NonBIO		TurkishSportsNER-Child-BIO		
O	O	I-ALIAS	O	SPORT_NAME	O	B-SPORT_NAME	I-RACE_TERM
AGREEMENT	B-AGREEMENT	I-EQUIPMENT	AGREEMENT	SPORT_ORGANIZATION	B-AGREEMENT	B-SPORT_ORGANIZATION	I-REFEREE
ALIAS	B-ALIAS	I-HEALTH_TERM	BASKETBALL_TERM	SPORT_PLAYER	B-BASKETBALL_TERM	B-SPORT_PLAYER	I-ROLE
EQUIPMENT	B-EQUIPMENT	I-NATION	COACH	SUPPORTER	B-COACH	B-SUPPORTER	I-SCORE_TERM
HEALTH_TERM	B-HEALTH_TERM	I-ROLE	EQUIPMENT	TEAM	B-EQUIPMENT	B-TEAM	I-SPORT_COMPETITION
NATION	B-NATION	I-SPORT_COMPETITION	FOOTBALL_TERM	TEAM_ALIAS	B-FOOTBALL_TERM	B-TEAM_ALIAS	I-SPORT_EVENT
ROLE	B-ROLE	I-SPORT_EVENT	HEALTH_TERM	TEAM_SPONSOR	B-HEALTH_TERM	B-TEAM_SPONSOR	I-SPORT_LEVEL
SPORT_COMPETITION	B-SPORT_COMPETITION	I-SPORT_LEVEL	HORSE_RACING_TERM	TENNIS_TERM	B-HORSE_RACING_TERM	B-TENNIS_TERM	I-SPORT_LOCATION
SPORT_EVENT	B-SPORT_EVENT	I-SPORT_LOCATION	MANAGER	VOLLEYBALL_TERM	B-MANAGER	B-VOLLEYBALL_TERM	I-SPORT_NAME
SPORT_LEVEL	B-SPORT_LEVEL	I-SPORT_NAME	NATION		B-NATION	I-AGREEMENT	I-SPORT_ORGANIZATION
SPORT_LOCATION	B-SPORT_LOCATION	I-SPORT_ORGANIZATION	PERSON_ALIAS		B-PERSON_ALIAS	I-BASKETBALL_TERM	I-SPORT_PLAYER
SPORT_NAME	B-SPORT_NAME	I-SPORT_PERSON	RACE_TERM		B-RACE_TERM	I-COACH	I-SUPPORTER
SPORT_ORGANIZATION	B-SPORT_ORGANIZATION	I-SPORT_TERM	REFEREE		B-REFEREE	I-EQUIPMENT	I-TEAM
SPORT_PERSON	B-SPORT_PERSON	I-SUPPORTER	ROLE		B-ROLE	I-FOOTBALL_TERM	I-TEAM_ALIAS
SPORT_TERM	B-SPORT_TERM	I-TEAM	SCORE_TERM		B-SCORE_TERM	I-HEALTH_TERM	I-TEAM_SPONSOR
SUPPORTER	B-SUPPORTER	I-TEAM_SPONSOR	SPORT_COMPETITION		B-SPORT_COMPETITION	I-HORSE_RACING_TERM	I-TENNIS_TERM
TEAM	B-TEAM		SPORT_EVENT		B-SPORT_EVENT	I-MANAGER	I-VOLLEYBALL_TERM
TEAM_SPONSOR	B-TEAM_SPONSOR		SPORT_LEVEL		B-SPORT_LEVEL	I-NATION	
	I-AGREEMENT		SPORT_LOCATION		B-SPORT_LOCATION	I-PERSON_ALIAS	

3.1.2 Preprocessing

Milliyet.com is an internet page of the Turkish newspaper Milliyet. It presents news about different categories such as economy, sports, politics, and magazines. Someone collected all the data for the period of 1997 to 2019. The data set is publicly available on Github.

In the source data set, each article has its own category as in Milliyet.com. Firstly, all categories related to sports were collected from the newspaper website. Even in some articles in sports-related categories, the content of them is different. After getting

Table 3.2: A Sample Sentence with Labels According to the Data Set

Sentence	TurkishSportsNER-Parent-NonBIO	TurkishSportsNER-Parent-BIO	TurkishSportsNER-Child-NonBIO	TurkishSportsNER-Child-BIO
cep	ALIAS	B-ALIAS	PERSON_ALIAS	B-PERSON_ALIAS
herkülü	ALIAS	I-ALIAS	PERSON_ALIAS	I-PERSON_ALIAS
naim	SPORT_PERSON	B-SPORT_PERSON	SPORT_PLAYER	B-SPORT_PLAYER
süleymanoğlu	SPORT_PERSON	I-SPORT_PERSON	SPORT_PLAYER	I-SPORT_PLAYER
ile	O	O	O	O
imparator	ALIAS	B-ALIAS	PERSON_ALIAS	B-PERSON_ALIAS
fatih	SPORT_PERSON	B-SPORT_PERSON	COACH	B-COACH
terim	SPORT_PERSON	I-SPORT_PERSON	COACH	I-COACH
cimbom	ALIAS	B-ALIAS	TEAM_ALIAS	B-TEAM_ALIAS
stadında	SPORT_LOCATION	B-SPORT_LOCATION	SPORT_LOCATION	B-SPORT_LOCATION
futboldaki	SPORT_NAME	B-SPORT_NAME	SPORT_NAME	B-SPORT_NAME
transferler	AGREEMENT	B-AGREEMENT	AGREEMENT	B-AGREEMENT
hakkında	O	O	O	O
konuşmak	O	O	O	O
için	O	O	O	O
buluştu	O	O	O	O

sport-related categories, all the articles were reviewed, and only sport-related articles were kept. After creating the initial data, the second step was preparing data in the proper data set format. That's why, firstly, all the unnecessary spaces, tabs, and new lines in the news text were removed. Also, all of the words in the news text were converted to lowercase. After that, words were split by punctuation. After preparing the words, all the news was split into two new lines.

At the end of the preprocessing steps, the data set contains news text that is separated by two blank lines. Moreover, each line of a word (token) has the Named Entity tag (label) next to it separated by a tab.

3.1.3 Annotation of Named Entities

Given the sports-centric nature of the data set, a pragmatic strategy was adopted to expedite the annotation process for this study. Annotators were judiciously selected from a pool of individuals possessing an athlete's background and familiarity with sports terminology. A panel of ten annotators was assembled for this purpose. Initially, a representative subset of the data set was meticulously scrutinized to inform the annotation decisions. This preliminary assessment led to a preliminary identification of labels, providing an initial structure for the annotation process. To facilitate this, a comprehensive Excel sheet was devised, encompassing annotations alongside contex-

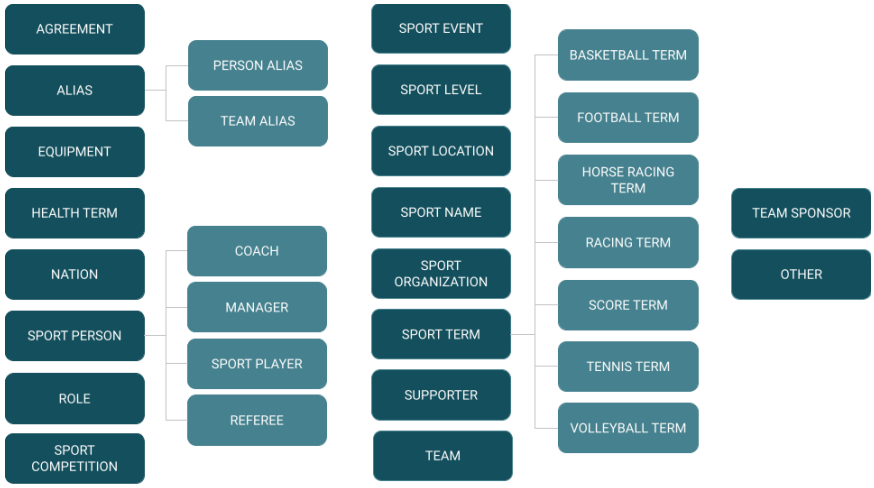


Figure 3.2: Schema of Parent and Child Entities (Dark rectangles represent parent entities, light rectangles represent child entities.)

tual samples indicative of the respective annotations. Notably, certain words exhibited polysemy and could warrant distinct annotations contingent on contextual nuances. As an illustration, the term *faul* could be annotated as either FOOTBALL_TERM or BASKETBALL_TERM based on the sports context. Subsequently, the parent entity names were systematically deduced with reference to their specific child entities.

During the annotation procedure, instances where annotators encountered ambiguity were addressed through a polling mechanism. In these instances, a collaborative voting process among annotators was instated to ascertain the most apt annotation. The final annotation was determined by selecting the annotation garnered the highest consensus among the annotators. This meticulous approach underscores the comprehensive and collaborative nature of the annotation process, aligning with the rigorous standards of data set curation within the domain of sports-related named entity recognition.

In this study, two different annotation formats were used: BIO format and Non-BIO format. BIO represents "beginning", "inside", and "outside". With the help of the BIO format, we can understand whether the word is at the beginning of the phrase or inside the phrase. Two versions of the TurkishSportsNER data set were constructed by using BIO format, and the other two data sets were constructed by using Non-BIO format.

In TurkishSportsNER data set, there are two types of entities such as child and parent. Two versions of the data set were constructed by using child entities and parent entities were used for the other constructed data sets. Most of the entities are the same for both child and parent; however, some of them are different. For the sake of example, although ALIAS is a parent entity, PERSON ALIAS or TEAM ALIAS is a child entity. The schema of parent and child entities can be found in Figure 3.2. All the detailed information about the annotations can be found in this section.

3.1.3.1 AGREEMENT

Sports clubs always have agreements with both players and sponsors. That's why, all words/phrases related to the agreement were annotated as $AGREEMENT$. "imzalamak" (sign), "transfer" (transfer) are some examples of this entity.

3.1.3.2 ALIAS

Since some of the well-known sports clubs and players/coaches have an alias, news content in TurkishSportsNER data set contains these aliases. Therefore, this tag was split into two different entities such as TEAM ALIAS, and PERSON ALIAS. Although one version of TurkishSportsNER data set contains only ALIAS label, the other one contains both TEAM ALIAS and PERSON ALIAS labels.

For instance, when "cimbom" represents the well-known team Galatasaray and it takes TEAM ALIAS, "cimbomlu" represents the supporter of Galatasary and it takes PERSON ALIAS. Also, the meaning of the word/phrase in the context was considered to decide for annotation. For example, "siyah beyazlılar" can represent both players of the team Beşiktaş and supporters of the team Beşiktaş.

PERSON ALIAS Mostly, coaches who have served Turkish sports for a long time have popular nicknames. For example, Fatih Terim was the coach of the national football team and worked as a coach for different teams. He acted like a real leader and got spacious achievements under even difficult conditions. Therefore, the people living in Turkey called him as "imparator" (emperor). Furthermore, some of the play-

ers have also nicknames. For instance, At 147 cm in height, Naim Süleymanoğlu's short stature and great strength led to him being nicknamed "cep herkülü" (pocket hercules).

TEAM ALIAS In Turkey, well-known sports clubs have aliases. These aliases can be based on their team colors, their emblems. For example, "sarı-kırmızılı" (yellow-red) was produced according to the colors of team Galatasaray. On the other hand, "kara kartallar" (black eagles) were produced according to both the color of the team and the emblem of the team Beşiktaş.

3.1.3.3 EQUIPMENT

All of the equipment belongs to this label. There is one difference between the Sport Term and equipment that is if the word/phrase is used in the meaning as furniture then it should have $EQUIPMENT$ label.

"forma" (uniform), "top" (ball), and "düdük" (whistle) are examples of the equipment.

3.1.3.4 HEALTH TERM

Some players can be injured. These kinds of terms are represented as health terms. "çapraz bağ" (cross-link), and "sakatlandı" (injured) are examples of this label.

3.1.3.5 NATION

In the realm of sports, athletes from diverse countries are celebrated, and their nationalities play a key role in news reporting. Terms like "Türk" (Turkish), "İsviçreli" (Swiss), and "Fransız" (French) are prominent examples of this practice. Mentioning an athlete's nationality adds depth to the narrative, offering readers a profound understanding of their cultural background and national identity. It fosters a sense of pride and unity among sports enthusiasts in the athlete's home country, as they celebrate the athlete's achievements on the international stage. Through this practice, sports news becomes a testament to the global spirit of competition and camaraderie among

athletes from various nations, transcending geographical boundaries and promoting mutual respect among diverse cultures.

All kinds of branches of sports have players that are from various countries. In order to specify the player, the nation is also used a lot in the news. "Türk" (Turkish), "İsviçreli" (Swiss), and "Fransız", (French) are examples of this entity.

3.1.3.6 ROLE

There are various roles in the sports domain and mentioned a lot in the news. "hakem" (referee), "teknik direktör" (coach), "futbolcu" (football player) are examples of the roles. Words/phrases with this label do not represent a specific person, they represent a profession.

3.1.3.7 SPORT COMPETITION

This entity covers all the competitions related to sports such as football, basketball leagues, and tournaments. While "Şampiyonlar Ligi" (Champions League), "Süper Lig" (Super League) were tagged as $SPORT\_COMPETITION$, "Spor Toto Süper Ligi" (Super League) was also tagged as $SPORT\_COMPETITION$.

3.1.3.8 SPORT EVENT

All kind of events related to sports represents with this tag such as "maç" (match), "lig" (league). This tag does not contain special events, it contains only event names.

3.1.3.9 SPORT LEVEL

Sports are done at different levels such as amateur, and professional. Sometimes, these levels are named according to age groups like U16, U18. All kinds of levels related to sports were tagged as $SPORT\_LEVEL$.

3.1.3.10 SPORT LOCATION

$SPORT\_LOCATION$ represents all the location-related words. Not only special names were annotated as $SPORT\_LOCATION$, but also direct sport locations were annotated as $SPORT\_LOCATION$. For instance as a special sports location name, "Ali Sami Yen Stadyumu" (Ali Sami Yen Stadium) was annotated as "$B\_SPORT\_LOCATION\ I\_SPORT\_LOCATION\ I\_SPORT\_LOCATION$" and "$SPORT\_LOCATION\ SPORT\_LOCATION\ SPORT\_LOCATION$". Also, "stad" (stadium) is an example of a non-special name of a sports location that was annotated as "$B\_SPORT\_LOCATION$" and $SPORT\_LOCATION$. Unlike typical named entities, in this data set, city names are not labeled as locations. This is because there are many entities that could be labeled as sports locations, like stadiums and fields, and city names are often used as team names. This approach was chosen due to the abundance of such entities and the common usage of city names in this context.

3.1.3.11 SPORT NAME

Names of the sports such as football, basketball, tennis, and F1 were labeled as $SPORT\_NAME$.

3.1.3.12 SPORT ORGANIZATION

All the sports organizations are annotated as $SPORT\_ORGANIZATION$. Sports organizations regulate sports-related things, and organize matches at different levels such as amateur, and professional, and have the right to penalize players or sports clubs. That's why, these organizations occupy a prominent place in the field of sports.

For instance, $SPORT\_ORGANIZATION$ label contains "Türkiye Futbol Federasyonu" (Turkish Football Federation), "Türkiye Milli Olmipiyat Komitesi" (Turkish National Olympic Committee).

3.1.3.13 SPORT PERSON

Name of a sport-related person. TurkishSportsNER data set was annotated as the first name with or without last name as a $SPORT\_PERSON$. One can have multiple names. All the names are annotated as $SPORT\_PERSON$. This entity covers all names of the people related to sports such as players, coaches, managers, and referees. Also, in this work, these entities were annotated as different entities such as $SPORT\_PLAYER$, $COACH$, $MANAGER$, and $REFEREE$. Detailed information about that can be seen in subsections.

COACH $COACH$ represents the name of all types of coach that works as a coach in all branches of sports. In the TurkishSportsNER data set, there are lots of interviews with the coaches. That's why there are lots of coach names represented in the TurkishSportsNER data set. Some examples of this entity from TurkishSportsNER data sets are, "Fatih Terim", "Ersun Yanal", "Şenol Güneş".

MANAGER $MANAGER$ represents the name of all types of manager that manages all type of sports clubs. "Ali Koç", "Nihat Özdemir" are examples of this group.

SPORT PLAYER $SPORT\_PLAYER$ represents the name of the players that play in all branches of sports such as basketball players, football players, tennis players, and volleyball players. Almost every news includes at least one player's name. Thus, $SPORT\_PLAYER$ named entity is another type that took place in TurkishSportsNER annotation.

Some examples of this entity from TurkishSportsNER data sets are, "Emre Belezoğlu", "Rodrygo Vinicius Jr.", "Mert Hakan Yandaş", "Cedi Osman".

REFEREE $REFEREE$ represents the names of all types of referees that work in all sports branches. "Cüneyt Çakır", "Ali Palabıyık" are examples of referees.

3.1.3.14 SPORT TERM

All the terms related to sports were tagged as $SPORT\_TERM$. Since there are lots of branches of sports, this tag has also separated sub-labels such as basketball term,

football term, and score term.

For $SPORT\_TERM$, the context of the news is also important in order to decide which sub-label the term belongs to. For the sake of example, "faul" (foul) can belong to basketball, football, or volleyball.

BASKETBALL TERM $BASKETBALL\_TERM$ represents all the terms regard to basketball. This entity contains not only terms that are used only in basketball like "periyot" (period), but also general terms like "şut" (shout).

FOOTBALL TERM $FOOTBALL\_TERM$ contains terms related to football. Also, FOOTBALL TERM is the most used sports term in TurkishSportsNER data sets.

HORSE RACING TERM $HORSE\_RACING\_TERM$ were constructed of horse racing-related terms. There are some horse names were tagged as HORSE RACING TERM.

RACING TERM $RACING\_TERM$ were constructed of racing-related terms. There is some F1-related news in TurkishSportsNER data sets.

TENNIS TERM $TENNIS\_TERM$ were constructed of tennis-related terms.

VOLLEYBALL TERM $VOLLEYBALL\_TERM$ were constructed of volleyball-related terms. The context of the news is important to decide like other branches of sports.

SCORE TERM Score term represents all the score-related terms like "galibiyet" (win), "mağlubiyet" (loss) etc.

3.1.3.15 SUPPORTER

Each team has its own supporters. This entity helps us to recognize which words/phrases belong to a supporter such as "taraftar" (supporter), "sporsever" (sports-lover).

3.1.3.16 TEAM

TEAM represents the name of the teams in all types branches of sports. "Galatasaray", "fenerbahçe", "anadolu efes" are the examples of TEAM.

3.1.3.17 TEAM SPONSOR

Almost every team has its own sponsor. These sponsors help to improve the team and supply budget. At the end, teams update their names with the name of the sponsor. For example, although "Çağdaş Faktoring" is part of the name of the team, "Tunç Holding" is a team sponsor in this sentence "Galatasaray Çağdaş Faktoring Kadın Basketbol Takımı şort yanı sponsoru Tunç Holding oldu." (The shorts side sponsor of Galatasaray Çağdaş Factoring Women's Basketball Team was Tunç Holding.).

3.1.3.18 OTHER

If the word does not belong to any of the tags above, then it was tagged as OTHER.

3.2 Deep Learning Models

In 2017, Vaswani et al. [33] introduced the transformer architecture, a relatively straightforward network design leveraging attention mechanisms. Notably, this architecture demands less training time compared to earlier recurrent neural architectures like long short-term memory (LSTM).

This architecture has found widespread usage such as natural language processing, and multi-modal processing. Additionally, it has played a pivotal role in the development of pre-trained systems like generative pre-trained transformers (GPTs) and BERT (Bidirectional Encoder Representations from Transformers).

In the realm of transformer-based architectures, there are common core components shared among various models. These components encompass tokenizers, vital in converting textual data into tokens, facilitating further processing. Following tokeniza-

tion, the embedding layers convert the tokens into semantically meaningful representations, capturing the essence of the input text. However, the crux of transformer-based models lies in their transformer layers, which house the reasoning capabilities. Within these transformer layers, Attention and MLP (Multi-Layer Perceptron) layers collaborate to enable effective and context-aware processing of the input data, making transformers particularly powerful in a wide range of natural language processing, computer vision, and multi-modal tasks.

Utilizing encoder and decoder layers in the original transformer architecture allowed for bidirectional and autoregressive processing, enabling comprehensive language understanding and generation capabilities. However, as subsequent models evolved, researchers began to explore the benefits of specialized models with either encoder-only or decoder-only configurations. For example, BERT focused on encoding contextual information bi-directionally. This diversity in transformer designs showcases the adaptability and versatility of the transformer architecture in addressing a wide range of natural language processing challenges.

3.2.1 BERT

BERT (Bidirectional Encoder Representations from Transformers), proposed by Devlin et al. [1] in 2018, is a groundbreaking transformer-based language model. Unlike traditional models BERT employs a bidirectional approach, enabling it to consider the entire context of a word by capturing information from both left and right contexts.

BERT utilizes WordPieces embeddings [34] instead of words. This strategy effectively reduces vocabulary size while augmenting the data available for each word. In contrast to predicting the subsequent word following a sequence of words, BERT employs random word masking within the sentence and endeavors to predict the masked words. This approach compels the model to learn how to leverage information from the entire sentence effectively.

BERT incorporates token embeddings, positional embeddings, and segment embeddings as input. The transformer architecture, however, faces a limitation concerning the consideration of input order. In response, BERT addresses this challenge by learn-

ing and employing positional embeddings, enabling the expression of word positions within a sentence. Additionally, it learns distinct embeddings for the first and second sentences, facilitating the model's ability to distinguish between the sentences effectively.

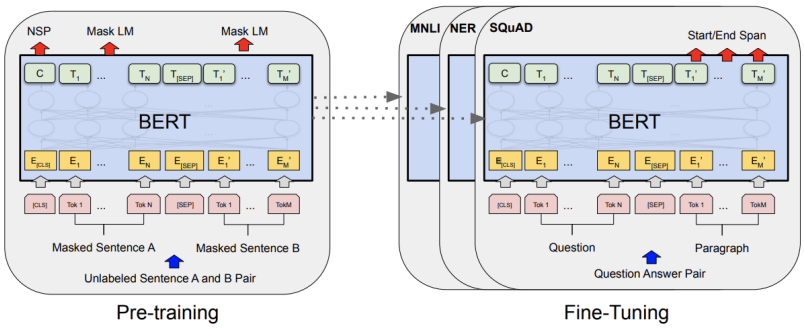


Figure 3.3: Pre-training and Fine-tuning Procedures for BERT. Sourced by [1]

Pre-training is a crucial phase in BERT's development, where the model is exposed to vast amounts of unlabeled text to learn general language representations. Fine-tuning involves training BERT on a smaller data set annotated for a specific task, such as sentiment analysis, named entity recognition, or question answering. This process allows BERT to specialize and excel in various natural language processing tasks.

To adapt BERT for NER, additional layers (as seen in Figure 3.3), often comprising conditional random fields or softmax classifiers, are appended to the model. These layers enable the prediction of entity labels for each token based on the contextualized representations generated by BERT. During the fine-tuning process, the model's weights are updated using annotated NER data to optimize its performance specifically for the NER task. By capitalizing on BERT's pre-trained language understanding and subsequently fine-tuning it on NER data, this approach harnesses the synergy of contextualized embeddings and domain-specific information, yielding state-of-the-art results in NER tasks across various languages and domains.

In this thesis work, two different versions of BERT were fine-tuned such as BERT-Base Multilingual Cased Model [35] and BERT-BASE Turkish Uncased Model [20]. Both models are publicly available on the Hugging Face platform.

BERT-Base Multilingual Cased Model [35], proposed in [1] and first released in [36],

trained on a large data set including 104 languages. The pre-trained model contains 12 layers, a hidden size of 768, 12 heads, and 110M parameters.

In conducting the experiments, the hyper-parameters of the BERT-Base Multilingual Cased Model [35] were predominantly maintained at their default settings, in alignment with the pre-configured specifications of the model [35]. Notably, the maximum sequence length was defined as 512, ensuring that longer sequences were suitably accommodated. Furthermore, a batch size of 8 was employed to regulate the computational load during training, optimizing the trade-off between memory efficiency and processing speed. Regarding the iterative optimization process, the model was trained for three epochs. This amalgamation of default settings and select parameter adjustments created a robust experimental framework to comprehensively evaluate the performance of the BERT-Base Multilingual Cased Model in the context of this thesis work.

BERT-BASE Turkish Uncased Model [20], also known as BERTurk, trained on Turkish OSCAR corpus, Wikipedia dump, different OPUS corpora, and a special corpus contributed by Kemal Oflazer. The total training corpus encompassed a substantial size of 35GB. For hyper-parameters, the model adopted the same configuration as the BERT-Base Multilingual Cased Model.

In the context of the conducted experiments, the hyper-parameters governing the behavior of the BERT-Base Turkish Uncased Model [20] were predominantly configured to their default values, aligning with the established specifications detailed in the model's configuration documentation [20]. It is important to note that while default settings formed the foundation of parameter choices, strategic adjustments were introduced to key aspects of the training process. Specifically, the maximum sequence length was established as 512, ensuring that input data samples of varying lengths were accommodated while maintaining a coherent training process. In addition, a batch size of 8 was selected, a value that strikes a balance between computational efficiency and memory utilization during the training procedure. The iterative training regimen, spanning a total of 3 epochs, signifies the frequency with which the entire training data set underwent optimization.

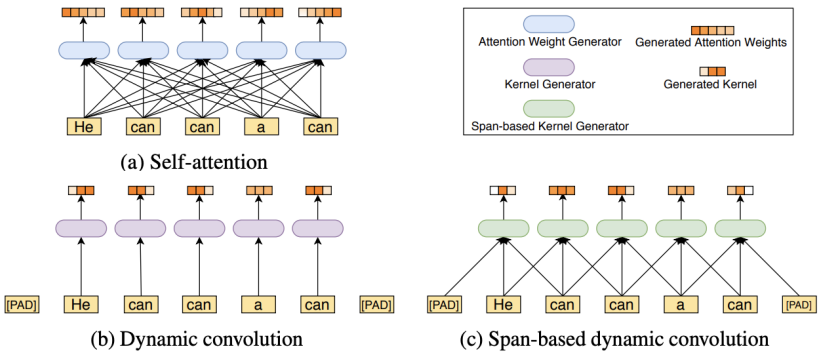


Figure 3.4: Processes of generating attention weights or convolution kernels. Sourced by [2]

3.2.2 ConvBERT

Jiang et al. [2] proposed ConvBERT model which is constructed as a version of BERT by integrating convolution into self-attention to form a mixed attention mechanism that is called span-based dynamic convolution. The architecture of self-attention, dynamic convolution and span-based dynamic convolution were represented in Figure 3.4. The driving incentive of the author stems from the resource-intensive nature of BERT in terms of memory and computational requirements. In their work, only some heads learn local dependencies. The results of ConvBERT significantly higher than BERT with reduced training cost.

ConvBERT-Base Turkish Uncased Model [37] was trained on Multilingual C4 (mC4) [38]. The training corpus has 31,240,963,926 tokens.

In the course of this thesis work, the ConvBERT-Base Turkish Uncased Model [37] underwent fine-tuning across different versions of TurkishSportsNER data sets. The model implemented in this study comprises a configuration of 12 layers.

In the course of conducting the experiments, the hyper-parameters governing the ConvBERT-Base Turkish Uncased Model's behavior [37] were predominantly configured according to the default specifications outlined in the model's configuration documentation [37]. Notably, the parameter regulating the maximum sequence length was set to 512. The batch size parameter was established at 8. The iterative training protocol was implemented over a span of 3 epochs, allowing for a comprehensive

exploration of the model's learning dynamics and convergence behavior. By thoughtfully configuring these hyper-parameters, encompassing both default values and deliberate adjustments, a robust experimental framework was established to assess the ConvBERT-Base Turkish Uncased Model's efficacy in the context of this thesis work.

3.2.3 DistilBERT

DistilBERT was introduced by Sanh et al. [39] titled "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.". The model has the same general architecture as BERT. In their study, they eliminated token-type embeddings and the pooler and reduced the number of layers by half. Furthermore, the authors utilized the student-teacher method during the distillation process.

DistilBERT-Base Turkish Cased Model [40], also known as DistilBERTurk, trained on a reduced data set of 7GB extracted from the original training data used for training BERTurk[20]. During this process, the cased version of BERTurk was employed as the teacher model to facilitate the distillation process of knowledge transfer.

In this thesis work, DistilBERT-Base Turkish Cased Model [40] was fine-tuned on variations of TurkishSportsNER data sets. The model employed in the study has 6 layers.

The experimental configuration of the DistilBERT-Base Turkish Cased Model [40] was meticulously designed to optimize its performance in the undertaken tasks. While keeping most hyperparameters at their default values as per the model's documentation [40], critical adjustments were made. The maximum sequence length was set to 512 to accommodate varying input lengths, while a batch size of 8 was chosen for computational efficiency. The training was conducted over 3 epochs to explore learning dynamics.

3.2.4 XLM-RoBERTa

XLM-RoBERTa, introduced by Conneau et al. [41] and initially unveiled in [42], adopts a structure aligned with the cross-lingual language model (XLM) [43] princi-

ples by using monolingual data. In this procedure, a sentence is subjected to a random masking of 15% of its words. Subsequently, the entire masked sentence is processed through the model, which is tasked with predicting the masked words.

To elaborate further, the model is pre-trained on a large corpus across 100 languages in a self-supervised approach. It is noteworthy that the authors emphasized the potential for multilingual modeling without compromising individual language performance.

This approach facilitates the acquisition of an inner representation encompassing 100 languages within the model, subsequently harnessed for extracting features for downstream tasks.

In this thesis work, XLM-RoBERTa-Base Multilingual Model [44] was fine-tuned on variations of TurkishSportsNER data set.

In the experimental phase, key hyper-parameters for the XLM-RoBERTa-Base Multilingual Model [44] were thoughtfully configured. While the default settings were adhered to for most hyper-parameters as recommended by the model's configuration [44], strategic adjustments were made to ensure optimal performance. The maximum sequence length was set to 512 to accommodate varying text lengths effectively. With a batch size of 8, a balance between computational efficiency and analytical accuracy was achieved. The training spanned 3 epochs, striking a balance between capturing learning patterns and avoiding over-fitting. These deliberate hyper-parameter choices laid a solid foundation for evaluating the model's effectiveness in the context of this thesis work.

3.3 Explainability Methods

3.3.1 LIME

Local Interpretable Model-Agnostic Explanations (LIME), proposed by Piberio et al. [45] is a method that provides an approximation for explaining individual predictions of any black box machine learning model.

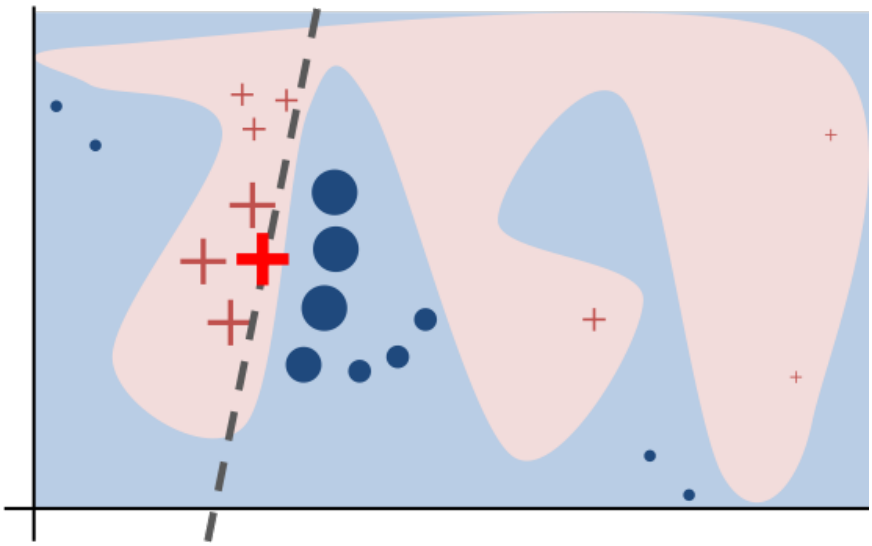


Figure 3.5: LIME Process. Sourced by [3]

An illustration of the implementation of LIME was given in 3.5 “The original model’s decision function is represented by the blue/pink background, and is clearly nonlinear. The bright red cross is the instance being explained (let’s call it X). We sample perturbed instances around X, and weight them according to their proximity to X (weight here is represented by size). We get original model’s prediction on these perturbed instances, and then learn a linear model (dashed line) that approximates the model well in the vicinity of X. Note that the explanation in this case is not faithful globally, but it is faithful locally around X.”

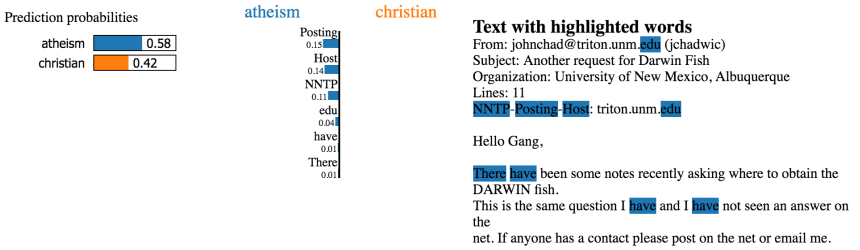


Figure 3.6: Explanation example of a text classification by LIME. Sourced by [3]

LIME extracts the list of tokens that impact the prediction of the model. Example output of LIME on text classification was given in 3.6. This methodology facilitates the identification of words that play a pivotal role in shaping the model’s prediction.

The LIME algorithm aims to create an interpretable model that can approximate the

behavior of a complex, non-interpretable machine learning model (referred to as the 'black-box' model) for individual predictions. It does this by sampling and analyzing the local space around a given instance. The core idea is to gain insights into why a black-box model makes certain predictions. The algorithm of LIME according to [45] is as follows:

- **Sampling for Local Exploration**: The algorithm begins by generating a new data set around a specific instance x that needs explanation. This is achieved by drawing perturbed samples z that are variations of x. These perturbations involve tweaking nonzero elements of x randomly to explore the local behavior of the black-box model f around x.

- **Defining the Loss Function**: LIME utilizes a locally weighted square loss function defined as:

$$L(f, g, \Pi_x) = \sum_{z,z' \in Z} \Pi_x(z) \times (f(z) - g(z'))^2 \tag{3.1}$$

  3.2, $f(z)$ is the prediction of the black-box model on the perturbed sample z, $g(z)$ is the prediction of the interpretable model, and $\Pi_x(z)$ is a similarity kernel, typically an exponential kernel, weighting the samples based on their proximity to the original instance x.

- **Creating an Interpretable Model**: The objective is to find an interpretable model g from a class of models G that minimizes the loss function $L(f, g, \Pi_x)$ while also keeping the complexity of g low. This is expressed as:

$$\xi(x) = argmin_{g \in G}[L(f, g, \Pi_x) + \Omega(g)] \tag{3.2}$$

  The interpretable model is trained using the generated data set to closely approximate the black-box model's behavior near x.

- **Algorithm Execution**: LIME requires the original classifier f, the number of samples N, the instance x, its interpretable version, the similarity kernel Π_x, and the length of explanation K. It iteratively samples around x, and computes the similarity weight and the label using the black-box model, forming the data set Z. A sparse linear model is then trained using techniques like Lasso with Z as the data set to obtain the explanation weights w.

This thesis work leveraged the eli5 library in Python to facilitate the implementation of the LIME method. The practical execution was conducted within the Google Colab environment, benefiting from its high RAM resources. The Python version 3.10 was employed to ensure compatibility. The transformers package, version 4.27.4, was employed to load saved models effectively. The implementation process encompassed several key steps. Initial steps involved loading the pre-trained model and its corresponding tokenizer. Subsequently, a coherent pipeline was constructed by integrating the model and tokenizer components. The MaskingTextSampler from the eli5 package was instrumental in generating a comprehensive set of samples through text masking techniques. The utilization of this sampler contributed to an enhanced understanding of the model's prediction behaviors. Following the sampler's creation, a crucial step involved the instantiation of the TextExplainer, a pivotal component of LIME methodology. This TextExplainer was configured with a setting n_samples=200. Subsequently, the TextExplainer was re-fitted with the prediction probability function to align with the model's prediction process. Finally, the explain_prediction function was employed to present the contributions attributed to individual words. Across all experimental scenarios, the parameter top_targets remained consistent at a value of 5, facilitating an analysis of the model's performance across multiple target labels.

3.3.2 SHAP

SHapley Additive exPlanations (SHAP) is a method introduced by Lundberg and Lee [46] that offers a unified framework for interpreting the predictions of machine learning models. Rooted in cooperative game theory, SHAP provides a principled approach to attributing the contribution of each feature to a prediction. Through the computation of Shapley values, which quantify the average incremental contribution of each feature across all conceivable feature permutations, SHAP ensures an equitable and consistent apportionment of the prediction's value among the contributing features.

SHAP values hold the property of local accuracy, ensuring that the sum of the attributions matches the difference between the model's prediction and the expected value. This enables SHAP to provide individual instance-level explanations and global fea-

ture importance rankings. The method's versatility extends to various types of models, including complex ensemble models, and it has found application in diverse domains such as text classification, image recognition, and healthcare. SHAP's comprehensive and intuitive framework contributes to enhancing the interpretability of machine learning models and fostering a deeper understanding of their decision-making processes.

SHAP is designed to explain the predictions made by machine learning models by assigning importance values to each feature involved in the prediction. This approach is especially crucial in scenarios where the rationale behind a model's decisions is just as important as its accuracy. The foundation of SHAP lies in game theory, more specifically in the concept of Shapley values. SHAP employs a game-theory to quantify each participant's role in the final result. In the context of machine learning, every feature is given a significance score that reflects its impact on the model's output. This approach leads to the creation of a novel class of additive feature importance measures. Features that exhibit positive SHAP values have positive contribution on the prediction, whereas those with negative SHAP values have negative contribution. The size of these values indicates the strength of their impact. These measures stand out due to their adherence to three principal properties: local accuracy, missingness, and consistency.

- Local Accuracy: This property guarantees that the sum of SHAP values for a given prediction, combined with a base value, precisely equals the model's output.
- Missingness: This principle states that if a feature is not present in the input, it should have no impact on the prediction, ensuring the integrity of the model's decision-making process.
- Consistency: This ensures that the importance attributed to a feature does not decrease if the model is altered in a way that either increases or maintains that feature's contribution to the prediction.

SHAP's innovative approach includes the unification of six distinct model interpretation methods, including LIME. This amalgamation simplifies the process of un-

derstanding model predictions, providing a comprehensive framework that integrates various interpretation approaches. The SHAP methodology is applicable to a broad spectrum of models, including those with complex architectures like deep neural networks. From a computational perspective, estimating SHAP values involves both model-specific and model-agnostic methods. [46] delves into the computational challenges associated with these estimations and proposes efficient solutions for calculating SHAP values in practical scenarios. An important aspect highlighted in [46] is the alignment of SHAP values with human intuition. This characteristic is essential in applications where understanding and trusting the predictions made by machine learning models is critical for effective decision-making.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left(f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right) \tag{3.3}$$

In the Equation 3.3 in [46], F represents the set of all features, and S is a subset of features excluding i. $|S|$ is the number of features in set S. $|F|$ is the total number of features. For the purpose of evaluating the impact of feature i, two distinct models are trained: $f_{S \cup \{i\}}$ and f_S. f_S is the model trained without feature i. This approach facilitates the comparative analysis of the models' predictions with respect to the presence or absence of feature i. The differential impact of feature i is quantified by comparing the predictions of these two models on a given input. This comparison is represented as $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$, where $x_{S \cup \{i\}}$ and x_S represent the values of the input features in the sets $S \cup \{i\}$ and S, respectively. This formula calculates the average marginal contribution of feature i across all possible combinations of features, giving a measure of its importance in the model's prediction.

In a scenario with a BERT model trained for Named Entity Recognition (NER), aiming to identify entities like PERSON, ORGANIZATION, and LOCATION, in a sentence, SHAP values can provide significant insights. Consider a sentence example: "Alice and Bob visited Paris last week." We aim to understand how the BERT model identifies "Paris" as a LOCATION. Initially, the BERT model tokenizes the sentence into tokens ["Alice", "and", "Bob", "visited", "Paris", "last", "week"], and each token is transformed into a feature vector using BERT's embedding layers. After the model predicts entity labels for each token and correctly labels "Paris" as LOCATION, we

apply a SHAP explainer. The SHAP explainer assesses the impact of each token (and their combinations) on the model's output by calculating the marginal contribution of each token towards labeling "Paris" as LOCATION. It does this by evaluating the model's output with various combinations of tokens present and absent, thereby determining how each token shifts the model's prediction.

In this thesis, the SHAP method is implemented using the shap library in Python. The experiments were conducted within the Google Colab environment, which offers robust GPU resources. The Python version used is 3.10, and the transformers package version is 4.35.2, which aids in efficiently loading saved models. The implementation process encompasses several steps. It begins with loading a pre-trained model and its associated tokenizer. Next, a function is developed to compute the probabilities of named entities. After that, an explainer is then created using the shap.Explainer method. Finally, to visualize the results, bar charts and text plots are generated using shap.plots.bar and shap.plots.text methods, respectively.

CHAPTER 4

EXPERIMENTS AND RESULTS

4.1 Information About Data Set

As explained in Section 3.1.1, the TurkishSportsNER data set was constructed by manually tagging articles from a Turkish newspaper internet page Milliyet.com. For the experiments, the TurkishSportsNER data set was first split into 5 splits. Each split was divided in 3 parts: train, development, and test. The train part for each partition contains 80% of the data from the partition, the test part for each partition contains 20% of the data from the partition, and the development part is the same for each partition and contains approximately 20% of the data.

The number of named entity tags for each split of TurkishSportsNER for different versions of the data set is represented in the tables below.

- TurkishSportsNER-Parent-NonBIO is the most general data set. It contains only parent named entities without BIO formatting. The number of labels is represented in Table 4.1.
- TurkishSportsNER-Parent-BIO is the second general data set. It contains only parent named entities with BIO formatting. The number of labels is represented in Table 4.2.
- TurkishSportsNER-Child-NonBIO is the second specific data set. It contains only child named entities without BIO formatting. The number of labels is represented in Table 4.3.
- TurkishSportsNER-Child-BIO is the most specific data set. It contains only

child named entities with BIO formatting. The number of labels is represented in Table 4.4 and Table 4.5.

Table 4.1: Number of Named Entity Tags For Each Split of TurkishSportsNER-Parent-NonBIO Data Set

Splits	**k=1**		**k=2**		**k=3**		**k=4**		**k=5**	
Labels	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**
AGREEMENT	2608	385	2037	956	2681	312	2377	616	2269	724
ALIAS	2617	577	2500	694	2694	500	2537	657	2428	766
EQUIPMENT	1886	355	1911	330	1583	658	1690	551	1894	347
HEALTH_TERM	504	196	561	139	554	146	587	113	594	106
NATION	1621	355	1614	362	1647	329	1455	521	1567	409
ROLE	12462	3444	12256	3650	13342	2564	13011	2895	12553	3353
SPORT_COMPETITION	4066	898	3828	1136	3951	1013	3917	1047	4094	870
SPORT_EVENT	10106	2334	10062	2378	9753	2687	9716	2724	10123	2317
SPORT_LEVEL	2506	523	2396	633	2408	621	2345	684	2461	568
SPORT_LOCATION	2764	550	2795	519	2540	774	2601	713	2556	758
SPORT_NAME	2340	696	2484	552	2361	675	2435	601	2524	512
SPORT_ORGANIZATION	2303	735	2479	559	2503	535	2397	641	2470	568
SPORT_PERSON	26288	6867	25378	7777	26646	6509	27084	6071	27224	5931
SPORT_TERM	15721	2972	15667	3026	13398	5295	14086	4607	15900	2793
SUPPORTER	1329	381	1308	402	1441	269	1386	324	1376	334
TEAM	15886	3656	15267	4275	15826	3716	15457	4085	15732	3810
TEAM_SPONSOR	183	51	140	94	223	11	211	23	179	55
TOTAL	105190	24975	102683	27482	103551	26614	103292	26873	105944	24221
TOTAL_COUNT	456178	114071	456592	113658	456133	114117	455777	114473	456322	113928
TOTAL_NEWS_COUNT	634	150	611	173	628	156	635	149	631	158

Table 4.2: Number of Named Entity Tags For Each Split of TurkishSportsNER-Parent-BIO Data Set

Splits	**k=1**		**k=2**		**k=3**		**k=4**		**k=5**	
Labels	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**
B-AGREEMENT	2483	370	1927	926	2560	293	2278	575	2164	689
B-ALIAS	1560	336	1483	413	1576	320	1514	382	1451	445
B-EQUIPMENT	1855	339	1880	314	1546	648	1645	549	1850	344
B-HEALTH_TERM	495	170	533	132	519	146	552	113	561	104
B-NATION	1598	351	1593	356	1624	325	1434	515	1547	402
B-ROLE	10852	2874	10547	3179	11488	2238	11214	2512	10803	2923
B-SPORT_COMPETITION	1991	453	1922	522	1942	502	1911	533	2010	434
B-SPORT_EVENT	9468	2142	9399	2211	9085	2525	9053	2557	9435	2175
B-SPORT_LEVEL	1388	285	1327	346	1335	338	1285	388	1357	316
B-SPORT_LOCATION	2209	465	2250	424	2041	633	2115	558	2080	594
B-SPORT_NAME	2304	678	2436	546	2319	663	2396	586	2473	509
B-SPORT_ORGANIZATION	1680	470	1772	378	1771	379	1683	467	1694	456
B-SPORT_PERSON	19181	4761	18279	5663	18981	4961	19562	4380	19765	4177
B-SPORT_TERM	13148	2437	13091	2494	11204	4381	11730	3855	13167	2418
B-SUPPORTER	1316	368	1288	396	1419	265	1362	322	1351	333
B-TEAM	14195	3241	13839	3597	14000	3436	13689	3747	14021	3415
B-TEAM_SPONSOR	137	31	95	73	160	8	151	17	129	39
I-AGREEMENT	125	15	110	30	121	19	99	41	105	35
I-ALIAS	1057	241	1017	281	1118	180	1023	275	977	321
I-EQUIPMENT	31	16	31	16	37	10	45	2	44	3
I-HEALTH_TERM	9	26	28	7	35	0	35	0	33	2
I-NATION	23	4	21	6	23	4	21	6	20	7
I-ROLE	1610	570	1709	471	1854	326	1797	383	1750	430
I-SPORT_COMPETITION	2075	445	1906	614	2009	511	2006	514	2084	436
I-SPORT_EVENT	638	192	663	167	668	162	663	167	688	142
I-SPORT_LEVEL	1118	238	1069	287	1073	283	1060	296	1104	252
I-SPORT_LOCATION	555	85	545	95	499	141	485	155	476	164
I-SPORT_NAME	36	18	48	6	42	12	39	15	51	3
I-SPORT_ORGANIZATION	623	265	707	181	732	156	714	174	776	112
I-SPORT_PERSON	7107	2106	7099	2114	7665	1548	7522	1691	7459	1754
I-SPORT_TERM	2573	535	2576	532	2194	914	2356	752	2733	375
I-SUPPORTER	13	13	20	6	22	4	24	2	25	1
I-TEAM	1691	415	1428	678	1826	280	1768	338	1711	395
I-TEAM_SPONSOR	46	20	45	21	63	3	60	6	50	16
TOTAL	105190	24975	102683	27482	103551	26614	103292	26873	105944	24221
TOTAL_COUNT	456178	114071	456592	113658	456133	114117	455777	114473	456322	113928
TOTAL_NEWS_COUNT	634	150	611	173	628	156	635	149	631	153

Table 4.3: Number of Named Entity Tags For Each Split of TurkishSportsNER-Child-NonBIO Data Set

Splits	**k=1**		**k=2**		**k=3**		**k=4**		**k=5**	
Labels	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**
AGREEMENT	2608	385	2037	956	2681	312	2377	616	2269	724
BASKETBALL_TERM	260	71	218	113	292	39	264	67	290	41
COACH	5976	1394	5242	2128	6275	1095	6117	1253	5870	1500
EQUIPMENT	1886	355	1911	330	1583	658	1690	551	1894	347
FOOTBALL_TERM	12026	2277	12085	2218	10176	4127	10649	3654	12276	2027
HEALTH_TERM	504	196	561	139	554	146	587	113	594	106
HORSE_RACING_TERM	210	38	245	3	47	201	242	6	248	0
MANAGER	3183	1509	3754	938	4177	515	4100	592	3554	1138
NATION	1621	355	1614	362	1647	329	1455	521	1567	409
PERSOM_ALIAS	645	76	574	147	581	140	557	164	527	194
RACE_TERM	56	0	25	31	40	16	47	9	56	0
REFEREE	1101	341	1347	95	982	460	1097	345	1241	201
ROLE	12462	3444	12256	3650	13342	2564	13011	2895	12553	3353
SCORE_TERM	3147	575	3068	654	2823	899	2852	870	2998	724
SPORT_COMPETITION	4066	898	3828	1136	3951	1013	3917	1047	4094	870
SPORT_EVENT	10106	2334	10062	2378	9753	2687	9716	2724	10123	2317
SPORT_LEVEL	2506	523	2396	633	2408	621	2345	684	2461	568
SPORT_LOCATION	2764	550	2795	519	2540	774	2601	713	2556	758
SPORT_NAME	2340	696	2484	552	2361	675	2435	601	2524	512
SPORT_ORGANIZATION	2303	735	2479	559	2503	535	2397	641	2470	568
SPORT_PLAYER	16028	3623	15035	4616	15212	4439	15770	3881	16559	3092
SUPPORTER	1329	381	1308	402	1441	269	1386	324	1376	334
TEAM	15886	3656	15267	4275	15826	3716	15457	4085	15732	3810
TEAM_ALIAS	1972	501	1926	547	2113	360	1980	493	1901	572
TEAM_SPONSOR	183	51	140	94	223	11	211	23	179	55
TENNIS_TERM	12	0	12	0	0	12	12	0	12	0
VOLLEYBALL_TERM	10	11	14	7	20	1	20	1	20	1
TOTAL	105190	24975	102683	27482	103551	26614	103292	26873	105944	24221
TOTAL_COUNT	456178	114071	456592	113658	456133	114117	455777	114473	456322	113928
TOTAL_NEWS_COUNT	634	150	611	173	628	156	635	149	631	153

Table 4.4: Number of Named Entity Tags For Each Split of TurkishSportsNER-Child-BIO Data Set

Splits	**k=1**		**k=2**		**k=3**		**k=4**		**k=5**	
Labels	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**
B-AGREEMENT	2483	370	1927	926	2560	293	2278	575	2164	689
B-BASKETBALL_TERM	202	64	177	89	236	30	218	48	231	35
B-COACH	4126	973	3643	1456	4301	798	4249	850	4077	1022
B-EQUIPMENT	1855	339	1880	314	1546	648	1645	549	1850	344
B-FOOTBALL_TERM	9735	1791	9779	1747	8196	3330	8566	2960	9828	1698
B-HEALTH_TERM	495	170	533	132	519	146	552	113	561	104
B-HORSE_RACING_TERM	169	31	197	3	40	160	194	6	200	0
B-MANAGER	1940	841	2230	551	2466	315	2415	366	2073	708
B-NATION	1598	351	1593	356	1624	325	1434	515	1547	402
B-PERSON_ALIAS	492	61	439	114	427	126	444	109	410	143
B-RACE_TERM	45	0	23	22	31	14	36	9	45	0
B-REFEREE	594	194	726	52	536	252	610	178	676	112
B-ROLE	10852	2874	10547	3179	11488	2238	11214	2512	10803	2923
B-SCORE_TERM	2982	541	2896	627	2683	840	2692	831	2839	684
B-SPORT_COMPETITION	1991	453	1922	522	1942	502	1911	533	2010	434
B-SPORT_EVENT	9468	2142	9399	2211	9085	2525	9053	2557	9435	2175
B-SPORT_LEVEL	1388	285	1327	346	1335	338	1285	388	1357	316
B-SPORT_LOCATION	2209	465	2250	424	2041	633	2116	558	2080	594
B-SPORT_NAME	2304	678	2436	546	2319	663	2396	586	2473	509
B-SPORT_ORGANIZATION	1680	470	1772	378	1771	379	1683	467	1694	456
B-SPORT_PLAYER	12521	2753	11670	3604	11678	3596	12288	2986	12939	2335
B-SUPPORTER	1316	368	1288	396	1419	265	1362	322	1351	333
B-TEAM	14195	3241	13839+	3597	14000	3436	13689	3747	14021	3415
B-TEAM_ALIAS	1068	275	1044	299	1149	194	1070	273	1041	302
B-TEAM_SPONSOR	137	31	95	73	160	8	151	17	129	39
B-TENNIS_TERM	6	0	6	0	0	6	6	0	6	0
B-VOLLEYBALL_TERM	9	10	13	6	18	1	18	1	18	1
I-AGREEMENT	125	15	110	30	121	19	99	41	105	35
I-BASKETBALL_TERM	58	7	41	24	56	9	46	19	59	6
I-COACH	1850	421	1599	672	1974	297	1868	403	1793	478
I-EQUIPMENT	31	16	31	16	37	10	45	2	44	3
I-FOOTBALL_TERM	2291	486	2306	471	1980	797	2083	694	1448	329
I-HEALTH_TERM	9	26	28	7	35	0	35	0	33	2
I-HORSE_RACING_TERM	41	7	48	0	7	41	48	0	48	0
I-MANAGER	1243	668	1524	387	1711	200	1685	226	1481	430
I-NATION	23	4	21	6	23	4	21	6	20	7
I-PERSON_ALIAS	153	15	135	33	154	14	113	55	117	51

Table 4.5: Number of Named Entity Tags For Each Split of TurkishSportsNER-Child-BIO Data Set (Table 4.4 continued)

Splits	**k=1**		**k=2**		**k=3**		**k=4**		**k=5**	
Labels	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**	**train**	**test**
I-RACE_TERM	11	0	2	9	9	2	11	0	11	0
I-REFEREE	507	147	611	43	446	208	487	167	565	89
I-ROLE	1610	570	1709	471	1854	326	1797	383	1750	430
I-SCORE_TERM	165	34	172	27	140	59	160	39	159	40
I-SPORT_COMPETITION	2075	445	1906	614	2009	511	2006	514	2084	436
I-SPORT_EVENT	638	192	663	167	668	162	663	167	688	142
I-SPORT_LEVEL	1118	238	1069	287	1073	283	1060	296	1104	252
I-SPORT_LOCATION	555	85	545	95	499	141	485	155	476	164
I-SPORT_NAME	36	18	48	6	42	12	39	15	51	3
I-SPORT_ORGANIZATION	623	265	707	181	732	156	714	174	776	112
I-SPORT_PLAYER	3507	870	3365	1012	3534	843	3482	895	3620	757
I-SUPPORTER	13	13	20	6	22	4	24	2	25	1
I-TEAM	1691	415	1428	678	1826	280	1768	338	1711	395
I-TEAM_ALIAS	904	226	882	248	964	166	910	220	860	270
I-TEAM_SPONSOR	46	20	45	21	63	3	60	6	50	16
I-TENNIS_TERM	6	0	6	0	0	6	6	0	6	0
I-VOLLEYBALL_TERM	1	1	1	1	2	0	2	0	2	0
TOTAL	105190	24975	102683	27482	103551	26614	103292	26873	105944	24221
TOTAL_COUNT	456178	114071	456592	113658	456133	114117	455777	114473	456322	113928
TOTAL_NEWS_COUNT	634	150	611	173	628	156	635	149	631	153

4.2 Evaluation

In this study, the evaluation of the experiments was conducted using three important metrics: Precision, Recall, and F1-Score. These metrics are commonly used in the field of machine learning, deep learning, and information retrieval to assess the performance of a model or system.

- **Precision** measures the accuracy of the positive predictions made by a model. It is calculated as the number of true positive predictions (relevant retrieved instances) divided by the total number of instances that the model predicted as positive (all retrieved instances). In other words, it answers the question: "Of all the instances predicted as positive, how many are actually relevant?"
  $Precision = \frac{TruePositives}{TruePositives+FalsePositives}$

- **Recall**, also known as sensitivity or true positive rate, measures the ability of

the model to find all the relevant instances in the data set. It is calculated as the number of true positive predictions (relevant retrieved instances) divided by the total number of instances that are actually positive (all relevant instances). In other words, it answers the question: "Of all the relevant instances, how many did the model correctly identify?"

$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$

- **F1-Score** is the harmonic mean of Precision and Recall. It is used when both Precision and Recall are important, and it provides a balance between the two metrics. F1-Score gives equal weight to Precision and Recall and is useful when there is an uneven class distribution or when false positives and false negatives have different impacts.

  $F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$

Table 4.6: Sample Sentence With Ground Truth Label And Predicted Label

Words	**Ground Truth Label**	**Predicted Label**
fenerbahçe	TEAM	TEAM
beko	TEAM	TEAM_SPONSOR
ile	O	O
anadolu	TEAM	TEAM
efes	TEAM	TEAM
maçı	SPORT_EVENT	SPORT_EVENT
istanbul'da	O	TEAM
uzatmalara	SPORT_TERM	SPORT_TERM
gitti	O	O

Through the evaluation of the experiments using these metrics, researchers can gain a thorough comprehension of the model's performance with regard to its ability to correctly identify relevant instances (Precision) and capture all relevant instances (Recall). The F1-Score plays a crucial role in evaluating the balance between these two aspects, thereby offering an overall measure of the model's effectiveness.

As Precision, Recall, and F1-Score values increase, the model's performance improves. Researchers leverage these metrics to conduct comparisons among various

Table 4.7: Confusion Matrix for TEAM Label According to Example in Table 4.6

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	True Positive TP=3	False Negative FN=1
	Negative	False Positive FP=1	True Negative TN=0

models, algorithms, or configurations, aiming to ascertain the optimal performer for a specific task or data set. This enables them to make informed decisions about which approach to pursue and deploy in real-world scenarios.

In Table 4.6, an example was given. The first column of the table is the word in the sentence, the second column is the ground truth label (manually annotated), and the third column represents the predicted label by the model. For this example, although the model can predict *fenerbahçe, anadolu, efes* correctly as TEAM, *beko*, and *istanbul'da* were not predicted correctly. The model would have the following entity-level evaluation, for the TEAM label:

- Since *fenerbahçe, anadolu*, and *efes* instances were correctly predicted as TEAM, true positive value for TEAM label is 3.
- Since *beko* was incorrectly predicted as TEAM_SPONSOR while it should have been TEAM, false negative value is 1.
- Since *istanbul'da* was incorrectly predicted as TEAM while it should have been O, false positive value is 1.

The confusion matrix according to entity-level evaluation can be found in Table 4.7. According to Table 4.7, Precision, Recall, and F1-Scores can be calculated as below.

$$Precision = \frac{TruePositives=3}{TruePositives=3+FalsePositives=1} = \frac{3}{4} = 0.75$$
$$Recall = \frac{TruePositives=3}{TruePositives=3+FalseNegatives=1} = \frac{3}{4} = 0.75$$
$$F1 - Score = \frac{2*Precision=0.75*Recall=0.75}{Precision=0.75+Recall=0.75} = 0.75$$

In this study, the model-level evaluation metrics were derived by aggregating the entity-level true positive, false positive, and false negative values. Specifically, the sum of all entity-level true positive values contributes to the model-level true positive value, while the sum of all entity-level false positive values contributes to the model-level false positive value. Likewise, the sum of all entity-level false negative values contributes to the model-level false negative value.

By performing these calculations, the model-level evaluation metrics were obtained. The metrics described earlier, namely Precision, Recall, and F1-Score, were utilized for assessing all the experiments conducted in this study.

4.2.1 Environments and Package Versions of Experiments

In order to ensure that other researchers can replicate our experiments using the same software versions, reducing potential discrepancies due to version differences and facilitating direct comparisons between different studies, the software environment was specified and explained explicitly.

In this study, all the experiments were conducted using the Google Colab environment. The central processing unit (CPU) utilized for the experiments was an Intel(R) Xeon(R) CPU clocked at 2.20GHz. The CPU was a dual-core processor, offering computing capabilities on two independent cores. The study also made use of approximately 24 gigabytes (GB) of random-access memory (RAM) to support the computational requirements.

Furthermore, the experiments were conducted using specific versions of Python and libraries. In this study, the experiments were implemented using Python 3.10.

The following versions of external libraries were utilized:

- **transformers (v.4.27.4)** is an open-source library developed by Hugging Face, and it provides pre-trained models, tokenization, and utilities for natural language processing (NLP) tasks, such as text classification, language translation, and question-answering.
- **torch (v.1.13.1)** refers to PyTorch, an open-source machine learning framework

developed by Meta AI. PyTorch is widely used for building and training deep learning models, including neural networks.

In this study, all the models used for the experiments are available in the Hugging Face model repository. The source of each model can be found in the below sections. In order to optimize the model training process, three crucial hyperparameters were specifically set:

- **Maximum Sequence Length** determines the maximum number of tokens allowed in each input sequence
- **Batch Size** refers to the number of input samples that are processed together in parallel during each iteration of the training process.
- **Number of Epochs** represents a complete pass through the entire training data set. During each epoch, the model iteratively updates its parameters based on the training data to improve its performance.

4.3 Experiments with BERT-Base Multilingual Cased Model

BERT-Base Multilingual Cased Model [35][1] is a pre-trained model. This model was trained on the top 104 languages including Turkish and it contains the largest Wikipedia data sets according to [47]. For this experiment, maximum sequence length was set to 512, batch size was set to 8, and the number of epochs was set to 3. The other hyper-parameters were set to default according to the configuration of the model [35][1].

Table 4.8, Table 4.9, Table 4.10, and Table 4.11 present the evaluation results of the BERT-Base Multilingual Cased model on multiple test data sets for the variations of TurkishSportsNER. The model's performance is assessed based on precision, recall, f1-score, and loss values, and the evaluation is conducted for different data splits represented by k=1 to k=5, as well as the average performance across these splits. The highest results for each variation of the data set were marked in bold in all tables.

Table 4.8 represents the model's performance on the TurkishSportsNER-Parent-NonBIO

Table 4.8: Precision, Recall, F1-Score, Loss Of BERT-Base Multilingual Cased Model On TurkishSportsNER-Parent-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	82.55	91.80	86.93	0.20
k=2	89.73	93.19	91.42	0.11
k=3	90.22	92.17	91.18	0.11
k=4	91.11	93.19	92.14	0.10
k=5	89.81	92.06	90.92	0.11
Average	88.68	92.48	90.51	0.12

Table 4.9: Precision, Recall, F1-Score, Loss Of BERT-Base Multilingual Cased Model On TurkishSportsNER-Parent-BIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	79.41	89.30	84.06	0.22
k=2	87.44	91.33	89.34	0.13
k=3	88.06	89.82	88.93	0.12
k=4	89.65	90.62	90.13	0.11
k=5	88.14	89.78	88.95	0.12
Average	86.54	90.17	88.28	0.14

Table 4.10: Precision, Recall, F1-Score, Loss Of BERT-Base Multilingual Cased Model On TurkishSportsNER-Child-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	80.30	88.86	84.36	0.22
k=2	87.98	91.11	89.52	0.13
k=3	88.03	89.89	88.95	0.13
k=4	89.73	90.89	90.31	0.11
k=5	88.10	90.03	89.05	0.12
Average	86.82	90.15	88.43	0.14

Table 4.11: Precision, Recall, F1-Score, Loss Of BERT-Base Multilingual Cased Model On TurkishSportsNER-Child-BIO Test Data Set

Splits	Precision	Recall	F1-Score	Loss
k=1	77.43	87.03	81.95	0.25
k=2	85.61	89.33	87.43	0.16
k=3	85.64	88.24	86.92	0.15
k=4	88.28	89.50	88.88	0.13
k=5	86.60	88.50	87.54	0.14
Average	84.71	88.52	86.54	0.16

test data set. The results reveal that the model achieves promising precision values, ranging from 82.55% to 91.11%. Additionally, the recall values range from 91.80% to 93.19%. Consequently, the f1-scores lie between 86.93% and 92.14%, reflecting the overall effectiveness of the model in finding named entities. Moreover, the loss values are relatively low, ranging from 0.10 to 0.20. Although the second fold and the fourth fold have the same and highest recall scores, the model is more successful on the fourth fold according to precision and f1-score. The highest f1-score (92.14) belongs to the fourth fold. On the other hand, the average f1-score of the model on the TurkishSportsNER-Parent-NonBIO test data set is 90.51%.

The evaluation results of the model on TurkishSportsNER-Parent-BIO were given

in Table 4.9. The results demonstrate consistent performance, with precision values varying from 79.41% to 89.65%, recall ranging from 89.30% to 91.33%, and f1-scores between 84.06% and 90.13%. The loss values for this data set range from 0.11 to 0.22, which, is similar to Table 4.8.

In Table 4.10, the results of the model on the TurkishSportsNER-Child-NonBIO test data set. The precision values vary from 80.30% to 89.73%, recall ranges from 88.86% to 91.11%, and f1-scores are between 84.36% and 90.31%. The corresponding loss values for this data set range from 0.11 to 0.22.

Finally, Table 4.11 represents the model's performance on the TurkishSportsNER-Child-BIO test data set. The precision values range from 77.43% to 88.28%, recall varies from 87.03% to 89.50%, and f1-scores are between 81.95% and 88.88%. The loss values for this data set range from 0.13 to 0.25.

Comparing the scores of the nth-fold training performance of the model in general, it is observed that the model has the highest recognition performance on the fourth fold of the data set and has the significantly lowest performance on the first fold of the data set according to f1-scores on all tables.

Comparing the results according to annotation format (whether the data set was constructed by using BIO format or not), it is observed that the model has higher performance on non-BIO data sets than BIO data sets. Although the performance of the model on TurkishSportsNER-Child-NonBIO is slightly higher than TurkishSportsNER-Parent-BIO, the model on TurkishSportsNER-Child-NonBIO has approximately 2 points higher than the results on TurkishSportsNER-Child-BIO.

According to the nth-fold results, BERT-Base Multilingual Cased Model has the most successful NER results on TurkishSportsNER-Parent-NonBIO and has the lowest recognition performance on TurkishSportsNER-Child-BIO.

Average precision, recall, and f1-scores per named entity type for each data set are represented in the tables below. In general, we can say that the highest precision, recall, and f1-scores mostly belong to the O label for each split of each data set. Moreover, it is observed that BERT-Base Multilingual Cased Model has the highest recognition performance on the TurkishSportsNER-Parent-NonBIO data set. The in-

crease in labels decreases the performance of the model mostly. Also, the scores of each split are very close to each other, except k is equal to 1.

Table 4.12 represents average 5-fold precision, recall and f1-scores on TurkishSports-NER-Parent-NonBIO per named entity. In the context of our model's performance, it is observed that the average recall scores generally exceed the average precision scores. This suggests that the model is more inclined to identify true positives but with the caveat of a higher false positive rate. However, a particular anomaly arises with the TEAM_SPONSOR entity, where the precision score substantially surpasses the recall score. Such a disparity indicates a high level of accuracy when the model predicts an instance as a TEAM_SPONSOR, yet it simultaneously fails to recognize all actual instances of this entity.

Table 4.13 represents average 5-fold precision, recall and f1-scores on TurkishSports-NER-Parent-BIO per named entity. A notable deficiency in the model's performance is observed in accurately identifying entities predominantly prefixed with "I-". This issue is caused by an imbalance in the training data set, where there are far fewer examples of entities labeled with "I-" compared to those labeled with "B-". As a result, the model tends to misclassify entities starting with "I-" as those starting with "B-".

Table 4.14 represents average 5-fold precision, recall, and f1-scores on TurkishSports-NER-Child-NonBIO per named entity. The model often struggles to correctly identify child entities of SPORT_TERM, except for FOOTBALL_TERM. This problem mainly arises because common sports terms like "savunma" (defense) and "faul" are used across different sports, and the model doesn't have enough varied examples in its training data to learn the subtle differences between these terms in different sports contexts.

Table 4.15 represents average 5-fold precision, recall, and f1-scores on TurkishSports-NER-Child-BIO per named entity. In examining the performance of the model on both TurkishSportsNER-Parent-BIO and TurkishSportsNER-Child-NonBIO data sets, a consistent pattern emerges: the model fails to accurately identify child entities of SPORT_TERM, with the exception of FOOTBALL_TERM, and also struggles to recognize entities labeled with the "I-" prefix. The observed behavior of the model on the

TurkishSportsNER-Child-BIO data set can be attributed to the fact that this data set is a composite of the TurkishSportsNER-Parent-BIO and TurkishSportsNER-Child-NonBIO data sets, thereby inheriting the characteristics and tendencies of both.

Table 4.16 represents f1-scores on TurkishSportsNER-Parent-NonBIO per named entity. The f1-score of the ROLE (94.40) entity is significantly higher than other entities except for O. Although SPORT_PERSON, SPORT_TERM, TEAM entities have more instances in both train and test data sets, the closest f1-score is 93.34 that is about 2 points behind ROLE. Also, the lowest recognition performance belongs to TEAM_SPONSOR. It is obviously shown that the number of train instances is important to recognize these entities.

The f1-scores of the model per named entity on TurkishSportsNER-Parent-BIO are shown in Table 4.17. Since I-EQUIPMENT, I-HEALTH_TERM, I-SPORT_NAME, I-NATION, I-SUPPORTER, and I-TEAM_SPONSOR have less number of instances in the train data set, the model has failed to recognize these entities. B-SPORT_NAME (95.94) entity has the second highest f1-scores in this experiment since most of the names of the sports are one word. Comparing the results of TurkishSportsNER-Parent-NonBIO and TurkishSportsNER-Parent-BIO, the model cannot perform successfully for B-ROLE, I-ROLE entities. It is observed that using the BIO schema decreased the performance of the model.

Table 4.18 represents f1-scores on TurkishSportsNER-Child-NonBIO per named entity. It is observed that the model has failed for child entities of SPORT_TERM (BASKETBALL_TERM, HORSE_RACING_TERM, RACE_TERM, TENNIS_TERM, VOLLEYBALL_TERM) except FOOTBALL_TERM. In the TurkishSportsNER-Child-NonBIO data set, some of the sports terms are the same. However, they were tagged according to the context. For instance, *faul* can be tagged as FOOTBALL-_TERM or BASKETBALL_TERM. That's why, the model cannot recognize the child entities of sports terms. Similar to the performance of the model on the TurkishSports-NER-Parent-NonBIO data set, ROLE (95.33) has the second-highest f1-score.

The f1-scores of the model per named entity on TurkishSportsNER-Child-BIO are shown in Table 4.19. Similar to the performance of the model on the TurkishSports-NER-Child-NonBIO, the model has failed on sports terms excluding FOOTBALL_-

TERM. Furthermore, I-EQUIPMENT, I-NATION, I-SPORT_NAME, I-SUPPORTER, I-HEALTH_TERM, and I-TEAM_SPONSOR entities have significant roles in the low results of the model. On contrary of the performances of the model on other data sets, B-AGREEMENT (93.99) has the second highest f1-score for the TurkishSports-NER-Child-BIO data set.

In conclusion, the recognition performance of the model decreases when using BIO schema, and child entities. This is an expected result since the complexity of the data set is increased.

In conclusion, the BERT-Base Multilingual Cased Model demonstrates promising performance on the TurkishSportsNER data sets, particularly in recognizing parent-level non-BIO entities. The model demonstrates a high level of precision, recall, and f1-scores, indicating its effectiveness in accurately identifying named entities in the sports domain. However, its performance varies depending on the data set and the type of named entity. The model performs better on non-BIO data sets, suggesting that it benefits from the simplicity of the non-BIO annotation format. Additionally, it shows better recognition of parent-level entities compared to child-level entities, likely due to the increased complexity involved in identifying the latter. Notably, the model achieves its peak performance in the fourth fold of the data set, while the first fold exhibits the lowest performance, suggesting a correlation between data distribution and model performance variation. Nevertheless, additional fine-tuning and optimization are necessary to improve the model's recognition capabilities for a broader spectrum of named entities, especially those with limited occurrences in the data sets and child-level entities following the BIO schema. The model's performance remains relatively consistent across different data splits, with the highest recognition performance observed on the TurkishSportsNER-Parent-NonBIO data set and the lowest performance on TurkishSportsNER-Child-BIO. In general, the model's ability to recognize named entities declines when confronted with the intricacies of the BIO schema and child entities, as anticipated.

Table 4.12: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	98.64	97.45	98.04
AGREEMENT	86.47	94.65	90.16
ALIAS	88.50	89.70	89.06
EQUIPMENT	89.90	91.36	90.54
HEALTH_TERM	82.97	84.98	83.81
NATION	90.53	88.67	89.39
ROLE	92.44	96.46	94.40
SPORT_COMPETITION	88.13	92.28	90.13
SPORT_EVENT	90.95	95.24	92.99
SPORT_LEVEL	90.60	93.75	92.14
SPORT_LOCATION	86.54	92.55	89.26
SPORT_NAME	93.83	94.49	94.15
SPORT_ORGANIZATION	88.11	91.22	89.61
SPORT_PERSON	89.47	94.31	91.83
SPORT_TERM	83.45	88.67	85.70
SUPPORTER	92.27	90.29	91.23
TEAM	92.25	93.50	92.85
TEAM_SPONSOR	79.26	56.94	63.25

Table 4.13: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	98.64	97.57	98.10
B-AGREEMENT	87.02	95.44	90.80
B-ALIAS	85.33	87.09	86.08

Table 4.13 continued from previous page

B-EQUIPMENT	90.16	91.74	90.88
B-HEALTH_TERM	83.86	86.58	85.16
B-NATION	90.12	89.79	89.74
B-ROLE	92.19	95.28	93.71
B-SPORT_COMPETITION	85.77	88.39	87.03
B-SPORT_EVENT	91.55	94.39	92.87
B-SPORT_LEVEL	87.07	90.22	88.59
B-SPORT_LOCATION	85.91	91.58	88.46
B-SPORT_NAME	94.10	95.53	94.80
B-SPORT_ORGANIZATION	89.21	90.28	89.74
B-SPORT_PERSON	88.72	91.65	90.16
B-SPORT_TERM	83.36	88.93	85.76
B-SUPPORTER	91.33	90.25	90.75
B-TEAM	92.12	93.88	92.94
B-TEAM_SPONSOR	88.39	52.68	65.55
I-AGREEMENT	87.02	62.26	72.38
I-ALIAS	92.69	91.89	92.20
I-EQUIPMENT	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00
I-ROLE	81.98	91.68	86.52
I-SPORT_COMPETITION	87.44	93.46	90.34
I-SPORT_EVENT	69.35	74.07	71.35
I-SPORT_LEVEL	87.81	91.61	89.64
I-SPORT_LOCATION	73.35	78.64	75.27
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	76.52	83.83	79.79
I-SPORT_PERSON	83.80	90.55	87.02
I-SPORT_TERM	74.93	80.73	77.41
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	83.55	80.86	82.12

Table 4.13 continued from previous page

I-TEAM_SPONSOR	20.00	1.00	1.90

Table 4.14: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	98.59	97.58	98.08
AGREEMENT	86.97	93.92	90.06
BASKETBALL_TERM	31.67	8.03	12.78
COACH	84.88	85.01	84.90
EQUIPMENT	90.52	90.75	90.56
FOOTBALL_TERM	84.07	91.86	87.69
HEALTH_TERM	83.16	83.94	83.21
HORSE_RACING_TERM	5.35	10.00	6.97
MANAGER	76.43	84.88	80.25
NATION	89.62	88.70	88.92
PERSON_ALIAS	67.91	66.53	65.98
RACE_TERM	0.00	0.00	0.00
REFEREE	73.97	81.13	77.17
ROLE	92.19	96.25	94.17
SCORE_TERM	78.89	85.55	81.46
SPORT_COMPETITION	88.23	92.86	90.48
SPORT_EVENT	91.44	94.94	93.10
SPORT_LEVEL	90.72	93.48	92.06
SPORT_LOCATION	86.53	93.37	89.66
SPORT_NAME	93.70	94.46	94.06
SPORT_ORGANIZATION	86.82	91.47	89.04
SPORT_PLAYER	86.87	90.18	88.47
SUPPORTER	92.66	89.82	91.19
TEAM	92.36	93.50	92.90

Table 4.14 continued from previous page

TEAM_ALIAS	89.09	89.33	89.08
TEAM_SPONSOR	73.98	45.37	54.25
TENNIS_TERM	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00

Table 4.15: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
B-AGREEMENT	87.83	95.45	91.26
B-BASKETBALL_TERM	20.00	0.57	1.11
B-COACH	84.23	82.65	83.40
B-EQUIPMENT	89.97	91.09	90.49
B-FOOTBALL_TERM	83.75	92.21	87.62
B-HEALTH_TERM	83.19	86.35	84.68
B-HORSE_RACING_TERM	10.91	3.87	5.71
B-MANAGER	71.62	80.21	75.49
B-NATION	89.89	88.75	89.16
B-PERSON_ALIAS	69.75	68.00	67.30
B-RACE_TERM	0.00	0.00	0.00
B-REFEREE	69.35	74.14	71.36
B-ROLE	92.26	95.07	93.64
B-SCORE_TERM	76.99	85.88	80.62
B-SPORT_COMPETITION	84.68	87.06	85.83
B-SPORT_EVENT	91.52	94.79	93.08
B-SPORT_LEVEL	87.74	90.38	89.00
B-SPORT_LOCATION	85.55	91.13	88.08
B-SPORT_NAME	93.72	95.16	94.41
B-SPORT_ORGANIZATION	86.66	90.27	88.40
B-SPORT_PLAYER	84.93	87.67	86.23

Table 4.15 continued from previous page

B-SUPPORTER	91.85	89.76	90.77
B-TEAM	92.22	93.72	92.92
B-TEAM_ALIAS	88.24	88.23	88.09
B-TEAM_SPONSOR	78.18	39.78	52.22
B-TENNIS_TERM	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00
I-AGREEMENT	86.81	63.77	72.84
I-BASKETBALL_TERM	0.00	0.00	0.00
I-COACH	81.27	83.63	82.23
I-EQUIPMENT	0.00	0.00	0.00
I-FOOTBALL_TERM	76.31	85.58	80.48
I-HEALTH_TERM	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00
I-MANAGER	76.09	85.28	80.13
I-NATION	0.00	0.00	0.00
I-PERSON_ALIAS	70.37	53.02	57.67
I-RACE_TERM	0.00	0.00	0.00
I-REFEREE	71.11	81.87	75.58
I-ROLE	81.34	91.63	86.09
I-SCORE_TERM	72.98	50.28	57.42
I-SPORT_COMPETITION	86.92	92.92	89.80
I-SPORT_EVENT	69.42	72.27	70.37
I-SPORT_LEVEL	88.21	91.32	89.71
I-SPORT_LOCATION	73.71	75.67	73.93
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	76.10	84.48	79.93
I-SPORT_PLAYER	76.33	82.40	79.21
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	82.95	81.24	82.01
I-TEAM_ALIAS	88.56	90.62	89.33
I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.15 continued from previous page

I-TENNIS_TERM	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00
O	98.58	97.59	98.08

Table 4.16: F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.22	98.10	98.27	98.35	98.26	98.04
AGREEMENT	83.39	92.91	90.66	90.99	92.84	90.16
ALIAS	85.96	92.08	85.20	92.50	89.54	89.06
EQUIPMENT	85.83	89.81	91.15	92.73	93.18	90.54
HEALTH_TERM	72.63	85.09	89.40	83.95	87.96	83.81
NATION	91.11	88.48	89.25	91.07	87.05	89.39
ROLE	91.98	94.84	94.71	95.40	95.08	94.40
SPORT_COMPETITION	86.91	91.86	90.83	89.52	91.55	90.13
SPORT_EVENT	88.33	93.94	93.40	94.66	94.64	92.99
SPORT_LEVEL	87.64	94.08	92.90	93.41	92.67	92.14
SPORT_LOCATION	82.78	90.33	89.90	91.98	91.30	89.26
SPORT_NAME	92.97	95.58	92.32	95.00	94.86	94.15
SPORT_ORGANIZATION	87.60	87.16	91.73	92.11	89.44	89.61
SPORT_PERSON	91.14	92.51	93.51	92.60	89.36	91.83
SPORT_TERM	74.59	85.86	87.21	91.12	89.69	85.70
SUPPORTER	85.22	93.73	93.09	90.73	93.37	91.23
TEAM	91.36	93.36	93.50	93.34	92.67	92.85
TEAM_SPONSOR	62.40	81.48	57.14	58.82	56.41	63.25

Table 4.17: F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.23	98.16	98.35	98.44	98.33	98.10
B-AGREEMENT	83.33	92.88	90.47	92.95	94.35	90.80
B-ALIAS	82.25	89.71	82.70	90.23	85.48	86.08
B-EQUIPMENT	84.55	91.75	91.40	93.30	93.39	90.88
B-HEALTH_TERM	75.86	84.17	91.28	87.39	87.08	85.16
B-NATION	90.91	88.71	89.39	92.14	87.53	89.74
B-ROLE	90.35	94.44	93.95	94.66	95.14	93.71
B-SPORT_COMPETITION	83.32	89.69	85.63	87.07	89.46	87.03
B-SPORT_EVENT	88.38	93.38	93.74	94.44	94.43	92.87
B-SPORT_LEVEL	83.61	90.52	90.04	89.09	89.70	88.59
B-SPORT_LOCATION	82.26	89.80	88.36	92.25	89.65	88.46
B-SPORT_NAME	93.29	95.86	93.79	95.94	95.11	94.80
B-SPORT_ORGANIZATION	84.56	88.92	92.11	91.53	91.57	89.74
B-SPORT_PERSON	89.89	91.20	91.65	91.07	86.97	90.16
B-SPORT_TERM	74.47	85.93	87.47	90.80	90.11	85.76
B-SUPPORTER	83.83	93.44	92.45	90.65	93.37	90.75
B-TEAM	91.28	92.78	93.97	93.46	93.22	92.94
B-TEAM_SPONSOR	59.57	80.31	46.15	78.57	63.16	65.55
I-AGREEMENT	78.57	66.67	77.42	71.43	67.80	72.38
I-ALIAS	90.09	95.76	88.36	94.01	92.79	92.20
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	83.30	85.26	86.20	88.09	89.73	86.52
I-SPORT_COMPETITION	87.66	91.05	90.58	90.08	92.31	90.34
I-SPORT_EVENT	56.99	76.11	68.87	78.19	76.58	71.35
I-SPORT_LEVEL	85.31	90.69	91.20	91.33	89.66	89.64
I-SPORT_LOCATION	65.92	77.83	68.27	84.35	80.00	75.27

Table 4.17 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	75.05	77.33	84.85	83.61	78.13	79.79
I-SPORT_PERSON	87.06	87.61	88.08	87.77	84.60	87.02
I-SPORT_TERM	61.96	81.57	79.73	82.99	80.82	77.41
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	74.81	88.79	84.07	82.73	80.20	82.12
I-TEAM_SPONSOR	9.52	0.00	0.00	0.00	0.00	1.90

Table 4.18: F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.31	98.10	98.30	98.41	98.30	98.08
AGREEMENT	82.44	92.53	90.74	91.53	93.04	90.06
BASKETBALL_TERM	0.00	0.00	37.50	0.00	26.42	12.78
COACH	84.65	89.72	79.77	87.47	82.86	84.90
EQUIPMENT	84.76	89.14	91.49	93.13	94.29	90.56
FOOTBALL_TERM	79.55	86.40	89.84	91.54	91.11	87.69
HEALTH_TERM	71.91	83.77	90.43	85.25	84.68	83.21
HORSE_RACING_TERM	34.86	0.00	0.00	0.00	0.00	6.97
MANAGER	82.30	78.03	78.43	82.75	79.76	80.25
NATION	90.73	87.12	88.08	92.19	86.48	88.92
PERSON_ALIAS	34.82	70.63	67.19	74.83	82.44	65.98
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	80.86	81.77	75.51	74.58	73.14	77.17
ROLE	91.44	94.68	94.39	95.33	95.01	94.17
SCORE_TERM	65.39	83.65	83.21	87.33	87.72	81.46
SPORT_COMPETITION	86.75	92.88	90.67	89.68	92.43	90.48
SPORT_EVENT	88.26	94.12	93.93	94.45	94.73	93.10
SPORT_LEVEL	87.97	93.67	92.77	92.83	93.09	92.06

Table 4.18 Continued

SPORT_LOCATION	82.13	90.08	90.59	93.01	92.49	89.66
SPORT_NAME	92.81	95.26	91.95	95.25	95.04	94.06
SPORT_ORGANIZATION	84.96	87.28	91.96	90.94	90.07	89.04
SPORT_PLAYER	86.61	89.66	89.62	90.38	86.09	88.47
SUPPORTER	84.88	94.03	93.38	90.18	93.47	91.19
TEAM	90.90	93.62	93.65	93.47	92.86	92.90
TEAM_ALIAS	83.24	92.38	87.04	93.41	89.30	89.08
TEAM_SPONSOR	38.81	70.34	50.00	54.05	58.06	54.25
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.19: F1-Scores of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.25	98.12	98.32	98.42	98.31	98.08
B-AGREEMENT	84.25	92.71	91.35	93.99	93.98	91.26
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	5.56	1.11
B-COACH	80.72	88.29	81.09	85.63	81.29	83.40
B-EQUIPMENT	83.31	91.45	91.09	92.35	94.24	90.49
B-FOOTBALL_TERM	79.90	85.47	90.03	91.63	91.05	87.62
B-HEALTH_TERM	76.16	84.05	89.11	88.24	85.85	84.68
B-HORSE_RACING_TERM	28.57	0.00	0.00	0.00	0.00	5.71
B-MANAGER	78.67	76.82	71.27	78.02	72.65	75.49
B-NATION	90.73	87.98	90.20	90.58	86.33	89.16
B-PERSON_ALIAS	40.70	69.52	69.64	77.49	79.13	67.30
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	79.12	71.79	69.98	70.08	65.82	71.36
B-ROLE	90.49	94.02	93.95	94.66	95.09	93.64
B-SCORE_TERM	64.14	83.38	81.85	86.52	87.23	80.62

Table 4.19 Continued

B-SPORT_COMPETITION	82.88	85.99	83.70	87.04	89.56	85.83
B-SPORT_EVENT	88.59	93.51	94.00	94.32	94.97	93.08
B-SPORT_LEVEL	84.16	91.17	90.28	89.88	89.49	89.00
B-SPORT_LOCATION	81.87	89.79	88.77	91.32	88.68	88.08
B-SPORT_NAME	91.57	95.79	94.21	95.95	94.54	94.41
B-SPORT_ORGANIZATION	84.23	86.85	89.41	91.18	90.34	88.40
B-SPORT_PLAYER	83.76	87.52	88.58	87.84	83.46	86.23
B-SUPPORTER	84.17	92.88	92.54	89.89	94.37	90.77
B-TEAM	91.22	93.04	93.71	93.50	93.13	92.92
B-TEAM_ALIAS	82.72	90.91	86.58	91.88	88.34	88.09
B-TEAM_SPONSOR	47.62	76.27	0.00	74.07	63.16	52.22
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	77.42	72.00	77.42	69.57	67.80	72.84
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	81.86	85.18	75.80	84.81	83.52	82.23
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	65.78	86.08	82.90	83.36	84.30	80.48
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	83.22	79.85	79.91	79.43	78.24	80.13
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	19.23	64.41	60.87	58.97	84.85	57.67
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	78.62	76.47	76.73	75.00	71.09	75.58
I-ROLE	83.60	83.76	86.12	87.35	89.63	86.09
I-SCORE_TERM	57.14	70.59	55.91	51.72	51.72	57.42
I-SPORT_COMPETITION	87.37	90.05	89.94	89.93	91.72	89.80
I-SPORT_EVENT	56.18	77.19	68.44	75.35	74.69	70.37
I-SPORT_LEVEL	86.42	90.29	91.07	90.73	90.06	89.71

Table 4.19 Continued

I-SPORT_LOCATION	61.96	79.63	69.17	81.90	77.01	73.93
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	75.00	78.42	83.18	82.87	80.16	79.93
I-SPORT_PLAYER	77.20	79.38	76.44	82.36	80.65	79.21
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	76.23	88.86	82.94	82.50	79.55	82.01
I-TEAM_ALIAS	83.57	92.22	87.89	91.07	91.89	89.33
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

4.4 Experiments with BERT-Base Turkish Uncased Model

Table 4.20: Precision, Recall, F1-Score, Loss Of BERT-Base Turkish Uncased Model On TurkishSportsNER-Parent-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	83.21	92.77	87.73	0.18
k=2	91.60	95.18	93.36	0.08
k=3	91.76	93.95	92.84	0.08
k=4	92.57	94.39	93.47	0.08
k=5	90.92	93.80	92.34	0.09
Average	90.01	94.01	91.94	0.10

Table 4.21: Precision, Recall, F1-Score, Loss Of BERT-Base Turkish Uncased Model On TurkishSportsNER-Parent-BIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	81.04	90.85	85.67	0.20
k=2	89.47	93.53	91.45	0.09
k=3	89.53	91.98	90.74	0.10
k=4	91.04	92.29	91.66	0.09
k=5	89.43	91.41	90.41	0.10
Average	88.10	92.01	89.98	0.11

Table 4.22: Precision, Recall, F1-Score, Loss Of BERT-Base Turkish Uncased Model On TurkishSportsNER-Child-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	82.55	90.73	86.44	0.19
k=2	89.74	93.06	91.37	0.09
k=3	89.35	91.35	90.34	0.10
k=4	91.33	92.68	92.00	0.09
k=5	89.22	91.53	90.36	0.10
Average	88.43	91.87	90.10	0.11

Table 4.23: Precision, Recall, F1-Score, Loss Of BERT-Base Turkish Uncased Model On TurkishSportsNER-Child-BIO Test Data Set

Splits	Precision	Recall	F1-Score	Loss
k=1	79.53	89.59	84.26	0.22
k=2	88.26	92.14	90.15	0.11
k=3	88.04	90.30	89.16	0.12
k=4	90.03	91.10	90.56	0.10
k=5	88.11	90.06	89.07	0.12
Average	86.79	90.63	88.64	0.13

BERT-Base Turkish Uncased Model [20] is a pre-trained model. This model was trained on Turkish OSCAR corpus [48], Wikipedia dump, OPUS corpora [49], and a special corpus. For this experiment, the maximum sequence length was set to 512, batch size was set to 8, and the number of epochs was set to 3. The other hyper-parameters were set to default according to the configuration of the model [20].

The results of the experiments of the BERT-Base Turkish Uncased model on several test data sets for the varieties of TurkishSportsNER are shown in Table 4.20, Table 4.21, Table 4.22, and Table 4.23. Precision, recall, f1-score, and loss values are used

to measure the model's performance. Experiments are done for several data splits, denoted by k=1 to k=5, as well as the average performance over these splits. In all tables, the highest results for each variation of the data set were marked in bold in all tables.

The performance of the BERT-Base Turkish Uncased model on TurkishSportsNER-Parent-NonBIO represents in Table 4.20. The f1-score ranges from 87.73% to 93.36% across different splits, with an average f1-score of 91.94%. Notably, both Precision and Recall values demonstrate consistently high values, affirming the model's proficiency in effectively identifying named entities within the sports domain. The model has the highest f1-score on the fourth split (93.44%).

Table 4.21 represents the evaluation results of the model on TurkishSportsNER-Parent-BIO. The f1-score fluctuates from 85.67% to 91.66% across different splits, and the average f1-score stands at 89.98%. On the other hand, recall ranges from 90.85% to 93.53%. It indicates that the model tends to retrieve a higher number of relevant instances compared to the number of incorrect instances it retrieved.

The evaluation results of the model on TurkishSportsNER-Child-NonBIO were given in Table 4.22. The precision values vary from 82.55% to 91.33%, recall ranges from 90.73% to 93.06%, and f1-scores are between 86.44% and 92.00% and the average f1-score is 90.10%. The lowest recognition performance of the model belongs to the first split, and the highest score belongs to the fourth split.

Finally, Table 4.23 indicates the performance of the model on the TurkishSportsNER-Child-BIO test data set. The precision values range from 79.53% to 90.03%, recall varies from 89.59% to 92.14%, and f1-scores are between 84.26% and 90.56%. The loss values for this data set range from 0.11 to 0.22.

When the f1-scores of the model's nth-fold training performance are compared across all tables, the model is more successful on the fourth fold of the data set and the lowest performance on the first split of the data set.

When the results are compared based on annotation format, the model performs better on non-BIO data sets than on BIO data sets. Since the model is trained on the Turkish data sets, there is not much difference between the lowest and highest average f1-

scores.

BERT-Base Turkish Uncased model has the best results on TurkishSportsNER-Parent-NonBIO and has the lowest recognition performance on TurkishSportsNER-Child-BIO according to the nth-fold results. In addition, the results of the experiments are relatively close for each split of the data set excluding the first fold.

In the below tables, average precision, recall, and f1-scores per named entity type for each data set on each split were given. The model, in general, has the most successful results for the Other (O) named entity. Furthermore, the highest recognition performance of the model in on the TurkishSportsNER-Parent-NonBIO data set. Mostly, the model's performance weakens as the number of labels rises.

In Table 4.24, average precision, recall, and f1-scores on TurkishSportsNER-Parent-NonBIO data set were given. The model performs better results for each named entity, but there may be specific challenges associated with accurately identifying TEAM_SPONSOR entities. The model's higher recall scores compared to precision scores indicate a tendency toward identifying a larger number of instances that belong to a particular label.

Table 4.25 represents average precision, recall, and f1-scores on TurkishSportsNER-Parent-BIO. The model faces challenges in accurately recognizing entities with the "I-" prefix, for example, I-HEALTH_TERM, I-NATION, and I-SUPPORTER. This challenge is caused by the lesser frequency of these entities in both the training and testing data sets.

Average precision, recall and f1-scores on TurkishSportsNER-Child-NonBIO are represented in Table 4.26. The model fails when identifying terms that belong to a branch of sports such as RACE_TERM, TENNIS_TERM, and VOLLEYBALL_TERM. Additionally, the model faces challenges with BASKETBALL_TERM and HORSE_-RACING_TERM. It shows that the model has difficulties when recognizing sports-related terminology within certain contexts.

In Table 4.27, average precision, recall, and f1-scores on TurkishSportsNER-Child-BIO were given. It is observed that the model has difficulties in accurately recognizing sports-related terms. Additionally, the model struggles with terms prefixed

with "I-". These observations are consistent with its performance patterns on both TurkishSportsNER-Parent-BIO and TurkishSportsNER-Child-BIO data sets.

In Table 4.28, f1-scores on TurkishSportsNER-Parent-NonBIO per named entity were given. Some named entity classes, such as NATION, TEAM, and SPORT_NAME, also show strong f1-scores, ranging from 91.25% to 96.98%. On the other hand, TEAM_SPONSOR presents the lowest f1-score on the third fold (26.67%). The variability in performance could be attributed to the complexity and ambiguity of terms within classes.

Table 4.29 represents f1-scores of the model per named entity on TurkishSportsNER-Parent-BIO. The model has failed to identify some of these entities such as I-NATION, I-EQUIPMENT, I-HEALTH_TERM, I-SUPPORTER, I-SPORT_NAME, and I-TEAM_SPONSOR. The second highest f1-score belongs to B-SPORT_NAME 96.47. Similar to the results on TurkishSportsNER-Parent-BIO, the model has failed on B-TEAM_SPONSOR and I-TEAM_SPONSOR named entities on the third split. It is observed that the model performed worse when the BIO schema was used.

The f1-scores of the model per named entity on TurkishSportsNER-Child-NonBIO were given in Table 4.30. Since the model is pre-trained on Turkish data, its ability to recognize terms according to the context is higher than BERT-Base Multilingual Cased model. As a result, some child entities of SPORT_TERM such as BASKETBALL_TERM and HORSE_RACING_TERM were recognized by the model. Also, NATION (97.46) entity has the second-highest f1-score on the fifth fold.

The f1-scores on TurkishSportsNER-Child-BIO per named entity were given in Table 4.31. Similar to the TurkishSportsNER-Child-NonBIO data set, B-NATION (97.06) named entity in the fifth fold has the second highest results after O. On the other hand, the model has failed on the I-NATION entity. It is important to bear to mind that the I-NATION entity is very less instances in the test data sets. Furthermore, the top 3 highest f1-scores belong to B-SPORT_NAME, B-TEAM, and I-TEAM_ALIAS.

In summary, the BERT-Base Turkish Uncased Model demonstrates promising performance on the TurkishSportsNER data sets, particularly in recognizing parent-level non-BIO entities. It achieves relatively high precision, recall, and f1-scores, indicat-

ing its proficiency in accurately identifying named entities within the sports domain. However, its performance varies across different data sets and named entity types. The model's superiority on non-BIO data sets compared to BIO data sets suggests an advantage in dealing with the simpler non-BIO annotation format. Additionally, it exhibits better recognition of parent-level entities compared to child-level entities, likely due to the increased complexity involved in identifying the latter. Notably, the model performs best in the fourth fold of the data set, while the first fold shows lower performance, suggesting a correlation between data distribution and model performance variation. However, further fine-tuning and optimization are required to enhance its recognition capabilities for a wider range of named entities, especially those with limited instances in the data sets and child-level entities under the BIO schema. The model's performance remains relatively consistent across different data splits, with the highest recognition performance observed on the TurkishSportsNER-Parent-NonBIO data set and the lowest on TurkishSportsNER-Child-BIO. Despite facing challenges in recognizing certain named entity types, such as TEAM_SPONSOR and I-NATION, the model exhibits remarkable proficiency in identifying entities like NATION, TEAM, and SPORT_NAME. Overall, the model's recognition performance diminishes when handling the complexity of the BIO schema and child entities, as anticipated.

Table 4.24: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
O	98.98	97.69	98.33
AGREEMENT	88.22	93.68	90.67
ALIAS	90.86	92.00	91.40
EQUIPMENT	90.55	93.09	91.73
HEALTH_TERM	81.53	86.02	83.44
NATION	91.66	96.65	94.06
ROLE	92.73	96.55	94.60
SPORT_COMPETITION	89.11	93.78	91.37

Table 4.24 continued from previous page

SPORT_EVENT	91.43	95.10	93.19
SPORT_LEVEL	89.90	93.53	91.66
SPORT_LOCATION	87.54	95.00	90.96
SPORT_NAME	95.27	95.44	95.34
SPORT_ORGANIZATION	88.80	92.85	90.73
SPORT_PERSON	92.69	96.41	94.51
SPORT_TERM	83.06	90.62	86.26
SUPPORTER	91.35	91.16	91.24
TEAM	93.89	95.40	94.62
TEAM_SPONSOR	70.07	66.28	66.27

Table 4.25: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	98.96	97.71	98.33
B-AGREEMENT	88.61	94.56	91.30
B-ALIAS	87.61	88.66	88.08
B-EQUIPMENT	90.09	92.95	91.42
B-HEALTH_TERM	83.55	87.64	85.45
B-NATION	91.36	97.66	94.38
B-ROLE	92.65	95.18	93.90
B-SPORT_COMPETITION	86.51	88.86	87.66
B-SPORT_EVENT	91.48	94.84	93.07
B-SPORT_LEVEL	88.59	90.36	89.42
B-SPORT_LOCATION	87.20	93.10	89.88
B-SPORT_NAME	95.29	96.48	95.87
B-SPORT_ORGANIZATION	89.15	92.96	90.97
B-SPORT_PERSON	91.65	94.31	92.96
B-SPORT_TERM	83.44	89.63	86.02

Table 4.25 continued from previous page

B-SUPPORTER	91.33	91.12	91.20
B-TEAM	93.60	95.67	94.59
B-TEAM_SPONSOR	75.65	50.05	59.42
I-AGREEMENT	81.78	63.49	71.01
I-ALIAS	93.97	94.04	93.94
I-EQUIPMENT	20.00	5.00	8.00
I-HEALTH_TERM	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00
I-ROLE	83.68	93.10	88.09
I-SPORT_COMPETITION	86.40	94.42	90.22
I-SPORT_EVENT	67.81	76.01	71.56
I-SPORT_LEVEL	86.62	93.98	90.14
I-SPORT_LOCATION	77.11	88.37	82.12
I-SPORT_NAME	20.00	3.33	5.71
I-SPORT_ORGANIZATION	78.18	87.78	82.22
I-SPORT_PERSON	87.61	94.48	90.84
I-SPORT_TERM	75.43	84.68	79.53
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	84.42	87.31	85.77
I-TEAM_SPONSOR	46.25	13.11	17.15

Table 4.26: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
O	98.93	97.79	98.35
AGREEMENT	88.56	93.44	90.73
BASKETBALL_TERM	64.25	46.18	50.88
COACH	87.44	89.31	88.32
EQUIPMENT	91.15	91.85	91.40

Table 4.26 continued from previous page

FOOTBALL_TERM	86.04	92.83	89.21
HEALTH_TERM	84.71	84.34	84.27
HORSE_RACING_TERM	21.71	20.34	5.74
MANAGER	79.78	87.61	83.22
NATION	92.30	96.89	94.52
PERSON_ALIAS	71.08	70.84	69.60
RACE_TERM	0.00	0.00	0.00
REFEREE	79.72	86.21	82.77
ROLE	92.79	96.45	94.58
SCORE_TERM	79.88	87.29	82.82
SPORT_COMPETITION	89.06	93.30	91.11
SPORT_EVENT	91.40	95.25	93.24
SPORT_LEVEL	90.07	93.90	91.94
SPORT_LOCATION	86.89	94.25	90.26
SPORT_NAME	94.18	95.34	94.74
SPORT_ORGANIZATION	88.29	93.04	90.52
SPORT_PLAYER	91.01	92.73	91.86
SUPPORTER	91.12	91.19	91.13
TEAM	93.70	95.55	94.59
TEAM_ALIAS	92.72	91.77	92.07
TEAM_SPONSOR	62.95	55.01	57.52
TENNIS_TERM	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00

Table 4.27: 5-Fold Average Precision, Recall and F1-Scores of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
B-AGREEMENT	88.34	94.60	91.18
B-BASKETBALL_TERM	53.68	25.74	30.35
B-COACH	87.17	89.09	88.09

Table 4.27 continued from previous page

B-EQUIPMENT	89.90	92.73	91.22
B-FOOTBALL_TERM	85.82	91.83	88.57
B-HEALTH_TERM	83.52	87.17	85.22
B-HORSE_RACING_TERM	1.34	19.35	2.50
B-MANAGER	79.16	84.53	81.50
B-NATION	91.32	97.42	94.25
B-PERSON_ALIAS	70.20	69.38	68.77
B-RACE_TERM	0.00	0.00	0.00
B-REFEREE	76.56	80.58	78.47
B-ROLE	92.29	94.96	93.60
B-SCORE_TERM	78.36	86.61	81.71
B-SPORT_COMPETITION	86.40	88.84	87.58
B-SPORT_EVENT	91.54	94.62	92.99
B-SPORT_LEVEL	87.94	89.49	88.63
B-SPORT_LOCATION	86.49	91.66	88.81
B-SPORT_NAME	94.75	96.23	95.47
B-SPORT_ORGANIZATION	89.87	92.48	91.13
B-SPORT_PLAYER	90.45	91.39	90.91
B-SUPPORTER	90.51	91.20	90.82
B-TEAM	93.80	95.59	94.65
B-TEAM_ALIAS	92.64	90.35	91.35
B-TEAM_SPONSOR	85.05	58.72	68.12
B-TENNIS_TERM	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00
I-AGREEMENT	81.83	63.97	71.41
I-BASKETBALL_TERM	0.00	0.00	0.00
I-COACH	85.45	89.45	87.28
I-EQUIPMENT	20.00	5.00	8.00
I-FOOTBALL_TERM	76.73	88.66	82.22
I-HEALTH_TERM	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00

Table 4.27 continued from previous page

I-MANAGER	78.68	88.33	82.86
I-NATION	0.00	0.00	0.00
I-PERSON_ALIAS	70.45	63.46	65.48
I-RACE_TERM	0.00	0.00	0.00
I-REFEREE	78.34	86.18	81.95
I-ROLE	82.10	92.56	86.93
I-SCORE_TERM	67.42	68.12	66.93
I-SPORT_COMPETITION	86.50	94.71	90.42
I-SPORT_EVENT	69.26	74.97	71.77
I-SPORT_LEVEL	86.19	94.05	89.94
I-SPORT_LOCATION	74.84	85.98	79.81
I-SPORT_NAME	20.00	3.33	5.71
I-SPORT_ORGANIZATION	77.86	88.02	82.28
I-SPORT_PLAYER	82.52	89.24	85.68
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	84.22	88.03	86.00
I-TEAM_ALIAS	93.33	92.85	92.90
I-TEAM_SPONSOR	48.33	15.17	22.78
I-TENNIS_TERM	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00
O	98.94	97.77	98.35

Table 4.28: F1-Scores of BERT-Base Turkish Uncased Model on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.35	98.56	98.55	98.67	98.52	98.33
AGREEMENT	84.64	93.77	91.30	92.01	91.61	90.67
ALIAS	87.97	92.43	88.40	94.33	93.84	91.40
EQUIPMENT	87.38	90.91	92.60	92.95	94.81	91.73
HEALTH_TERM	73.51	85.29	90.13	83.79	84.44	83.44

Table 4.28 Continued

NATION	94.10	92.64	91.25	95.35	96.98	94.06
ROLE	92.46	95.21	94.91	95.59	94.84	94.60
SPORT_COMPETITION	87.70	93.39	92.67	90.37	92.72	91.37
SPORT_EVENT	88.62	94.37	93.22	94.97	94.75	93.19
SPORT_LEVEL	86.94	94.44	92.59	92.70	91.64	91.66
SPORT_LOCATION	84.79	90.45	92.56	93.63	93.39	90.96
SPORT_NAME	94.80	95.91	94.97	95.39	95.62	95.34
SPORT_ORGANIZATION	87.12	90.29	94.11	91.31	90.82	90.73
SPORT_PERSON	93.78	96.09	95.87	95.12	91.68	94.51
SPORT_TERM	71.52	87.96	88.89	92.02	90.91	86.26
SUPPORTER	85.86	93.47	93.60	90.54	92.72	91.24
TEAM	93.47	95.01	95.57	95.26	93.79	94.62
TEAM_SPONSOR	70.27	85.26	26.67	78.43	70.71	66.27

Table 4.29: F1-Scores of BERT-Base Turkish Uncased Model on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.42	98.49	98.54	98.66	98.54	98.33
B-AGREEMENT	84.46	94.45	90.64	93.71	93.25	91.30
B-ALIAS	83.53	90.27	84.35	91.66	90.61	88.08
B-EQUIPMENT	85.91	91.27	92.47	93.65	93.78	91.42
B-HEALTH_TERM	77.95	85.93	89.63	90.08	83.64	85.45
B-NATION	93.70	93.14	92.17	95.24	97.66	94.38
B-ROLE	91.27	94.71	93.84	95.29	94.38	93.90
B-SPORT_COMPETITION	83.79	87.61	88.56	88.75	89.58	87.66
B-SPORT_EVENT	88.57	94.07	93.04	94.77	94.91	93.07
B-SPORT_LEVEL	83.08	93.71	91.91	90.58	87.80	89.42
B-SPORT_LOCATION	83.29	90.16	91.35	92.23	92.40	89.88
B-SPORT_NAME	94.84	96.94	95.47	96.47	95.62	95.87

Table 4.29 Continued

B-SPORT_ORGANIZATION	86.85	90.27	93.49	92.05	92.21	90.97
B-SPORT_PERSON	92.08	95.01	94.42	93.54	89.74	92.96
B-SPORT_TERM	73.23	86.71	88.44	91.51	90.23	86.02
B-SUPPORTER	84.41	93.94	92.94	90.65	94.07	91.20
B-TEAM	93.51	94.69	95.56	94.95	94.25	94.59
B-TEAM_SPONSOR	66.67	94.29	0.00	74.07	62.07	59.42
I-AGREEMENT	70.97	67.92	78.79	69.57	67.80	71.01
I-ALIAS	90.87	95.65	92.23	95.36	95.61	93.94
I-EQUIPMENT	40.00	0.00	0.00	0.00	0.00	8.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.17	88.76	87.52	90.39	89.63	88.09
I-SPORT_COMPETITION	86.99	91.08	92.16	88.42	92.46	90.22
I-SPORT_EVENT	56.53	77.13	69.82	76.06	78.23	71.56
I-SPORT_LEVEL	86.59	91.85	90.82	91.03	90.43	90.14
I-SPORT_LOCATION	74.75	83.57	79.42	87.73	85.14	82.12
I-SPORT_NAME	0.00	28.57	0.00	0.00	0.00	5.71
I-SPORT_ORGANIZATION	77.48	84.38	87.16	83.56	78.52	82.22
I-SPORT_PERSON	89.98	92.98	92.13	90.86	88.25	90.84
I-SPORT_TERM	62.68	84.41	81.36	85.01	84.17	79.53
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	80.50	89.86	87.68	86.99	83.85	85.77
I-TEAM_SPONSOR	50.00	24.00	0.00	0.00	11.76	17.15

Table 4.30: F1-Scores of BERT-Base Turkish Uncased Model on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.50	98.56	98.52	98.65	98.52	98.35
AGREEMENT	84.19	93.50	91.82	91.84	92.31	90.73

Table 4.30 Continued

BASKETBALL_TERM	57.35	28.99	55.26	59.13	53.66	50.88
COACH	89.02	92.55	82.34	90.52	87.20	88.32
EQUIPMENT	87.64	90.53	90.98	94.45	93.41	91.40
FOOTBALL_TERM	82.27	87.72	91.14	93.07	91.88	89.21
HEALTH_TERM	72.58	86.69	88.00	89.54	84.55	84.27
HORSE_RACING_TERM	15.69	0.00	13.02	0.00	0.00	5.74
MANAGER	84.05	86.34	82.76	85.02	77.94	83.22
NATION	94.12	93.24	92.33	95.45	97.46	94.52
PERSON_ALIAS	37.84	73.61	72.50	79.35	84.67	69.60
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	86.38	83.42	81.72	81.01	81.31	82.77
ROLE	92.31	95.24	95.04	95.59	94.74	94.58
SCORE_TERM	66.22	86.80	83.49	88.37	89.24	82.82
SPORT_COMPETITION	87.42	92.56	92.08	90.88	92.61	91.11
SPORT_EVENT	88.40	94.39	93.14	95.14	95.14	93.24
SPORT_LEVEL	87.60	95.14	91.81	93.06	92.07	91.94
SPORT_LOCATION	83.92	89.94	91.72	93.34	92.40	90.26
SPORT_NAME	94.01	95.58	93.28	95.67	95.14	94.74
SPORT_ORGANIZATION	86.90	89.46	93.47	92.28	90.46	90.52
SPORT_PLAYER	91.48	93.80	92.65	92.68	88.69	91.86
SUPPORTER	85.42	93.06	92.59	90.68	93.89	91.13
TEAM	93.52	94.76	95.34	95.15	94.19	94.59
TEAM_ALIAS	86.27	94.17	89.12	95.21	95.57	92.07
TEAM_SPONSOR	64.15	82.93	0.00	78.43	62.07	57.52
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.31: F1-Scores of BERT-Base Turkish Uncased Model on TurkishSportsNER-Child-BIO Data Set

Labels	k=1	k=2	k=3	k=4	k=5	Avg.
O	97.46	98.54	98.58	98.68	98.50	98.35
B-AGREEMENT	84.53	93.90	91.54	93.30	92.62	91.18
B-BASKETBALL_TERM	52.63	16.00	26.09	3.92	53.13	30.35
B-COACH	87.57	92.55	83.71	90.36	86.26	88.09
B-EQUIPMENT	86.31	91.30	92.27	93.30	92.93	91.22
B-FOOTBALL_TERM	82.18	86.76	91.01	92.32	90.55	88.57
B-HEALTH_TERM	75.88	87.60	88.29	90.08	84.26	85.22
B-HORSE_RACING_TERM	12.50	0.00	0.00	0.00	0.00	2.50
B-MANAGER	82.61	83.53	80.85	85.19	75.35	81.50
B-NATION	93.55	92.87	92.44	95.31	97.06	94.25
B-PERSON_ALIAS	41.21	70.64	74.77	79.61	77.63	68.77
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	83.13	81.13	77.17	77.01	73.91	78.47
B-ROLE	90.99	94.99	93.44	94.67	93.93	93.60
B-SCORE_TERM	65.43	85.24	82.58	88.56	86.74	81.71
B-SPORT_COMPETITION	84.50	88.29	87.70	87.85	89.58	87.58
B-SPORT_EVENT	88.83	93.77	93.28	94.68	94.39	92.99
B-SPORT_LEVEL	82.39	93.66	91.25	88.43	87.44	88.63
B-SPORT_LOCATION	82.67	89.74	89.88	91.94	89.80	88.81
B-SPORT_NAME	94.20	96.69	94.94	96.39	95.14	95.47
B-SPORT_ORGANIZATION	87.25	91.14	93.63	92.19	91.43	91.13
B-SPORT_PLAYER	89.83	93.22	92.74	91.53	87.22	90.91
B-SUPPORTER	84.53	93.35	93.01	90.79	92.42	90.82
B-TEAM	94.02	94.52	95.54	94.76	94.40	94.65
B-TEAM_ALIAS	86.84	93.27	87.53	94.78	94.36	91.35
B-TEAM_SPONSOR	71.70	93.43	20.00	82.76	72.73	68.12
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.31 Continued

I-AGREEMENT	73.33	70.59	80.00	67.61	65.52	71.41
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	86.83	91.11	77.95	91.00	89.53	87.28
I-EQUIPMENT	40.00	0.00	0.00	0.00	0.00	8.00
I-FOOTBALL_TERM	71.60	84.64	84.59	85.02	85.27	82.22
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	85.18	84.38	83.99	83.30	77.44	82.86
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	21.28	68.85	60.87	85.15	91.26	65.48
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	83.18	81.32	81.43	80.48	83.33	81.95
I-ROLE	83.63	87.65	86.24	88.78	88.37	86.93
I-SCORE_TERM	47.73	76.92	69.92	74.36	65.71	66.93
I-SPORT_COMPETITION	87.25	92.61	91.35	88.79	92.09	90.42
I-SPORT_EVENT	60.66	77.01	68.29	76.22	76.69	71.77
I-SPORT_LEVEL	86.12	92.64	90.51	90.73	89.72	89.94
I-SPORT_LOCATION	71.79	81.11	78.14	85.80	82.22	79.81
I-SPORT_NAME	0.00	28.57	0.00	0.00	0.00	5.71
I-SPORT_ORGANIZATION	77.91	87.12	86.22	83.29	76.87	82.28
I-SPORT_PLAYER	83.15	87.58	82.82	88.44	86.40	85.68
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	80.25	90.65	87.15	86.62	85.33	86.00
I-TEAM_ALIAS	87.08	95.32	89.97	96.79	95.36	92.90
I-TEAM_SPONSOR	42.86	50.00	0.00	0.00	21.05	22.78
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

4.5 Experiments with ConvBERT-Base Turkish Uncased Model

Table 4.32: Precision, Recall, F1-Score, Loss Of ConvBERT-Base Turkish Uncased Model On TurkishSportsNER-Parent-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	82.59	93.16	87.55	0.18
k=2	91.07	95.77	93.36	0.09
k=3	91.77	94.19	92.96	0.09
k=4	92.69	94.74	93.70	0.08
k=5	90.56	94.56	92.52	0.09
Average	89.73	94.48	92.01	0.10

Table 4.33: Precision, Recall, F1-Score, Loss Of ConvBERT-Base Turkish Uncased Model On TurkishSportsNER-Parent-BIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	80.69	91.68	85.84	0.21
k=2	89.18	94.65	91.83	0.11
k=3	90.18	92.49	91.32	0.11
k=4	90.96	93.01	91.97	0.10
k=5	89.16	92.18	90.64	0.11
Average	88.03	92.80	90.32	0.12

Table 4.34: Precision, Recall, F1-Score, Loss Of ConvBERT-Base Turkish Uncased Model On TurkishSportsNER-Child-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	82.21	90.73	86.26	0.20
k=2	89.62	93.89	91.71	0.11
k=3	89.64	91.69	90.65	0.12
k=4	91.43	92.76	92.09	0.10
k=5	88.96	91.99	90.45	0.11
Average	88.37	92.21	90.23	0.12

Table 4.35: Precision, Recall, F1-Score, Loss Of ConvBERT-Base Turkish Uncased Model On TurkishSportsNER-Child-BIO Test Data Set

Splits	Precision	Recall	F1-Score	Loss
k=1	80.20	89.75	84.71	0.23
k=2	88.17	93.31	90.67	0.13
k=3	87.75	90.50	89.10	0.14
k=4	89.90	91.92	90.90	0.12
k=5	87.60	90.73	89.14	0.14
Average	86.72	91.24	88.90	0.15

The ConvBERT-Base Turkish Uncased Model, as cited in [37][2], is a pre-trained model explicitly designed for the Turkish language. It was trained on the Turkish section of the MC4 corpus. Key experiment parameters included a maximum sequence length of 512, batch size 8, and 3 epochs. All other hyperparameters defaulted to the model's configuration according to [37].

Tables 4.32, 4.33, 4.34, and 4.35 display the evaluation results of the ConvBERT-Base Turkish Uncased Model for various TurkishSportsNER test data sets. The model's performance is measured using precision, recall, f1-score, and loss values. Evalua-

tions were conducted on different data splits represented by k=1 to k=5, as well as the model's average performance across these splits. The highest results for each data set variation are highlighted in bold within the tables.

Table 4.32 displays the model's performance on the TurkishSportsNER-Parent-Non-BIO test data set. The results show promising precision values, ranging from 82.59% to 92.69%, and recall values between 93.16% and 95.77%. As a result, the f1-scores fall between 87.55% and 93.70%, indicating the model's overall effectiveness in identifying named entities. The loss values are relatively low, varying from 0.10 to 0.18. The highest f1-score (93.70%) belongs to the fourth fold. The average f1-score for this data set is 92.01%.

The evaluation results for TurkishSportsNER-Parent-BIO are presented in Table 4.33. The results exhibit consistent performance, with precision values between 80.69% and 90.96%, recall ranging from 91.68% to 93.01%, and f1-scores from 85.84% to 91.97%. The loss values for this data set range from 0.10 to 0.21. The average f1-score of 90.32% demonstrates the model's proficiency in recognizing sports-related named entities while adhering to the BIO annotation format.

Table 4.34 shows the model's performance on the TurkishSportsNER-Child-NonBIO test data set. The model exhibits consistent and robust performance, with f1-scores in the range of 86.26% to 92.09%. The average f1-score of 90.23% showcases the model's effectiveness in identifying named entities within the child domain, even without the use of BIO labels.

Lastly, Table 4.35 highlights the model's performance on the TurkishSportsNER-Child-BIO test data set. The precision values range from 80.20% to 89.90%, recall varies from 89.75% to 91.92%, and f1-scores lie between 814.71% and 90.90%. The loss values for this data set range from 0.13 to 0.23.

Upon comparing the nth-fold training performance scores of the model, it becomes evident that the highest recognition performance occurs in the fourth fold of the data set, while the first fold exhibits significantly lower performance based on f1-scores across all tables.

When examining results according to annotation format (i.e., whether or not the data

set utilized BIO format), it's clear that the model performs better with non-BIO data sets than with BIO data sets. The model's average f1-scores consistently hover around 90%, indicating its strong performance and effectiveness in recognizing named entities across all test data sets.

Regarding nth-fold results, the ConvBERT-Base Turkish Uncased Model achieves its most successful NER outcomes on TurkishSportsNER-Parent-NonBIO and has its lowest recognition performance on TurkishSportsNER-Child-BIO. Contrary to the BERT-Base Multilingual Cased model and BERT-Base Turkish uncased model, ConvBERT-Base Turkish Uncased model is more successful on TurkishSportsNER-Parent-BIO than TurkishSportsNER-Child-NonBIO.

Table 4.36 represents average precision, recall and f1-scores of the ConvBERT-Base Turkish Uncased Model on TurkishSportsNER-Parent-NonBIO per named entity. It is observed that the model has a successful recognition performance, yet it notably under-performs in accurately identifying the TEAM_SPONSOR entity.

Average precision, recall and f1-scores of the model on TurkishSportsNER-Parent-BIO were given in Table 4.37. The model generally exhibits strong performance in recognizing various entities. However, it faces difficulties with the I-AGREEMENT entity. Additionally, the model fails to identify entities such as I-EQUIPMENT, I-HEALTH_TERM, I-NATION, I-SPORT_NAME, I-SUPPORTER, and I-TEAM_-SPONSOR.

In Table 4.38, average precision, recall, and f1-scores on TurkishSportsNER-Child-NonBIO are represented. The model demonstrates difficulties in accurately recognizing sports-related terms, with the exception of FOOTBALL_TERM, where its performance is notably better. Furthermore, it has a significantly low recognition capability for HORSE_RACING_TERM. The model also faces challenges in correctly identifying entities such as PERSON_ALIAS and TEAM_SPONSOR.

Average precision, recall, and f1-scores of the model on TurkishSportsNER-Child-BIO are represented in Table 4.39. The model exhibits failure in recognizing entities starting with "I-". Additionally, it demonstrates notably low recall scores for entities like B-PERSON_ALIAS, B-TEAM_SPONSOR, I-AGREEMENT, and I-SCORE_-

TERM. This pattern indicates the model's struggle to recognize entities that are underrepresented in the training data set.

Table 4.40 represents the f1-scores of the model on TurkishSportsNER-Parent-NonBIO per named entity. NATION, ROLE, SPORT_PERSON, and TEAM entities have higher results comparing the other named entities. The model performs better since it is pre-trained on a large Turkish data set. As a result, notable names of people can be recognized by the model more precisely. The lowest f1-scores belong to TEAM_SPONSOR which is mostly under 60. It shows the importance of the number of instances in train data sets.

The f1-scores of the model on TurkishSportsNER-Parent-BIO per named entity were given in Table 4.41. B-NATION, B-ROLE, and B-SPORT_NAME have the top 3 highest f1-scores. On the other hand, the model has unsuccessful results for B-TEAM_SPONSOR, I-AGREEMENT, I-EQUIPMENT, I-SUPPORTER, I-NATION, I-SPORT_NAME, I-HEALTH_TERM, I-TEAM_SPONSOR. Since the number of B-NATION and B-SPORT_NAME samples is higher than I-SPORT_NAME and I-NATION samples, the model can recognize them.

In Table 4.42, the f1-scores of TurkishSportsNER-Child-NonBIO were given. The table indicates that the model achieves exceptionally high f1-scores for certain named entity labels. Notably, the label *O* (representing non-entity tokens) consistently obtains f1-scores above 98%, showcasing the model's ability to identify non-entity tokens with high precision and recall correctly. Additionally, labels such as NATION, ROLE, and SPORT_NAME also demonstrate outstanding performance, with f1-scores ranging from 92% to 97%. However, the model faces challenges in correctly identifying some other entity labels. For instance, the labels BASKETBALL_TERM, HORSE_RACING_TERM, RACE_TERM, TENNIS_TERM, and VOLLEYBALL-_TERM receive f1-scores of 0% in all five splits, indicating the model's difficulty in recognizing these specific terms related to sports disciplines. Moreover, the labels MANAGER, PERSON_ALIAS, REFEREE, TEAM_SPONSOR, and HEALTH-_TERM instances exhibit relatively lower f1-scores, suggesting that the model struggles to identify entities within these categories.

The f1-scores of the model on TurkishSporstNER-Child-BIO per named entity are

represented in Table 4.43. Labels such as B-NATION, B-ROLE, B-SPORT_NAME exhibit excellent f1-scores ranging from 91% to 97%, indicating the model's effectiveness in recognizing named entities related to nations, roles, and sports names. On the other hand, the model faces challenges in accurately identifying specific named entity labels. For instance, labels like B-TEAM_SPONSOR, I-AGREEMENT, I-EQUIPMENT, I-SUPPORTER, I-NATION, I-PERSON_ALIAS, I-SPORT_NAME, I-HEALTH_TERM, I-TEAM_SPONSOR, and children entities of sports terms excluding football terms obtain f1-scores of 0% in all five splits. This suggests that the model struggles with correctly identifying entities within these categories, which may require further investigation and fine-tuning to improve performance.

In conclusion, the ConvBERT-Base Turkish Uncased Model exhibits promising results on the TurkishSportsNER data sets, particularly concerning parent-level non-BIO entities. The model demonstrates relatively high precision, recall, and f1-scores, indicating its efficacy in accurately identifying named entities. Nevertheless, its performance varies across different data sets and named entity types. The model's superiority on non-BIO data sets compared to BIO data sets suggests an advantage in dealing with the simplicity of the non-BIO annotation format. Moreover, the model displays better proficiency in recognizing parent-level entities than child-level entities, which can be attributed to the increased complexity involved in identifying the latter. Remarkably, the model achieves its highest recognition performance in the fourth fold of the data set. In contrast, the first fold exhibits lower performance, implying a correlation between data distribution and model performance variation. Nonetheless, further fine-tuning and optimization are required to enhance its recognition capabilities for a broader range of named entities, particularly those with limited instances in the data sets and child-level entities following the BIO schema. Investigating the model's limitations and exploring potential model updates hold the potential for improving performance and achieving more robust named entity recognition in the domain of sports.

Table 4.36: 5-Fold Average Precision, Recall and F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	99.08	97.58	98.32
AGREEMENT	87.28	93.90	90.23
ALIAS	90.68	90.74	90.69
EQUIPMENT	89.92	92.11	90.90
HEALTH_TERM	84.45	83.15	83.53
NATION	92.07	97.15	94.50
ROLE	93.46	96.41	94.91
SPORT_COMPETITION	89.80	94.33	92.00
SPORT_EVENT	90.80	95.73	93.14
SPORT_LEVEL	89.49	94.58	91.96
SPORT_LOCATION	86.22	94.64	90.08
SPORT_NAME	94.61	95.95	95.26
SPORT_ORGANIZATION	88.28	92.79	90.39
SPORT_PERSON	92.57	96.94	94.70
SPORT_TERM	81.97	91.54	86.06
SUPPORTER	88.83	92.09	90.39
TEAM	94.01	96.00	94.96
TEAM_SPONSOR	73.38	37.78	48.42

Table 4.37: 5-Fold Average Precision, Recall and F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	99.12	97.55	98.33
B-AGREEMENT	85.34	95.23	89.83
B-ALIAS	86.22	87.99	87.04

Table 4.37 continued from previous page

B-EQUIPMENT	87.76	94.32	90.84
B-HEALTH_TERM	83.61	86.07	84.72
B-NATION	92.14	97.67	94.80
B-ROLE	92.83	95.97	94.37
B-SPORT_COMPETITION	84.94	90.49	87.61
B-SPORT_EVENT	91.03	95.78	93.27
B-SPORT_LEVEL	87.36	90.30	88.76
B-SPORT_LOCATION	85.61	91.78	88.45
B-SPORT_NAME	94.76	96.27	95.50
B-SPORT_ORGANIZATION	86.08	94.05	89.76
B-SPORT_PERSON	91.76	95.39	93.53
B-SPORT_TERM	82.58	90.75	86.09
B-SUPPORTER	88.26	91.13	89.64
B-TEAM	93.81	96.15	94.92
B-TEAM_SPONSOR	40.00	2.70	4.97
I-AGREEMENT	36.00	8.49	13.45
I-ALIAS	91.70	95.04	93.27
I-EQUIPMENT	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00
I-ROLE	84.25	93.52	88.58
I-SPORT_COMPETITION	87.67	95.33	91.33
I-SPORT_EVENT	72.23	77.20	74.55
I-SPORT_LEVEL	86.24	94.68	90.24
I-SPORT_LOCATION	73.80	85.35	78.86
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	79.74	87.54	83.22
I-SPORT_PERSON	88.77	96.07	92.22
I-SPORT_TERM	75.19	85.04	79.55
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	87.28	90.81	88.99

Table 4.37 continued from previous page

I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.38: 5-Fold Average Precision, Recall and F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	99.03	97.74	98.38
AGREEMENT	87.21	94.09	90.25
BASKETBALL_TERM	0.00	0.00	0.00
COACH	87.88	91.03	89.37
EQUIPMENT	89.38	91.95	90.57
FOOTBALL_TERM	84.66	93.57	88.82
HEALTH_TERM	86.02	82.37	83.91
HORSE_RACING_TERM	0.91	0.53	0.67
MANAGER	78.05	88.59	82.70
NATION	91.09	97.79	94.27
PERSON_ALIAS	70.74	51.84	58.03
RACE_TERM	0.00	0.00	0.00
REFEREE	66.48	87.33	74.90
ROLE	92.83	96.54	94.65
SCORE_TERM	77.63	87.98	81.93
SPORT_COMPETITION	88.81	94.25	91.43
SPORT_EVENT	90.99	95.28	93.03
SPORT_LEVEL	88.69	93.84	91.18
SPORT_LOCATION	86.28	95.10	90.30
SPORT_NAME	94.17	96.03	95.06
SPORT_ORGANIZATION	86.12	93.35	89.48
SPORT_PLAYER	92.21	93.51	92.85
SUPPORTER	88.68	90.85	89.71
TEAM	94.26	95.87	95.03

Table 4.38 continued from previous page

TEAM_ALIAS	88.91	93.17	90.82
TEAM_SPONSOR	54.67	18.69	27.78
TENNIS_TERM	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00

Table 4.39: 5-Fold Average Precision, Recall and F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
B-AGREEMENT	84.79	95.29	89.54
B-BASKETBALL_TERM	0.00	0.00	0.00
B-COACH	87.27	90.35	88.74
B-EQUIPMENT	87.41	92.66	89.87
B-FOOTBALL_TERM	84.72	93.41	88.72
B-HEALTH_TERM	82.34	86.17	84.05
B-HORSE_RACING_TERM	0.00	0.00	0.00
B-MANAGER	72.50	86.87	78.89
B-NATION	90.96	97.58	94.12
B-PERSON_ALIAS	76.49	32.86	43.15
B-RACE_TERM	0.00	0.00	0.00
B-REFEREE	72.81	74.55	72.94
B-ROLE	92.60	95.89	94.21
B-SCORE_TERM	74.70	88.17	80.32
B-SPORT_COMPETITION	83.56	89.75	86.52
B-SPORT_EVENT	90.92	95.28	92.98
B-SPORT_LEVEL	86.24	90.25	88.13
B-SPORT_LOCATION	84.81	92.79	88.41
B-SPORT_NAME	94.39	96.33	95.34
B-SPORT_ORGANIZATION	88.82	94.16	91.39
B-SPORT_PLAYER	91.67	92.80	92.22

Table 4.39 continued from previous page

B-SUPPORTER	84.93	92.28	88.34
B-TEAM	93.68	96.19	94.88
B-TEAM_ALIAS	83.92	92.89	88.05
B-TEAM_SPONSOR	20.00	1.94	3.53
B-TENNIS_TERM	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00
I-AGREEMENT	20.00	1.33	2.50
I-BASKETBALL_TERM	0.00	0.00	0.00
I-COACH	86.52	91.30	88.74
I-EQUIPMENT	0.00	0.00	0.00
I-FOOTBALL_TERM	77.23	88.80	82.57
I-HEALTH_TERM	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00
I-MANAGER	74.66	90.68	81.57
I-NATION	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00
I-REFEREE	70.82	87.86	77.80
I-ROLE	84.07	93.08	88.25
I-SCORE_TERM	50.00	6.66	10.60
I-SPORT_COMPETITION	86.32	95.08	90.46
I-SPORT_EVENT	68.51	74.04	70.94
I-SPORT_LEVEL	86.72	94.21	90.30
I-SPORT_LOCATION	75.02	84.31	79.11
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	80.60	86.64	83.29
I-SPORT_PLAYER	87.11	90.64	88.77
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	86.15	90.95	88.48
I-TEAM_ALIAS	82.59	96.41	88.87
I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.39 continued from previous page

I-TENNIS_TERM	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00
O	99.07	97.69	98.38

Table 4.40: F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.30	98.52	98.55	98.69	98.55	98.32
AGREEMENT	82.94	92.52	91.37	91.71	92.63	90.23
ALIAS	87.51	92.60	85.83	94.74	92.79	90.69
EQUIPMENT	86.28	90.51	91.46	93.10	93.14	90.90
HEALTH_TERM	72.98	84.85	88.59	84.43	86.79	83.53
NATION	93.59	93.96	92.15	95.24	97.57	94.50
ROLE	92.29	95.53	95.26	95.98	95.50	94.91
SPORT_COMPETITION	87.75	94.32	93.38	91.96	92.59	92.00
SPORT_EVENT	87.70	94.40	93.90	94.91	94.79	93.14
SPORT_LEVEL	87.56	94.75	92.02	93.25	92.23	91.96
SPORT_LOCATION	83.68	89.67	92.34	92.43	92.28	90.08
SPORT_NAME	94.40	96.51	93.61	95.85	95.93	95.26
SPORT_ORGANIZATION	86.36	87.15	93.63	92.58	92.24	90.39
SPORT_PERSON	94.65	96.02	96.08	95.20	91.53	94.70
SPORT_TERM	70.17	87.72	89.19	92.22	91.00	86.06
SUPPORTER	84.54	93.66	91.81	89.69	92.24	90.39
TEAM	93.81	95.54	95.58	95.45	94.44	94.96
TEAM_SPONSOR	47.06	0.00	58.82	80.95	55.26	48.42

Table 4.41: F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.42	98.45	98.57	98.68	98.52	98.33
B-AGREEMENT	82.85	91.96	90.16	92.13	92.03	89.83
B-ALIAS	83.33	87.66	82.48	92.14	89.56	87.04
B-EQUIPMENT	84.18	92.38	91.66	93.10	92.90	90.84
B-HEALTH_TERM	77.58	83.85	88.96	86.42	86.79	84.72
B-NATION	94.09	93.78	94.00	95.59	96.56	94.80
B-ROLE	91.36	95.06	94.60	95.81	95.03	94.37
B-SPORT_COMPETITION	83.24	89.22	89.09	87.55	88.94	87.61
B-SPORT_EVENT	87.67	93.95	94.57	95.20	94.97	93.27
B-SPORT_LEVEL	84.58	93.16	89.02	89.81	87.26	88.76
B-SPORT_LOCATION	83.14	88.59	89.79	90.88	89.84	88.45
B-SPORT_NAME	94.56	96.04	95.28	96.30	95.32	95.50
B-SPORT_ORGANIZATION	85.48	86.66	94.69	91.02	90.97	89.76
B-SPORT_PERSON	93.35	95.58	95.11	93.84	89.78	93.53
B-SPORT_TERM	72.56	86.81	88.50	91.73	90.86	86.09
B-SUPPORTER	83.42	92.62	91.01	89.87	91.28	89.64
B-TEAM	93.54	94.91	95.58	95.50	95.09	94.92
B-TEAM_SPONSOR	6.25	0.00	0.00	0.00	18.60	4.97
I-AGREEMENT	40.00	0.00	27.27	0.00	0.00	13.45
I-ALIAS	90.68	94.30	89.53	96.73	95.14	93.27
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.81	91.30	87.87	88.62	90.27	88.58
I-SPORT_COMPETITION	86.90	93.51	92.73	91.40	92.09	91.33
I-SPORT_EVENT	55.32	77.75	77.78	80.45	81.48	74.55
I-SPORT_LEVEL	86.83	91.65	90.24	91.41	91.08	90.24
I-SPORT_LOCATION	69.79	75.68	78.55	85.36	84.93	78.86

Table 4.41 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	77.34	85.87	87.35	84.44	81.08	83.22
I-SPORT_PERSON	92.72	93.39	93.45	92.71	88.83	92.22
I-SPORT_TERM	59.50	84.57	82.95	85.03	85.71	79.55
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	85.99	92.90	89.77	89.31	86.96	88.99
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.42: F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.62	98.52	98.52	98.69	98.53	98.38
AGREEMENT	82.16	92.83	90.94	92.54	92.78	90.25
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	89.99	93.91	81.78	91.35	89.82	89.37
EQUIPMENT	86.44	90.20	91.23	93.24	91.76	90.57
FOOTBALL_TERM	80.88	87.32	91.24	92.70	91.97	88.82
HEALTH_TERM	73.35	83.02	88.59	89.57	85.05	83.91
HORSE_RACING_TERM	3.33	0.00	0.00	0.00	0.00	0.67
MANAGER	84.10	85.45	82.22	83.74	77.99	82.70
NATION	93.50	93.62	92.13	94.91	97.22	94.27
PERSON_ALIAS	33.96	61.16	67.80	62.95	64.31	58.03
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	70.68	74.89	80.13	78.76	70.04	74.90
ROLE	92.28	95.28	94.98	95.90	94.81	94.65
SCORE_TERM	64.98	84.75	84.54	87.61	87.80	81.93
SPORT_COMPETITION	86.60	93.88	93.31	91.46	91.91	91.43
SPORT_EVENT	87.68	93.99	93.80	94.86	94.80	93.03
SPORT_LEVEL	86.10	94.59	91.61	92.13	91.44	91.18

Table 4.42 Continued

SPORT_LOCATION	83.25	89.53	92.01	92.69	94.02	90.30
SPORT_NAME	93.09	96.78	94.43	95.67	95.34	95.06
SPORT_ORGANIZATION	85.75	85.88	93.27	92.39	90.12	89.48
SPORT_PLAYER	91.82	95.30	93.26	94.06	89.79	92.85
SUPPORTER	85.15	92.00	91.73	89.76	89.91	89.71
TEAM	93.91	95.66	95.27	95.43	94.90	95.03
TEAM_ALIAS	86.93	92.48	88.80	92.00	93.90	90.82
TEAM_SPONSOR	33.33	0.00	0.00	56.25	49.32	27.78
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.43: F1-Scores of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.65	98.50	98.52	98.70	98.51	98.38
B-AGREEMENT	82.34	92.31	90.61	91.21	91.21	89.54
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	87.99	93.91	82.26	91.58	87.97	88.74
B-EQUIPMENT	84.01	91.05	91.25	93.01	90.04	89.87
B-FOOTBALL_TERM	80.90	87.27	90.96	92.85	91.64	88.72
B-HEALTH_TERM	77.71	84.50	88.67	85.14	84.26	84.05
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	82.46	80.10	76.43	80.74	74.71	78.89
B-NATION	93.26	93.82	91.98	95.67	95.86	94.12
B-PERSON_ALIAS	35.14	42.86	52.57	38.30	46.88	43.15
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	74.09	74.58	70.53	73.91	71.60	72.94
B-ROLE	91.24	95.17	94.39	95.56	94.69	94.21
B-SCORE_TERM	63.09	83.42	83.52	86.16	85.41	80.32

Table 4.43 Continued

B-SPORT_COMPETITION	81.17	87.65	88.56	87.20	88.04	86.52
B-SPORT_EVENT	87.58	93.49	94.06	94.73	95.02	92.98
B-SPORT_LEVEL	81.23	92.79	90.96	88.17	87.52	88.13
B-SPORT_LOCATION	81.53	87.44	90.17	92.61	90.30	88.41
B-SPORT_NAME	93.75	96.75	94.68	96.64	94.87	95.34
B-SPORT_ORGANIZATION	86.21	90.98	94.47	92.37	92.91	91.39
B-SPORT_PLAYER	91.83	95.24	93.18	93.06	87.78	92.22
B-SUPPORTER	83.62	91.87	90.11	88.85	87.23	88.34
B-TEAM	93.48	94.97	95.48	95.52	94.98	94.88
B-TEAM_ALIAS	83.96	89.92	86.83	89.50	90.03	88.05
B-TEAM_SPONSOR	17.65	0.00	0.00	0.00	0.00	3.53
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	12.50	0.00	0.00	0.00	0.00	2.50
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	87.86	92.09	81.46	90.59	91.70	88.74
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	71.28	85.39	84.05	85.99	86.12	82.57
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	86.30	82.06	78.04	83.14	78.31	81.57
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	81.31	75.47	79.71	79.07	73.45	77.80
I-ROLE	84.57	91.46	86.68	89.95	88.58	88.25
I-SCORE_TERM	29.17	0.00	0.00	14.29	9.52	10.60
I-SPORT_COMPETITION	83.92	93.68	92.02	90.47	92.21	90.46
I-SPORT_EVENT	52.55	74.66	71.71	75.90	79.87	70.94
I-SPORT_LEVEL	83.73	93.55	91.57	91.23	91.40	90.30

Table 4.43 Continued

I-SPORT_LOCATION	67.35	74.53	78.83	88.18	86.67	79.11
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	77.32	87.19	85.98	84.85	81.10	83.29
I-SPORT_PLAYER	88.23	91.60	85.03	90.35	88.64	88.77
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	83.49	93.66	89.16	89.08	87.03	88.48
I-TEAM_ALIAS	89.80	90.26	85.71	88.20	90.36	88.87
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

4.6 Experiments with DistilBERT-Base Turkish Cased Model

Table 4.44: Precision, Recall, F1-Score, Loss Of DistilBERT-Base Turkish Cased Model On TurkishSportsNER-Parent-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	74.77	83.21	78.77	0.25
k=2	80.01	84.91	82.38	0.20
k=3	80.68	83.95	82.28	0.20
k=4	81.27	84.24	82.73	0.19
k=5	80.24	84.73	82.42	0.19
Average	79.39	84.20	81.71	0.20

Table 4.45: Precision, Recall, F1-Score, Loss Of DistilBERT-Base Turkish Cased Model On TurkishSportsNER-Parent-BIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	66.74	79.45	72.54	0.30
k=2	71.95	79.49	75.53	0.26
k=3	73.15	78.90	75.92	0.25
k=4	72.80	79.35	75.93	0.25
k=5	72.56	79.01	75.64	0.24
Average	71.44	79.24	75.11	0.26

Table 4.46: Precision, Recall, F1-Score, Loss Of DistilBERT-Base Turkish Cased Model On TurkishSportsNER-Child-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	66.57	76.37	71.13	0.31
k=2	72.39	78.32	75.24	0.27
k=3	73.57	78.42	75.92	0.27
k=4	74.46	79.91	77.09	0.25
k=5	72.93	78.99	75.84	0.24
Average	71.98	78.40	75.04	0.26

Table 4.47: Precision, Recall, F1-Score, Loss Of DistilBERT-Base Turkish Cased Model On TurkishSportsNER-Child-BIO Test Data Set

Splits	Precision	Recall	F1-Score	Loss
k=1	63.34	74.51	68.47	0.36
k=2	67.84	75.27	71.36	0.32
k=3	69.98	74.97	72.39	0.32
k=4	69.66	76.46	72.90	0.29
k=5	69.10	75.34	72.08	0.29
Average	67.98	75.31	71.44	0.31

DistilBERT-Base Turkish Cased Model [40][39] is a pretrained model. The model is a distilled version of the BERT-Base model [50]. It was trained on only an English data set. For this experiment, we set the maximum sequence length to 512, batch size to 8, and the number of epochs to 3. The remaining hyperparameters were kept at their default values, following the model's configuration [40].

The evaluation results of the BERT-Base Multilingual Cased Model on various test data sets for different versions of TurkishSportsNER are presented in Table 4.44, Table 4.45, Table 4.46, and Table 4.47. The model's performance is evaluated using

precision, recall, f1-score, and loss values. Assessments are conducted for varying data splits, represented by k=1 to k=5, as well as the average performance across these splits. The highest results for each data set variation are highlighted in bold throughout all tables.

Table 4.44 showcases the precision, recall, f1-score, and loss values of the DistilBERT-Base Turkish Cased Model on the TurkishSportsNER-Parent-NonBIO test data set. f1-scores range from 78.77% to 82.73% for different splits (k=1 to k=5). The average f1-score of 81.71% indicates its capability to identify various entities accurately. Additionally, the loss value of around 0.20 indicates the model's ability to minimize errors during training and prediction.

Table 4.45 presents the performance of the model on the TurkishSportsNER-Parent-BIO test data set. It achieves f1-scores ranging from 72.54% to 75.93% for different splits, with an average f1-score of 75.11%. The f1-scores are lower compared to the TurkishSportsNER-Parent-NonBIO data set.

The results of the model on TurkishSportsNER-Child-NonBIO were given in Table 4.46. The f1-scores vary from 71.13% to 77.09% for different splits, with an average f1-score of 75.04%. Also, precision values range from 66.57% to 74.46%, recall lies between 76.37% and 79.91%. The corresponding loss values are between 0.24 and 0.31.

The results of the model's performance on the TurkishSportsNER-Child-BIO are represented in Table 4.47. It achieves precision ranging from 63.44% to 69.10%, recall lies between 74.51% and 76.46%, and f1-scores are between 68.47% and 72.90% with an average f1-score of 71.44%.

Upon comparing the nth-fold training performance scores of the model, it becomes evident that the highest recognition performance occurs in the fourth fold of the data set, while the first fold demonstrates a significantly lower performance based on the f1-scores across all tables.

An examination of the results by annotation format reveals that the model performs better with non-BIO data sets compared to BIO data sets.

Considering the nth-fold outcomes, the DistilBERT-Base Turkish Cased Model achieves its most successful NER results on TurkishSportsNER-Parent-NonBIO and exhibits the lowest recognition performance on TurkishSportsNER-Child-BIO.

In the below tables, average precision, recall, and f1-scores per named entity for different versions of the data set are represented. Mainly, it is observed that the highest precision, recall, and f1-scores belong to the O label for each split of the data sets.

Table 4.48 represents average precision, recall, and f1-scores on TurkishSportsNER-Parent-NonBIO data set. The model performs relatively well for entities like AGREEMENT, ROLE, SPORT_EVENT, and SPORT_NAME, with balanced precision, recall, and f1-scores. However, certain entities like HEALTH_TERM, NATION, and particularly TEAM_SPONSOR show lower scores, indicating the model struggles in these areas. TEAM_SPONSOR, for example, has a very low recall, suggesting the model misses most of the actual instances of this entity.

Table 4.49 represents average precision, recall, and f1-scores on TurkishSportsNER-Parent-BIO data set. The model performs well in recognizing entities prefixed with "B-" with reasonably high scores across precision, recall, and f1-score. The model shows lower performance for entities like B-HEALTH_TERM, B-SPORT_COMPETITION, and B-SPORT_ORGANIZATION. A notable observation is that the model completely fails to recognize certain "I-" prefixed entities like I-AGREEMENT, I-EQUIPMENT, I-HEALTH_TERM, etc. This indicates a significant difficulty in identifying the continuation of entities within the text.

In Table 4.50, average precision, recall, and f1-scores of the model on TurkishSportsNER-Child-NonBIO data set per named entity. For entities like AGREEMENT, FOOTBALL_TERM, ROLE, and SPORT_EVENT, the model demonstrates a good performance with balanced scores across all three metrics. Notably, the model completely fails to recognize several specific sports terms. This indicates a significant challenge in identifying these specific sports categories. The model also shows poor performance for PERSON_ALIAS, with very low scores. For some entities like COACH, MANAGER, and SPORT_PLAYER, the model has moderate performance.

Average precision, recall, and f1-scores of the model on TurkishSportsNER-Child-

BIO data set per named entity were given in Table 4.51. The model has moderate performance in identifying entities with the "B-" prefix, while it demonstrates a notable shortfall in recognizing sports-related terms, with the exceptions of B-FOOTBALL_TERM and I-FOOTBALL_TERM. Furthermore, the model completely fails to accurately identify entities like B-PERSON_ALIAS and B-TEAM_SPONSOR. The model also struggles with 14 "I-" prefixed entities, showing moderate effectiveness in recognizing other entities.

Table 4.52 represents the f1-scores of the model on TurkishSportsNER-Parent-Non-BIO per named entity type. Comparing the f1-scores of the named entities, the model performs better for ROLE and SPORT_EVENT named entities. The data set has some English samples for these entities. It indicates that the existence of English samples has been influential in the success of the model. On the other hand, although the data set contains many foreign athlete names, the model cannot perform on SPORT_PERSON named entity.

In Table 4.53, f1-scores of the model on TurkishSportsNER-Parent-BIO per named entity. Looking at the results, we observe that the *O* label, which represents non-entity tokens, consistently achieves high f1-scores across all data splits, indicating that the model can effectively recognize non-entity tokens. For other entity types, such as B-ROLE, B-SPORT_EVENT, B-SPORT_NAME, and I-ALIAS, the model also demonstrates relatively high f1-scores, showing its proficiency in identifying these entities. On the other hand, the model faces challenges in recognizing certain entities, as evident from the low f1-scores for I-SPORT_EVENT, I-SPORT_LOCATION, and I-TEAM. These entities seem to pose difficulties for the model, resulting in lower f1-scores.

Table 4.54 presents the s1-scores of the DistilBERT-Base Turkish Cased Model, which was trained on the TurkishSportsNER-Child-NonBIO data set. The table exhibits the evaluation results for various labels corresponding to different data splits represented by k=1 to k=5. Upon analyzing the table, for specific named entities, we observe varying levels of performance. For instance, the COACH and REFEREE labels display lower f1-scores in comparison to other entities, suggesting that the model faces challenges in accurately identifying these specific people within the sports domain. On

the other hand, labels such as ROLE, SPORT_EVENT, and SPORT_NAME demonstrate consistently high f1-scores throughout all data splits, highlighting the model's effectiveness in recognizing these entities with f1-scores.

The f1-scores of the model on TurkishSportsNER-Child-BIO per named entity were given in Table 4.55. Specific entity labels exhibit varying levels of performance. For example, the B-REFEREE, I-SPORT_LOCATION, and I-SPORT_EVENT labels display relatively low f1-scores across all data splits, implying that the model struggles to identify and label these specific entities accurately. On the other hand, labels such as B-ROLE, B-SPORT_EVENT, and B-SPORT_NAME demonstrate consistently high f1-scores throughout all data splits. Also, the model has failed on child entities of sports terms, B-PERSON_ALIAS, B-TEAM_SPONSOR, I-AGREEMENT, I-EQUIPMENT, I-HEALTH_TERM, I-NATION, I-PERSON_ALIAS, I-REFEREE, I-SPORT_NAME, I-SUPPORTER, I-TEAM_SPONSOR.

In summary, the DistilBERT-Base Turkish Cased model's performance is comparatively lower as it was trained solely on English data. Moreover, the complexity of the task, with an increasing number of named entities, has impacted its recognition performance. The model demonstrates promising performance on the TurkishSportsNER data sets. It achieves relatively high precision, recall, and f1-scores, showcasing its ability to identify named entities accurately. However, its performance varies across different data sets and named entity types. The model exhibits better recognition of non-BIO data sets compared to BIO data sets, achieving higher f1-scores in the former. Additionally, it shows more excellent proficiency in recognizing parent-level entities than child-level entities. The model's performance also varies across different data splits, with the fourth fold performing the best, while the first fold lags behind. Overall, the DistilBERT-Base Turkish Cased Model shows promise for named entity recognition in TurkishSportsNER. It excels in identifying specific entity types while facing challenges with others. Further fine-tuning and optimization could enhance its recognition capabilities for a broader range of named entities.

Table 4.48: 5-Fold Average Precision, Recall and F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	97.59	96.27	96.92
AGREEMENT	87.01	88.11	87.45
ALIAS	87.49	74.12	80.15
EQUIPMENT	87.51	80.47	83.78
HEALTH_TERM	92.04	61.61	73.58
NATION	85.06	68.09	75.53
ROLE	88.30	92.67	90.44
SPORT_COMPETITION	72.39	76.96	74.57
SPORT_EVENT	88.38	90.83	89.54
SPORT_LEVEL	81.92	77.29	79.49
SPORT_LOCATION	83.16	69.67	75.52
SPORT_NAME	87.91	89.35	88.56
SPORT_ORGANIZATION	72.26	75.82	73.93
SPORT_PERSON	75.55	88.94	81.69
SPORT_TERM	79.08	80.83	79.79
SUPPORTER	94.65	81.06	87.29
TEAM	81.23	83.77	82.43
TEAM_SPONSOR	55.00	7.06	12.42

Table 4.49: 5-Fold Average Precision, Recall and F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	97.49	96.28	96.88
B-AGREEMENT	81.83	90.25	85.73
B-ALIAS	84.44	61.81	71.15

Table 4.49 continued from previous page

B-EQUIPMENT	83.43	84.19	83.72
B-HEALTH_TERM	90.88	44.06	59.07
B-NATION	77.64	66.86	71.67
B-ROLE	87.19	91.12	89.11
B-SPORT_COMPETITION	66.93	57.73	61.93
B-SPORT_EVENT	85.49	92.30	88.70
B-SPORT_LEVEL	77.39	67.16	71.82
B-SPORT_LOCATION	78.83	69.85	73.83
B-SPORT_NAME	86.39	90.79	88.49
B-SPORT_ORGANIZATION	64.59	76.87	69.93
B-SPORT_PERSON	69.37	83.01	75.55
B-SPORT_TERM	76.26	82.28	78.94
B-SUPPORTER	93.73	79.48	85.99
B-TEAM	80.55	84.03	82.20
B-TEAM_SPONSOR	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00
I-ALIAS	96.30	83.31	89.32
I-EQUIPMENT	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00
I-ROLE	73.68	76.83	75.17
I-SPORT_COMPETITION	65.99	78.31	71.46
I-SPORT_EVENT	61.63	15.07	23.98
I-SPORT_LEVEL	76.74	77.78	77.21
I-SPORT_LOCATION	75.00	6.08	10.97
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	72.64	41.59	52.30
I-SPORT_PERSON	62.37	71.04	66.39
I-SPORT_TERM	65.40	56.47	60.41
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	59.07	46.23	51.76

Table 4.49 continued from previous page

I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.50: 5-Fold Average Precision, Recall and F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
O	97.56	96.14	96.84
AGREEMENT	87.59	88.12	87.75
BASKETBALL_TERM	0.00	0.00	0.00
COACH	62.03	54.51	57.49
EQUIPMENT	86.50	80.21	83.16
FOOTBALL_TERM	78.88	82.20	80.43
HEALTH_TERM	92.16	57.24	70.39
HORSE_RACING_TERM	0.00	0.00	0.00
MANAGER	55.08	71.38	61.99
NATION	78.46	66.99	72.07
PERSON_ALIAS	4.98	0.89	1.46
RACE_TERM	0.00	0.00	0.00
REFEREE	80.00	2.84	5.32
ROLE	86.65	93.21	89.81
SCORE_TERM	74.51	75.19	73.93
SPORT_COMPETITION	66.03	78.84	71.85
SPORT_EVENT	86.53	90.82	88.55
SPORT_LEVEL	78.67	78.36	78.49
SPORT_LOCATION	82.73	67.82	74.18
SPORT_NAME	89.19	88.81	88.94
SPORT_ORGANIZATION	70.00	72.80	71.15
SPORT_PLAYER	62.17	78.79	69.37
SUPPORTER	93.50	79.96	86.18
TEAM	78.07	84.32	81.03

Table 4.50 continued from previous page

TEAM_ALIAS	84.36	79.05	81.44
TEAM_SPONSOR	0.00	0.00	0.00
TENNIS_TERM	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00

Table 4.51: 5-Fold Average Precision, Recall and F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-BIO Data Set per Named Entity

Label	Avg. Precision	Avg. Recall	Avg. F1-Score
B-AGREEMENT	85.06	89.08	86.91
B-BASKETBALL_TERM	0.00	0.00	0.00
B-COACH	63.42	45.46	52.75
B-EQUIPMENT	84.83	81.89	83.18
B-FOOTBALL_TERM	78.15	84.25	80.94
B-HEALTH_TERM	91.42	58.12	70.97
B-HORSE_RACING_TERM	0.00	0.00	0.00
B-MANAGER	50.58	56.02	53.02
B-NATION	78.46	67.15	72.29
B-PERSON_ALIAS	0.00	0.00	0.00
B-RACE_TERM	0.00	0.00	0.00
B-REFEREE	100.00	2.23	4.32
B-ROLE	85.75	91.88	88.70
B-SCORE_TERM	73.14	74.71	73.13
B-SPORT_COMPETITION	62.51	57.88	60.10
B-SPORT_EVENT	84.32	92.73	88.25
B-SPORT_LEVEL	75.21	66.71	70.56
B-SPORT_LOCATION	78.11	69.12	73.22
B-SPORT_NAME	86.77	90.12	88.35
B-SPORT_ORGANIZATION	66.36	74.08	69.97
B-SPORT_PLAYER	58.37	76.19	66.02

Table 4.51 continued from previous page

B-SUPPORTER	93.82	79.65	86.14
B-TEAM	76.70	85.46	80.74
B-TEAM_ALIAS	82.99	75.82	79.10
B-TEAM_SPONSOR	0.00	0.00	0.00
B-TENNIS_TERM	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00
I-BASKETBALL_TERM	0.00	0.00	0.00
I-COACH	70.50	47.15	56.10
I-EQUIPMENT	0.00	0.00	0.00
I-FOOTBALL_TERM	64.54	62.88	63.53
I-HEALTH_TERM	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00
I-MANAGER	44.17	64.84	52.01
I-NATION	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00
I-REFEREE	0.00	0.00	0.00
I-ROLE	74.15	76.08	74.98
I-SCORE_TERM	0.00	0.00	0.00
I-SPORT_COMPETITION	66.97	78.01	71.90
I-SPORT_EVENT	63.62	13.39	21.99
I-SPORT_LEVEL	74.13	76.50	75.26
I-SPORT_LOCATION	60.00	0.57	1.14
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	69.34	45.09	53.47
I-SPORT_PLAYER	43.77	50.72	46.92
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	65.23	41.25	50.49
I-TEAM_ALIAS	88.31	83.26	85.57
I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.51 continued from previous page

I-TENNIS_TERM	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00
O	97.18	96.59	96.88

Table 4.52: F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.20	96.83	97.01	97.21	97.35	96.92
AGREEMENT	82.73	90.51	85.81	89.18	89.01	87.45
ALIAS	79.85	82.88	73.67	84.30	80.06	80.15
EQUIPMENT	77.71	82.35	83.40	87.93	87.52	83.78
HEALTH_TERM	66.67	80.50	72.58	70.72	77.42	73.58
NATION	80.87	73.11	75.49	76.51	71.65	75.53
ROLE	87.81	90.89	90.77	91.19	91.52	90.44
SPORT_COMPETITION	71.14	76.01	76.66	72.84	76.21	74.57
SPORT_EVENT	85.75	90.47	90.44	90.24	90.81	89.54
SPORT_LEVEL	76.98	77.57	81.72	82.85	78.31	79.49
SPORT_LOCATION	75.09	77.29	73.68	74.03	77.49	75.52
SPORT_NAME	85.42	89.35	86.96	89.74	91.34	88.56
SPORT_ORGANIZATION	69.28	75.09	74.07	74.06	77.15	73.93
SPORT_PERSON	81.12	82.55	82.08	81.64	81.07	81.69
SPORT_TERM	71.43	79.34	81.65	84.31	82.22	79.79
SUPPORTER	83.98	90.11	89.33	85.03	88.03	87.29
TEAM	80.14	82.69	83.69	82.64	82.97	82.43
TEAM_SPONSOR	10.91	0.00	0.00	23.08	28.13	12.42

Table 4.53: F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.24	96.69	97.00	97.17	97.30	96.88
B-AGREEMENT	80.48	88.36	84.97	87.33	87.49	85.73
B-ALIAS	72.55	73.38	61.23	76.58	72.01	71.15
B-EQUIPMENT	78.33	83.54	82.55	88.03	86.13	83.72
B-HEALTH_TERM	56.91	52.41	56.34	61.08	68.64	59.07
B-NATION	78.18	69.87	70.65	72.34	67.34	71.67
B-ROLE	86.61	90.15	88.35	90.28	90.16	89.11
B-SPORT_COMPETITION	57.45	61.21	63.25	61.79	65.96	61.93
B-SPORT_EVENT	84.70	88.90	90.38	89.44	90.07	88.70
B-SPORT_LEVEL	70.67	66.99	73.25	78.68	69.50	71.82
B-SPORT_LOCATION	74.17	73.68	72.13	74.61	74.58	73.83
B-SPORT_NAME	85.80	89.52	86.50	89.63	91.01	88.49
B-SPORT_ORGANIZATION	68.06	75.32	68.87	71.98	65.40	69.93
B-SPORT_PERSON	75.60	77.36	77.57	74.10	73.13	75.55
B-SPORT_TERM	69.99	78.22	80.69	82.87	82.90	78.94
B-SUPPORTER	83.12	88.23	87.40	84.34	86.89	85.99
B-TEAM	80.83	81.47	83.53	82.25	82.91	82.20
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-ALIAS	88.29	91.35	87.35	91.54	88.07	89.32
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	73.99	77.05	70.88	73.08	80.85	75.17
I-SPORT_COMPETITION	66.12	71.99	73.68	71.85	73.65	71.46
I-SPORT_EVENT	25.64	23.74	22.11	17.73	30.69	23.98
I-SPORT_LEVEL	74.13	75.48	78.77	79.67	78.00	77.21
I-SPORT_LOCATION	16.16	11.01	12.00	12.12	3.57	10.97

Table 4.53 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	49.09	52.40	53.38	52.67	53.97	52.30
I-SPORT_PERSON	69.32	66.77	65.23	66.27	64.38	66.39
I-SPORT_TERM	54.63	59.19	62.78	65.42	60.03	60.41
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	46.74	65.94	46.64	51.46	48.05	51.76
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.54: F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.16	96.68	96.93	97.16	97.29	96.84
AGREEMENT	83.75	90.56	86.59	88.52	89.33	87.75
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	55.87	64.26	54.86	59.84	52.63	57.49
EQUIPMENT	78.39	81.79	81.65	88.44	85.55	83.16
FOOTBALL_TERM	74.34	78.32	83.14	84.79	81.55	80.43
HEALTH_TERM	63.27	78.81	67.23	70.00	72.63	70.39
HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
MANAGER	68.10	60.48	54.60	62.76	64.01	61.99
NATION	79.34	68.63	70.94	72.43	69.01	72.07
PERSON_ALIAS	6.32	0.00	0.00	0.00	0.98	1.46
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	0.58	17.31	3.84	0.00	4.85	5.32
ROLE	87.96	90.54	89.46	90.72	90.36	89.81
SCORE_TERM	61.61	75.64	73.34	77.15	81.90	73.93
SPORT_COMPETITION	66.07	72.43	74.42	71.55	74.81	71.85
SPORT_EVENT	83.87	89.08	89.41	90.09	90.30	88.55
SPORT_LEVEL	75.02	75.50	80.29	82.17	79.46	78.49

Table 4.54 Continued

SPORT_LOCATION	72.96	78.59	70.52	73.75	75.08	74.18
SPORT_NAME	85.65	90.20	86.66	90.14	92.03	88.94
SPORT_ORGANIZATION	67.65	74.38	71.05	72.73	69.94	71.15
SPORT_PLAYER	65.82	69.52	72.61	70.51	68.38	69.37
SUPPORTER	83.62	88.68	88.10	84.25	86.26	86.18
TEAM	78.79	80.84	82.51	81.21	81.80	81.03
TEAM_ALIAS	81.85	83.32	80.12	81.47	80.45	81.44
TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.55: F1-Scores of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.29	96.75	96.87	97.14	97.34	96.88
B-AGREEMENT	81.72	89.69	85.52	87.99	89.60	86.91
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	53.46	61.95	50.77	54.19	43.37	52.75
B-EQUIPMENT	78.46	83.68	80.81	88.13	84.82	83.18
B-FOOTBALL_TERM	75.45	79.27	82.38	84.73	82.85	80.94
B-HEALTH_TERM	68.66	75.68	67.78	68.89	73.86	70.97
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	62.25	54.19	41.01	53.30	54.35	53.02
B-NATION	77.61	69.82	74.27	72.81	66.95	72.29
B-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	1.03	7.41	0.79	5.46	6.90	4.32
B-ROLE	86.07	90.11	87.99	89.52	89.81	88.70
B-SCORE_TERM	59.36	76.27	71.29	77.67	81.07	73.13

Table 4.55 Continued

B-SPORT_COMPETITION	56.38	59.08	62.15	60.53	62.35	60.10
B-SPORT_EVENT	83.94	88.62	90.06	89.19	89.46	88.25
B-SPORT_LEVEL	67.61	66.56	72.56	77.49	68.57	70.56
B-SPORT_LOCATION	74.32	73.62	71.26	73.27	73.62	73.22
B-SPORT_NAME	85.36	89.43	87.10	89.30	90.55	88.35
B-SPORT_ORGANIZATION	66.80	72.99	67.08	70.70	72.29	69.97
B-SPORT_PLAYER	64.18	66.96	69.62	65.35	63.98	66.02
B-SUPPORTER	83.55	88.38	86.69	84.68	87.40	86.14
B-TEAM	80.38	77.60	82.10	81.53	82.07	80.74
B-TEAM_ALIAS	80.24	80.63	77.87	80.22	76.54	79.10
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	52.95	63.26	51.42	63.08	49.79	56.10
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	56.92	61.25	66.83	67.89	64.74	63.53
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	61.06	51.88	40.00	53.45	53.68	52.01
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	74.12	77.35	69.23	74.26	79.96	74.98
I-SCORE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_COMPETITION	66.81	72.76	73.96	71.63	74.35	71.90
I-SPORT_EVENT	23.93	23.36	22.46	21.26	18.93	21.99
I-SPORT_LEVEL	72.23	72.11	78.71	78.21	75.05	75.26

Table 4.55 Continued

I-SPORT_LOCATION	2.33	2.08	0.00	1.28	0.00	1.14
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	49.22	52.36	56.69	55.68	53.40	53.47
I-SPORT_PLAYER	44.11	45.84	47.56	48.97	48.09	46.92
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	49.69	65.90	45.71	47.02	44.13	50.49
I-TEAM_ALIAS	88.31	86.78	85.53	85.46	81.78	85.57
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

4.7 Experiments with XLM-RoBERTa-Base Multilingual Model

XLM-RoBERTa-Base Multilingual Model [44][41] is a pre-trained model. This model was trained on CommonCrawl data containing 100 languages including Turkish. In this particular experiment, the maximum sequence length was configured to be 512, the batch size was chosen as 8, and the number of epochs was set to 3. As for the rest of the hyper-parameters, they were kept at their default values following the model's configuration [44].

The evaluation results of the XLM-RoBERTa-Base Multilingual model on various test data sets for different variations of TurkishSportsNER are shown in Table 4.56, Table 4.57, Table 4.58, and Table 4.59. The performance of the model is evaluated based on metrics such as precision, recall, f1-score, and loss, and the evaluation is performed on different data splits indicated by k=1 to k=5, along with the average performance across these splits. The highest results for each variation of the data set are indicated in bold in all tables.

Table 4.56, the performance of the model on the TurkishSportsNER-Parent-NonBIO test data set. The model achieves reasonably high precision, recall, and f1-score values, ranging from 87.48% to 92.86%. These results suggest that the model is pro-

Table 4.56: Precision, Recall, F1-Score, Loss Of XLM-RoBERTa-Base Multilingual Model On TurkishSportsNER-Parent-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	83.10	92.36	87.48	0.19
k=2	90.87	94.68	92.73	0.09
k=3	91.17	93.01	92.08	0.10
k=4	91.68	94.08	92.86	0.09
k=5	90.52	93.35	91.91	0.09
Average	89.46	93.49	91.41	0.11

Table 4.57: Precision, Recall, F1-Score, Loss Of XLM-RoBERTa-Base Multilingual Model On TurkishSportsNER-Parent-BIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	80.26	91.00	85.07	0.21
k=2	88.16	92.80	90.42	0.11
k=3	88.66	90.99	89.81	0.11
k=4	89.91	91.68	90.78	0.10
k=5	88.84	90.41	89.62	0.11
Average	87.16	91.27	89.14	0.12

Table 4.58: Precision, Recall, F1-Score, Loss Of XLM-RoBERTa-Base Multilingual Model On TurkishSportsNER-Child-NonBIO Test Data Set

Splits	Precision	Recall	F1-Score	loss
k=1	80.91	90.10	85.26	0.21
k=2	88.93	92.70	90.78	0.11
k=3	88.68	90.15	89.41	0.13
k=4	90.01	92.35	91.16	0.10
k=5	88.34	90.97	89.63	0.11
Average	87.37	91.25	89.24	0.13

Table 4.59: Precision, Recall, F1-Score, Loss Of XLM-RoBERTa-Base Multilingual Model On TurkishSportsNER-Child-BIO Test Data Set

Splits	Precision	Recall	F1-Score	Loss
k=1	78.79	88.62	83.42	0.24
k=2	86.28	91.22	88.69	0.13
k=3	86.60	88.86	87.71	0.15
k=4	88.34	90.46	89.39	0.12
k=5	87.29	89.16	88.21	0.13
Average	85.46	89.66	87.48	0.15

ficient in identifying parent-level entities within the non-BIO format. Additionally, the low loss values indicate that the model's predictions closely align with the ground truth labels.

Table 4.57 displays the model's evaluation on the TurkishSportsNER-Parent-BIO test data set. Similar to the previous table, the XLM-RoBERTa model demonstrates good precision, recall, and f1-score values, ranging from 85.07% to 90.78% with an average f1-score of 89.14%.

Moving on to Table 4.58, it presents the model's performance on the TurkishSportsNER-Child-NonBIO test data set. Precision values range from 80.91% to 90.01%, recall

values lie between 90.10% and 92.35%, and the average f1-score is 89.24%. The corresponding loss values range from 0.10 to 0.21.

Finally, Table 4.59 provides the evaluation metrics for the XLM-RoBERTa model on the TurkishSportsNER-Child-BIO test data set. The model achieves reasonably low precision, recall, and f1-score values with precision ranging from 78.79% to 88.34%, recall in the range of 88.62%-90.46%, and an average f1-score of 87.48%.

When comparing the model's training performance scores across different folds, it can be observed that the model achieves its best recognition performance on the fourth fold of the data set. However, the model's performance is noticeably lower on the first fold of the data set, as indicated by the f1-scores in all tables.

When comparing the results based on the annotation format (BIO or non-BIO) of the data sets, it is evident that the model performs better on non-BIO data sets compared to BIO data sets. Although the model's performance on TurkishSportsNER-Child-NonBIO is slightly higher than TurkishSportsNER-Parent-BIO, the model's performance on TurkishSportsNER-Child-NonBIO surpasses the results on TurkishSports-NER-Child-BIO by approximately 2 points.

Based on the evaluation results for each fold, it can be observed that the XLM-RoBERTa-Base Multilingual Model performs most effectively on TurkishSportsNER-Parent-NonBIO while exhibiting the least accurate recognition performance on Turkish-SportsNER-Child-BIO.

The tables below present the average precision, recall, and f1-scores for each named entity type in different data sets. Generally, we observe that the O label consistently achieves the highest precision, recall, and f1-scores across all splits and data sets. Furthermore, the XLM-RoBERTa-Base Multilingual Model demonstrates its highest recognition performance on the TurkishSportsNER-Parent-NonBIO data set. As the number of labels increases, the model's overall performance decreases slightly. However, it is important to note that the scores for each split are generally very close to each other, except for when k is equal to 1, where some variations may occur.

In Table 4.60, average precision, recall, and f1-scores on TurkishSportsNER-Parent-NonBIO data set are represented. The model has a successful recognition perfor-

mance. However, it struggles to identify TEAM_SPONSOR entity.

Table 4.61 represents average precision, recall, and f1-scores on TurkishSportsNER-Parent-BIO data set. The model is unable to accurately recognize entities such as I-HEALTH_TERM, I-NATION, I-SPORT_NAME, I-SUPPORTER and I-TEAM_-SPONSOR. Conversely, it demonstrates moderate effectiveness in its overall recognition performance.

Average precision, recall, and f1-scores on TurkishSportsNER-Child-NonBIO data set were given in Table 4.62. The model has difficulties in accurately identifying entities such as BASKETBALL_TERM, HORSE_RACING_TERM, PERSON_ALIAS, TEAM_SPONSOR. Additionally, it is unable to recognize entities like TENNIS_-TERM and VOLLEYBALL_TERM effectively. In contrast, for the majority of other named entities, the model demonstrates proficient recognition performance.

In Table 4.64, we can see the f1-scores per named entity of the XLM-RoBERTa-Base Multilingual Model trained on the TurkishSportsNER-Parent-NonBIO data set. The table provides a detailed breakdown of f1-scores for different named entity types across various data splits (k=1 to k=5). Looking at specific named entity types, we observe high f1-scores for entities like SPORT_COMPETITION, SPORT_EVENT, AGREEMENT, NATION, ROLE, SPORT_LEVEL, SPORT_NAME, SPORT_ORGANIZATION, SPORT_PERSON, and TEAM. These entities show strong recognition by the model. However, there are some named entity types, such as TEAM_SPONSOR, that exhibit significantly lower f1-scores. For instance, in the second data split, the f1-score for TEAM_SPONSOR is 80.47, while for k=3, it is not recognized (f1-score 0.00). This indicates that the model struggles with accurately identifying this specific type of named entity.

Table 4.65 represents the f1-scores of the XLM-RoBERTa-Base Multilingual Model trained on the TurkishSportsNER-Parent-BIO data set. This table provides a comprehensive assessment of the model's f1-scores for various named entity types across different data splits, represented by k=1 to k=5. Upon thorough analysis, it becomes evident that the model achieves consistently high f1-scores for the O label, which denotes non-named entities, across all data splits. When examining individual named entity types, we notice significant f1-scores for various entities, such as B-ROLE,

B-SPORT_EVENT, B-SPORT_NAME, B-TEAM, and I-ALIAS. These specific entities demonstrate impressive recognition performance by the model. However, there are certain named entity types, for instance, B-TEAM_SPONSOR, that show significantly lower f1-scores. For example, in the second data split (k=2), the f1-score for TEAM_SPONSOR is 87.41, while it drops even further to 15.38 in the third data split (k=3). This suggests the model faces challenges in accurately identifying this specific type of named entity. Additionally, we notice that some named entity types, such as I-EQUIPMENT, I-HEALTH_TERM, I-NATION, I-SPORT_NAME, I-SUPPORTER, and I-TEAM_SPONSOR, have f1-scores of 0.00. This indicates that the model struggles to recognize these particular entities effectively.

Table 4.66 provides the f1-scores of the model when trained on the TurkishSportsNER-Child-NonBIO data set. The f1-score is a metric that combines precision and recall, providing a balanced measure of the model's performance for each named entity type. The evaluation is performed on different data splits represented by k=1 to k=5. Upon examining the results, it is evident that the model achieves impressive f1-scores for various named entity types. Entities such as EQUIPMENT, NATION, ROLE, SPORT_EVENT, SPORT_LEVEL, SPORT_NAME, and TEAM consistently receive high scores. However, some named entity types, including BASKETBALL_TERM, HORSE_RACING_TERM, RACE_TERM, TENNIS_TERM, and VOLLEYBALL-_TERM, exhibit significantly lower f1-scores. Nonetheless, specific named entity types like PERSON_ALIAS display notably lower f1-scores. For instance, in the fifth data split, the f1-score for PERSON_ALIAS is 78.92, whereas, for k=1, it goes unrecognized with an f1-score of only 36.61. This suggests that the model faces challenges in precisely identifying this particular type of named entity.

Table 4.67 displays the f1-scores of the XLM-RoBERTa-Base Multilingual Model trained on the TurkishSportsNER-Child-BIO data set, evaluated for different named entity types and varying values of k. Certain named entity types, such as B-NATION, B-SPORT_EVENT, B-SPORT_NAME, B-TEAM, and I-TEAM_ALIAS, achieve relatively high f1-scores. These results indicate the model's proficiency in recognizing entities related to nations, sports names, sports events, teams, and team aliases. However, the model's performance varies significantly across different named entity types. For instance, the labels B-BASKETBALL_TERM, I-HEALTH_TERM,

B-HORSE_RACING_TERM, I-NATION, B-RACE_TERM, I-SUPPORTER, and I-TENNIS_TERM show limited or no recognition at all, with f1-scores of 0.00%. This implies that the model struggles to identify these specific named entity types effectively.

In conclusion, the XLM-RoBERTa-Base Multilingual Model demonstrates promising performance on the TurkishSportsNER data sets. The model achieves high precision, recall, and f1-scores on various test data sets, indicating its proficiency in recognizing named entities. Notably, the model performs better on non-BIO data sets than BIO ones. The evaluation results reveal that the model's recognition performance is generally consistent across different data splits, except for some variations when k=1. Additionally, the model's performance is most robust on the TurkishSportsNER-Parent-NonBIO data set and weakest on the TurkishSportsNER-Child-BIO data set. Overall, the XLM-RoBERTa-Base Multilingual Model holds promise for named entity recognition in TurkishSportsNER, with the potential for further improvements through targeted optimization for specific named entity types.

Table 4.60: 5-Fold Average Precision, Recall and F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-NonBIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
O	98.88	97.60	98.23
AGREEMENT	86.93	94.24	90.23
ALIAS	89.85	90.52	90.11
EQUIPMENT	90.43	91.70	90.96
HEALTH_TERM	82.15	85.50	83.50
NATION	91.07	93.83	92.39
ROLE	93.09	96.52	94.77
SPORT_COMPETITION	88.71	92.75	90.67
SPORT_EVENT	91.29	95.64	93.37
SPORT_LEVEL	90.58	93.56	92.03
SPORT_LOCATION	87.23	93.47	90.07

Table 4.60 continued from previous page

SPORT_NAME	94.22	95.35	94.77
SPORT_ORGANIZATION	90.24	91.58	90.85
SPORT_PERSON	90.47	95.62	92.96
SPORT_TERM	83.84	90.12	86.54
SUPPORTER	90.94	91.90	91.37
TEAM	93.20	94.39	93.77
TEAM_SPONSOR	58.96	50.64	53.45

Table 4.61: 5-Fold Average Precision, Recall and F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-BIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
O	98.87	97.59	98.23
B-AGREEMENT	86.85	95.64	90.83
B-ALIAS	87.03	88.89	87.90
B-EQUIPMENT	90.56	92.63	91.56
B-HEALTH_TERM	81.09	86.93	83.74
B-NATION	90.24	94.28	92.16
B-ROLE	92.74	95.35	94.03
B-SPORT_COMPETITION	85.62	88.81	87.16
B-SPORT_EVENT	91.42	95.16	93.19
B-SPORT_LEVEL	87.55	90.75	89.05
B-SPORT_LOCATION	86.18	92.07	88.85
B-SPORT_NAME	94.36	96.18	95.24
B-SPORT_ORGANIZATION	89.73	92.75	91.15
B-SPORT_PERSON	89.30	93.16	91.18
B-SPORT_TERM	83.38	89.61	86.04
B-SUPPORTER	91.14	90.70	90.89
B-TEAM	92.80	94.71	93.71
B-TEAM_SPONSOR	78.21	50.63	60.81

Table 4.61 continued from previous page

I-AGREEMENT	85.97	62.43	72.10
I-ALIAS	93.51	94.05	93.74
I-EQUIPMENT	20.00	5.00	8.00
I-HEALTH_TERM	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00
I-ROLE	83.86	92.52	87.92
I-SPORT_COMPETITION	86.94	94.03	90.34
I-SPORT_EVENT	70.36	77.12	73.16
I-SPORT_LEVEL	87.45	91.44	89.38
I-SPORT_LOCATION	73.57	83.57	78.05
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	82.44	86.95	84.32
I-SPORT_PERSON	84.84	92.93	88.64
I-SPORT_TERM	77.71	82.60	79.90
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	84.66	82.10	83.30
I-TEAM_SPONSOR	0.00	0.00	0.00

Table 4.62: 5-Fold Average Precision, Recall and F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-NonBIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
O	98.84	97.60	98.22
AGREEMENT	86.01	94.54	89.88
BASKETBALL_TERM	40.00	2.00	3.81
COACH	86.08	86.42	86.16
EQUIPMENT	90.54	92.71	91.56
FOOTBALL_TERM	84.71	91.86	87.98
HEALTH_TERM	81.91	85.14	83.24
HORSE_RACING_TERM	6.60	18.42	9.72

Table 4.62 continued from previous page

MANAGER	73.21	87.14	79.09
NATION	91.56	93.63	92.55
PERSON_ALIAS	69.05	67.22	66.71
RACE_TERM	0.00	0.00	0.00
REFEREE	74.83	80.95	77.02
ROLE	92.90	96.52	94.67
SCORE_TERM	78.22	88.91	82.60
SPORT_COMPETITION	88.34	93.14	90.67
SPORT_EVENT	91.14	95.80	93.38
SPORT_LEVEL	90.58	93.86	92.18
SPORT_LOCATION	86.27	93.53	89.60
SPORT_NAME	94.14	95.67	94.88
SPORT_ORGANIZATION	89.92	92.28	91.02
SPORT_PLAYER	87.29	91.60	89.33
SUPPORTER	90.74	90.89	90.77
TEAM	93.39	94.29	93.81
TEAM_ALIAS	90.66	90.89	90.68
TEAM_SPONSOR	69.50	49.61	57.15
TENNIS_TERM	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00

Table 4.63: 5-Fold Average Precision, Recall and F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-BIO Data Set per Named Entity

Label	**Avg. Precision**	**Avg. Recall**	**Avg. F1-Score**
B-AGREEMENT	86.65	95.39	90.58
B-BASKETBALL_TERM	0.00	0.00	0.00
B-COACH	86.20	85.66	85.89
B-EQUIPMENT	89.83	91.69	90.69
B-FOOTBALL_TERM	84.74	91.70	87.89

Table 4.63 continued from previous page

B-HEALTH_TERM	81.83	87.16	84.26
B-HORSE_RACING_TERM	4.27	14.19	6.57
B-MANAGER	73.32	83.09	77.65
B-NATION	90.32	94.05	92.07
B-PERSON_ALIAS	67.90	67.76	66.53
B-RACE_TERM	0.00	0.00	0.00
B-REFEREE	72.42	77.01	74.05
B-ROLE	92.72	95.29	93.99
B-SCORE_TERM	76.91	87.65	81.33
B-SPORT_COMPETITION	85.72	88.81	87.23
B-SPORT_EVENT	91.44	95.18	93.22
B-SPORT_LEVEL	87.25	90.32	88.71
B-SPORT_LOCATION	85.53	92.50	88.72
B-SPORT_NAME	93.52	95.96	94.70
B-SPORT_ORGANIZATION	89.15	92.81	90.92
B-SPORT_PLAYER	85.98	89.78	87.80
B-SUPPORTER	90.57	90.59	90.54
B-TEAM	92.55	94.28	93.37
B-TEAM_ALIAS	91.16	88.91	89.96
B-TEAM_SPONSOR	76.92	42.82	54.63
B-TENNIS_TERM	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00
I-AGREEMENT	81.80	65.03	72.19
I-BASKETBALL_TERM	0.00	0.00	0.00
I-COACH	82.77	85.71	84.10
I-EQUIPMENT	20.00	3.75	6.32
I-FOOTBALL_TERM	76.15	86.05	80.66
I-HEALTH_TERM	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00
I-MANAGER	73.57	88.75	80.07
I-NATION	0.00	0.00	0.00

Table 4.63 continued from previous page

I-PERSON_ALIAS	71.37	52.15	58.22
I-RACE_TERM	0.00	0.00	0.00
I-REFEREE	74.89	84.44	79.03
I-ROLE	83.99	92.01	87.75
I-SCORE_TERM	63.41	54.02	56.94
I-SPORT_COMPETITION	86.50	94.54	90.33
I-SPORT_EVENT	68.14	75.71	71.44
I-SPORT_LEVEL	87.45	92.00	89.65
I-SPORT_LOCATION	74.45	82.69	78.20
I-SPORT_NAME	0.00	0.00	0.00
I-SPORT_ORGANIZATION	79.38	87.05	82.77
I-SPORT_PLAYER	76.69	83.71	79.89
I-SUPPORTER	0.00	0.00	0.00
I-TEAM	83.26	81.67	82.43
I-TEAM_ALIAS	90.97	91.01	90.86
I-TEAM_SPONSOR	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00
O	98.84	97.60	98.22

Table 4.64: F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.40	98.38	98.43	98.51	98.45	98.23
AGREEMENT	84.49	92.47	91.22	91.54	91.45	90.23
ALIAS	86.57	91.95	87.26	93.28	91.48	90.11
EQUIPMENT	86.99	88.96	91.36	94.27	93.22	90.96
HEALTH_TERM	74.46	83.03	89.40	83.27	87.32	83.50
NATION	92.46	92.59	89.91	93.05	93.94	92.39
ROLE	92.16	95.70	95.01	95.59	95.40	94.77

Table 4.64 Continued

SPORT_COMPETITION	87.26	92.41	91.03	90.24	92.41	90.67
SPORT_EVENT	88.55	94.27	94.31	95.09	94.65	93.37
SPORT_LEVEL	88.46	94.03	92.99	93.03	91.67	92.03
SPORT_LOCATION	83.11	89.42	90.65	93.42	93.74	90.07
SPORT_NAME	94.23	94.88	93.81	95.87	95.03	94.77
SPORT_ORGANIZATION	86.99	90.80	92.42	92.01	92.01	90.85
SPORT_PERSON	92.37	94.80	94.13	93.24	90.28	92.96
SPORT_TERM	74.33	87.08	88.79	91.75	90.72	86.54
SUPPORTER	86.65	93.28	92.23	89.97	94.70	91.37
TEAM	91.93	94.31	94.27	94.47	93.87	93.77
TEAM_SPONSOR	54.39	80.47	0.00	78.26	54.12	53.45

Table 4.65: F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.45	98.33	98.41	98.51	98.44	98.23
B-AGREEMENT	84.05	93.19	90.85	93.18	92.90	90.83
B-ALIAS	83.62	89.99	84.58	90.86	90.44	87.90
B-EQUIPMENT	85.88	93.08	91.71	94.17	92.95	91.56
B-HEALTH_TERM	78.18	84.21	87.58	85.38	83.33	83.74
B-NATION	92.27	92.66	90.43	93.01	92.44	92.16
B-ROLE	90.96	95.04	94.30	95.11	94.72	94.03
B-SPORT_COMPETITION	84.76	87.45	86.73	87.45	89.43	87.16
B-SPORT_EVENT	89.06	93.53	94.23	94.40	94.71	93.19
B-SPORT_LEVEL	83.06	91.30	91.68	89.85	89.35	89.05
B-SPORT_LOCATION	82.48	89.38	90.25	91.62	90.55	88.85
B-SPORT_NAME	94.49	94.69	95.68	96.40	94.92	95.24
B-SPORT_ORGANIZATION	85.27	91.61	93.30	92.34	93.25	91.15
B-SPORT_PERSON	90.65	92.91	92.51	91.72	88.11	91.18

Table 4.65 Continued

B-SPORT_TERM	74.42	86.12	87.79	91.56	90.30	86.04
B-SUPPORTER	84.87	93.61	92.05	90.59	93.35	90.89
B-TEAM	92.52	94.11	94.13	93.95	93.84	93.71
B-TEAM_SPONSOR	66.67	87.41	15.38	71.43	63.16	60.81
I-AGREEMENT	78.57	67.92	77.42	67.61	68.97	72.10
I-ALIAS	90.45	95.45	91.76	95.73	95.31	93.74
I-EQUIPMENT	40.00	0.00	0.00	0.00	0.00	8.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.64	87.32	87.86	89.76	90.02	87.92
I-SPORT_COMPETITION	87.80	92.61	90.18	89.69	91.40	90.34
I-SPORT_EVENT	58.76	78.24	75.97	76.62	76.19	73.16
I-SPORT_LEVEL	85.31	89.31	91.90	90.49	89.90	89.38
I-SPORT_LOCATION	63.69	79.28	81.00	82.78	83.52	78.05
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	76.10	88.33	88.62	86.21	82.35	84.32
I-SPORT_PERSON	89.07	89.97	89.37	88.88	85.91	88.64
I-SPORT_TERM	65.25	83.08	81.86	84.43	84.86	79.90
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	77.79	91.14	83.42	83.80	80.36	83.30
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.66: F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.36	98.36	98.39	98.54	98.44	98.22
AGREEMENT	84.04	92.56	89.56	91.62	91.61	89.88
BASKETBALL_TERM	0.00	0.00	9.76	0.00	9.30	3.81
COACH	87.00	91.91	78.88	87.94	85.05	86.16

Table 4.66 Continued

EQUIPMENT	87.38	92.52	91.07	93.50	93.33	91.56
FOOTBALL_TERM	78.78	86.95	90.79	92.33	91.04	87.98
HEALTH_TERM	73.07	82.18	89.77	85.49	85.71	83.24
HORSE_RACING_TERM	48.61	0.00	0.00	0.00	0.00	9.72
MANAGER	84.12	82.38	72.83	79.39	76.73	79.09
NATION	92.54	92.53	90.07	93.95	93.63	92.55
PERSON_ALIAS	36.61	71.78	70.49	75.77	78.92	66.71
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	86.56	71.73	73.49	82.37	70.97	77.02
ROLE	92.33	95.28	94.96	95.62	95.18	94.67
SCORE_TERM	65.27	85.86	85.34	87.90	88.61	82.60
SPORT_COMPETITION	87.22	93.14	90.84	89.88	92.27	90.67
SPORT_EVENT	88.29	94.60	93.95	94.95	95.10	93.38
SPORT_LEVEL	88.34	93.58	92.99	93.58	92.39	92.18
SPORT_LOCATION	83.53	88.95	90.11	93.18	92.23	89.60
SPORT_NAME	94.32	95.37	94.26	95.70	94.78	94.88
SPORT_ORGANIZATION	86.51	91.30	94.43	91.69	91.14	91.02
SPORT_PLAYER	86.62	91.71	90.34	90.82	87.16	89.33
SUPPORTER	86.16	92.81	91.46	89.62	93.80	90.77
TEAM	91.62	94.34	94.13	94.85	94.14	93.81
TEAM_ALIAS	84.75	94.10	87.99	93.24	93.34	90.68
TEAM_SPONSOR	61.33	82.93	0.00	93.02	48.48	57.15
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table 4.67: F1-Scores of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.46	98.32	98.39	98.52	98.39	98.22

Table 4.67 Continued

B-AGREEMENT	83.70	93.58	90.18	93.09	92.36	90.58
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	85.44	90.25	79.90	87.76	86.12	85.89
B-EQUIPMENT	85.71	92.13	90.85	92.65	92.11	90.69
B-FOOTBALL_TERM	80.51	85.72	90.68	91.93	90.61	87.89
B-HEALTH_TERM	76.42	83.65	89.47	85.38	86.38	84.26
B-HORSE_RACING_TERM	32.84	0.00	0.00	0.00	0.00	6.57
B-MANAGER	81.06	80.99	74.21	80.31	71.70	77.65
B-NATION	92.39	91.29	89.82	94.25	92.58	92.07
B-PERSON_ALIAS	37.50	73.21	73.21	69.74	78.96	66.53
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	85.06	75.59	73.19	76.16	60.25	74.05
B-ROLE	90.96	95.33	94.06	94.82	94.76	93.99
B-SCORE_TERM	64.11	85.34	83.46	86.93	86.83	81.33
B-SPORT_COMPETITION	85.31	88.42	85.69	87.60	89.11	87.23
B-SPORT_EVENT	88.74	93.58	94.14	94.88	94.76	93.22
B-SPORT_LEVEL	83.22	92.49	91.17	88.27	88.39	88.71
B-SPORT_LOCATION	81.84	88.79	90.02	92.48	90.48	88.72
B-SPORT_NAME	93.07	94.17	94.88	96.49	94.89	94.70
B-SPORT_ORGANIZATION	86.75	90.86	92.70	91.66	92.64	90.92
B-SPORT_PLAYER	86.72	89.94	89.53	87.98	84.82	87.80
B-SUPPORTER	85.06	92.72	91.61	89.84	93.49	90.54
B-TEAM	92.01	93.64	93.56	94.26	93.36	93.37
B-TEAM_ALIAS	84.17	93.02	87.43	91.85	93.31	89.96
B-TEAM_SPONSOR	50.00	81.30	0.00	74.07	67.80	54.63
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	73.33	66.67	84.85	69.44	66.67	72.19
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	84.45	87.70	72.22	86.88	89.24	84.10

Table 4.67 Continued

I-EQUIPMENT	31.58	0.00	0.00	0.00	0.00	6.32
I-FOOTBALL_TERM	65.35	85.65	83.92	83.96	84.43	80.66
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	84.83	81.37	77.73	79.84	76.61	80.07
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	20.00	65.52	60.87	63.41	81.32	58.22
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	87.95	76.47	78.12	80.00	72.64	79.03
I-ROLE	84.68	87.59	87.37	89.05	90.04	87.75
I-SCORE_TERM	50.60	61.82	66.04	50.00	56.25	56.94
I-SPORT_COMPETITION	87.82	92.86	89.93	89.53	91.54	90.33
I-SPORT_EVENT	56.02	76.67	73.25	74.16	77.11	71.44
I-SPORT_LEVEL	86.53	90.22	91.42	90.47	89.62	89.65
I-SPORT_LOCATION	62.03	81.13	79.42	84.54	83.85	78.20
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	76.77	84.74	86.98	84.42	80.93	82.77
I-SPORT_PLAYER	77.64	82.18	75.16	83.09	81.39	79.89
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	76.37	90.03	81.52	83.69	80.52	82.43
I-TEAM_ALIAS	84.01	93.44	89.88	93.16	93.82	90.86
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

4.8 Experiments with 5-Fold Cross-Validation Method

Table 4.68: Average Precision, Recall, F1-Score Of Models On TurkishSportsNER-Parent-NonBIO Test Data Set

	Presicison	Recall	F1-Score
BERT-Base Multilingual Cased	88.68	92.48	90.51
BERT-Base Turkish Uncased	**90.01**	94.01	91.94
ConvBERT-Base Turkish Uncased	89.73	**94.48**	**92.01**
DistilBERT-Base Turkish Cased	79.39	84.20	81.71
XLM-RoBERTa-Base Multilingual	89.46	93.49	91.41

Table 4.69: Average Precision, Recall, F1-Score Of Models On TurkishSportsNER-Parent-BIO Test Data Set

	Presicison	Recall	F1-Score
BERT-Base Multilingual Cased	86.54	90.17	88.28
BERT-Base Turkish Uncased	**88.10**	92.01	89.98
ConvBERT-Base Turkish Uncased	88.03	**92.80**	**90.32**
DistilBERT-Base Turkish Cased	71.44	79.24	75.11
XLM-RoBERTa-Base Multilingual	87.16	91.27	89.14

Table 4.70: Average Precision, Recall, F1-Score Of Models On TurkishSportsNER-Child-NonBIO Test Data Set

	Presicison	Recall	F1-Score
BERT-Base Multilingual Cased	86.82	90.15	88.43
BERT-Base Turkish Uncased	**88.43**	91.87	90.10
ConvBERT-Base Turkish Uncased	88.37	**92.21**	**90.23**
DistilBERT-Base Turkish Cased	71.98	78.40	75.04
XLM-RoBERTa-Base Multilingual	87.37	91.25	89.24

Table 4.71: Average Precision, Recall, F1-Score Of Models On TurkishSportsNER-Child-BIO Test Data Set

	Presicison	Recall	F1-Score
BERT-Base Multilingual Cased	84.71	88.52	86.54
BERT-Base Turkish Uncased	**86.79**	90.63	88.64
ConvBERT-Base Turkish Uncased	86.72	**91.24**	**88.90**
DistilBERT-Base Turkish Cased	67.98	75.31	71.44
XLM-RoBERTa-Base Multilingual	85.46	89.66	87.48

This section provides a detailed analysis of the performance of various models on different variations of the TurkishSportsNER data set, corresponding to each data split. It includes the evaluation of average precision, recall, and f1-scores. Notably, the models exhibit successful recognition capabilities when the partitioning parameter k is set to 4.

Table 4.68 and Figure 4.1 present a comprehensive overview of average precision, recall, and f1-scores attributed to various models applied to the TurkishSportsNER-Parent-NonBIO data set. Notably, the most successful results are highlighted in bold. As evident from Table 4.68, the BERT-Base Turkish Uncased model has the highest precision score, while the ConvBERT-Base Turkish Uncased model has the highest recall and f1-scores. In contrast, the DistilBERT-Base Turkish Cased model, despite its pre-training on Turkish data sets, demonstrates comparatively modest results, registering lower performance levels.

The provided table, denoted as Table 4.69, and Figure 4.2 encapsulate a summary

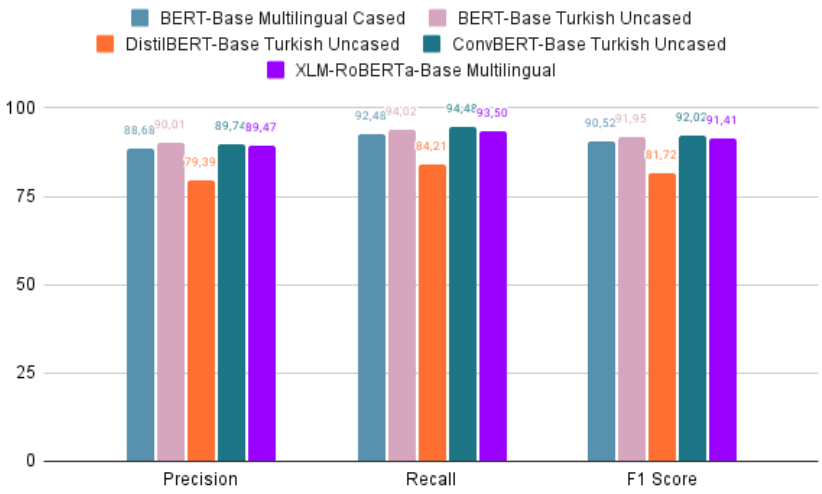


Figure 4.1: Precision, Recall, F1-Scores on TurkishSportsNER-Parent-NonBIO

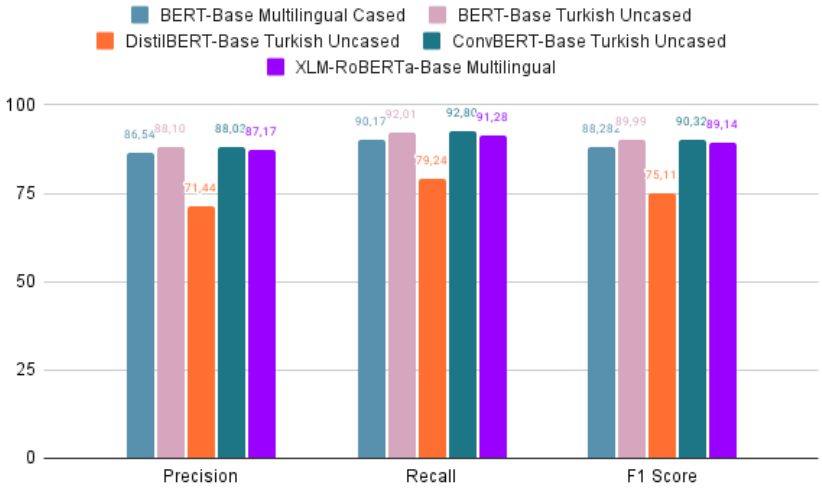


Figure 4.2: Precision, Recall, F1-Scores on TurkishSportsNER-Parent-BIO

of the average precision, recall, and f1-scores exhibited by different models when applied to the TurkishSportsNER-Parent-BIO test data set. Among the evaluated models, the BERT-Base Turkish Uncased Model achieves a slightly higher precision of 88.10%. Meanwhile, the ConvBERT-Base Turkish Uncased model registers an average precision of 88.03%, exhibiting particularly notable recall and f1-scores of 92.80% and 90.32%, respectively. In contrast, the DistilBERT-Base Turkish Cased model demonstrates comparatively modest performance.

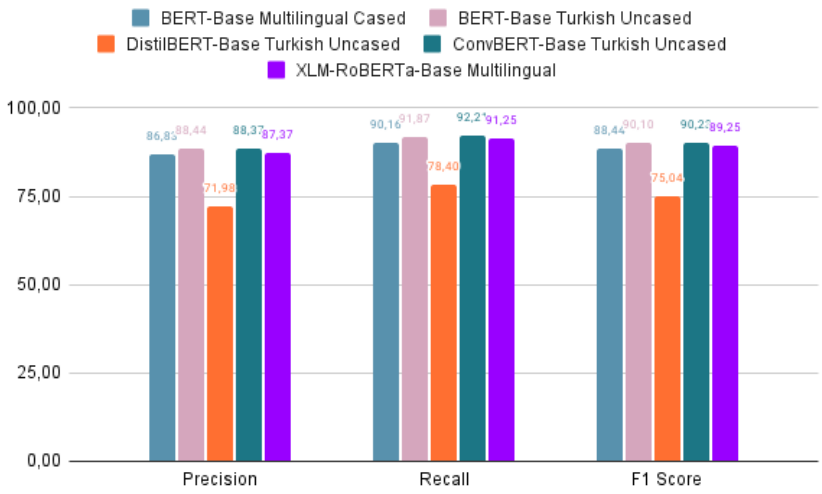


Figure 4.3: Precision, Recall, F1-Scores on TurkishSportsNER-Child-NonBIO

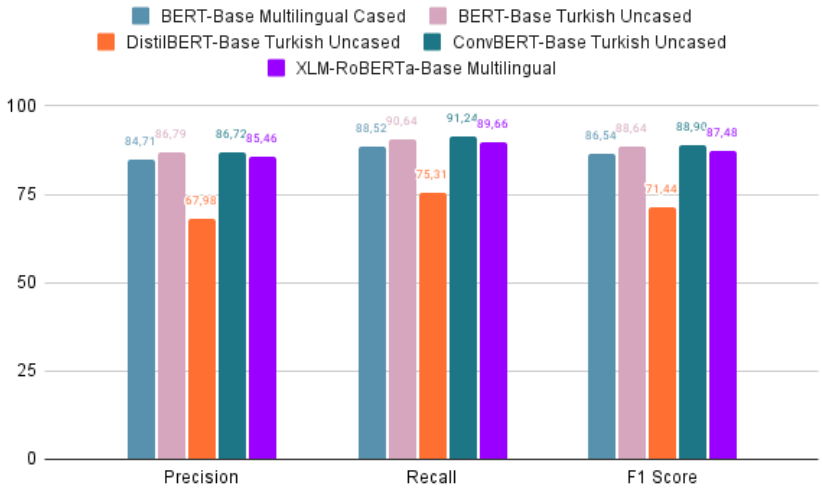


Figure 4.4: Precision, Recall, F1-Scores on TurkishSportsNER-Child-BIO

Table 4.70 and Figure 4.3 represent the average precision, recall, and f1-scores of different models on the TurkishSportsNER-Child-NonBIO test data set. Among the models evaluated, the BERT-Base Turkish Uncased Model emerges as a standout performer, boasting an average precision of 88.43%. Similarly, the ConvBERT-Base Turkish Uncased model demonstrates commendable results with notably high recall and f1-scores of 92.21% and 90.23%, respectively, underscoring its robust performance. In contrast, the DistilBERT-Base Turkish Cased model returns an average precision of 71.98%, coupled with recall of 78.40% and an f1-score of 75.04%, indi-

cating a comparatively moderate performance profile.

In Table 4.71 and Figure 4.4, average precision, recall, and f1-scores achieved by the evaluated models on the TurkishSportsNER-Child-BIO data set are represented. Remarkably, the BERT-Base Turkish Uncased Model stands out once again, registering the highest precision score among all evaluated models. In parallel, the ConvBERT-Base Turkish Uncased model achieves the most noteworthy recall and f1-scores. In contrast, the DistilBERT-Base Turkish Cased model demonstrates comparatively modest results across these performance metrics. Despite its pre-training on Turkish data sets, its precision, recall, and f1-scores are observed to be less robust than those of the other models evaluated.

Figures 4.5, 4.6, 4.7, 4.8, and 4.9 represents the precision, recall, and f1-scores for each model on distinct variants of the TurkishSportsNER data set. When considering the collective overview of average results, a general trend emerges in the ranking of model performance. Specifically, the models tend to exhibit a sequence of success, which aligns as follows: starting from the highest performing, we observe TurkishSportsNER-Parent-NonBIO, followed by TurkishSportsNER-Parent-BIO, then TurkishSportsNER-Child-NonBIO, and lastly, TurkishSportsNER-Child-BIO. This observed ranking order points to a significant insight into the complexity of the data sets. Essentially, an increase in data set complexity is associated with a discernible decrease in model performance. This observation highlights that models tend to exhibit reduced effectiveness when confronted with more complex data sets. This trend emphasizes the challenges presented by complex data sets and offers a valuable perspective on the relationship between data set complexity and model performance.

Table 4.72 represents a comparative analysis of various BERT model variations on the TurkishSportsNER-Parent-NonBIO data set. The data set involves the annotation of Turkish sports-related terms across distinct label categories. The models under examination encompass BERT-Base Multilingual Cased, BERT-Base Turkish Uncased, ConvBERT-Base Turkish Uncased, DistilBERT-Base Turkish Cased, and XLM-RoBERTa-Base Multilingual. The evaluation criterion is centered around the f1-score, an amalgamation of precision and recall metrics, signifying the models' labeling efficacy. Notably, Table 4.72 highlights the highest f1-scores per label cat-

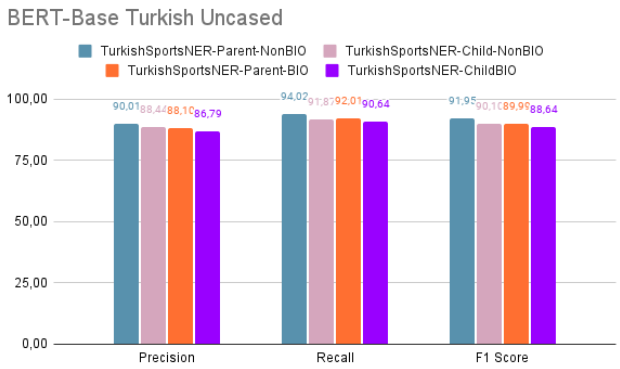


Figure 4.5: Precision, Recall, F1-Scores of BERT-Base Turkish Uncased Model on variants of TurkishSportsNER data set

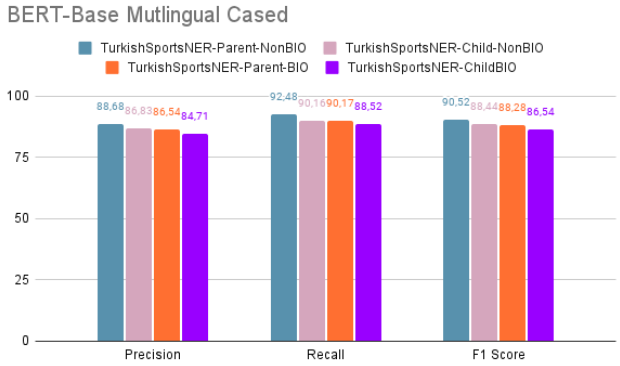


Figure 4.6: Precision, Recall, F1-Scores of BERT-Base Multilingual Cased Model on variants of TurkishSportsNER data set

egory, typically denoted in bold font. The ConvBERT-Base Turkish Uncased model emerges as a consistent performer across various categories, capturing the essence of its robust feature representation. All evaluated models demonstrate proficient performance in identifying the O (Other) named entity category, achieving scores exceeding 97.20%. Specifically, the BERT-Base Turkish Uncased model achieves noteworthy competence in recognizing NATION entities, attaining an accuracy rate of 95.35%. The ConvBERT-Base Turkish Uncased model excels in identifying entities such as ROLE, SPORT_PERSON, and TEAM, surpassing the 95% threshold. Similarly, the XLM-RoBERTa-Base Multilingual model exhibits remarkable f1-scores in the SPORT_EVENT and SPORT_NAME categories, both exceeding the 95% mark.

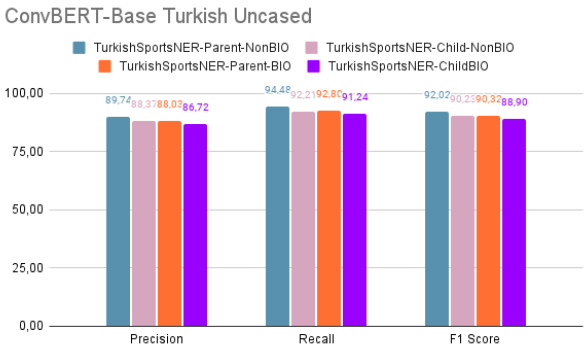


Figure 4.7: Precision, Recall, F1-Scores of ConvBERT-Base Turkish Uncased Model on variants of TurkishSportsNER data set

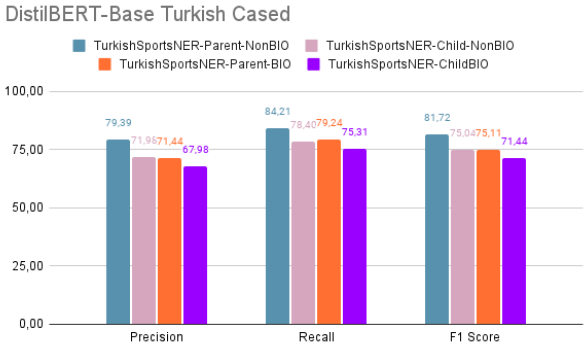


Figure 4.8: Precision, Recall, F1-Scores of DistilBERT-Base Turkish Cased Model on variants of TurkishSportsNER data set

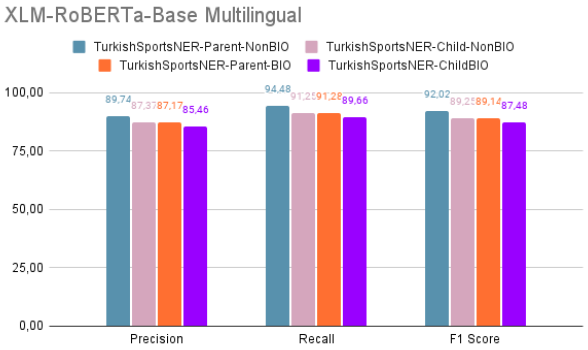


Figure 4.9: Precision, Recall, F1-Scores of XLM-RoBERTa-Base Multilingual Model on variants of TurkishSportsNER data set

Conversely, consistent across all models is the moderate proficiency observed in the TEAM_SPONSOR category, with all models registering recognition rates below 81%. Worth noting is the DistilBERT-Base Turkish Cased model, which demonstrates modest performance across all named entity categories.

Table 4.73 and Figures 4.14, 4.15 represent an in-depth analysis of several BERT model variants' performance on the TurkishSportsNER-Parent-BIO data set, utilizing a partitioning approach with k equals 4. The models are evaluated across various named entities, employing the f1-score as the primary quantitative metric for assessment. Primarily, all models consistently showcase substantial accuracy in recognizing the ubiquitous O (Other) named entity category, with f1-scores consistently surpassing the 97.17% threshold. This consistency underscores the robustness of these models across different contexts and categories. Notably, the ConvBERT-Base Turkish Uncased model emerges as a consistently high-performer across multiple of these categories, underscoring its versatility and efficacy. Certain categories stand out due to their exceptional f1-scores, particularly those associated with the B-NATION, B-ROLE, B-SPORT_EVENT, B-SPORT_NAME, B-TEAM, and I-ALIAS entities. Conversely, challenges become evident in certain categories such as I-EQUIPMENT, I-HEALTH_TERM, I-NATION, I-SPORT_NAME, and I-SUPPORTER as indicated by comparatively lower f1-scores. These categories appear to present intricate contextual nuances, posing difficulties in accurate identification and classification. Of significance, the BERT-Base Turkish Uncased model demonstrates exceptional capability in identifying both B-HEALTH_TERM and I-SPORT_LOCATION entities boasting a prominent f1-score of approximately 3 points over other models. While demonstrating commendable proficiency in recognizing common entities and certain categories, these models face complexity and variability in handling more intricate entity classifications.

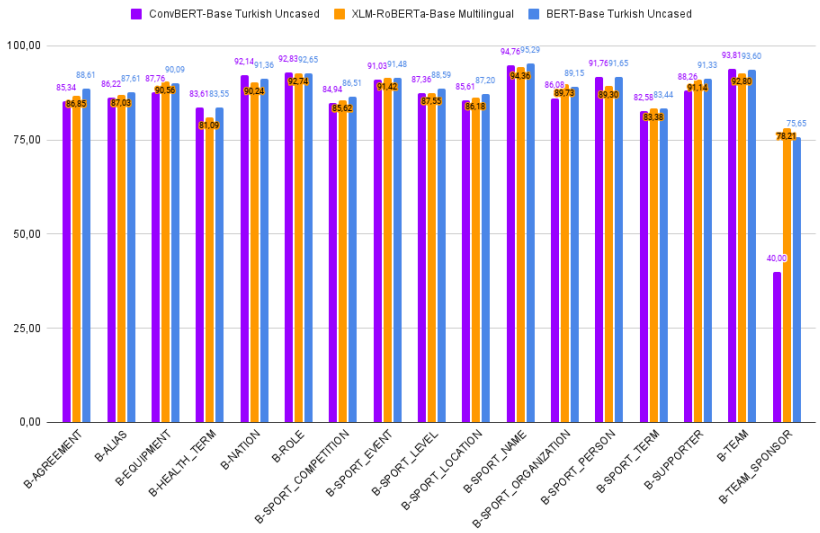


Figure 4.10: Label-wise Precision Values of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

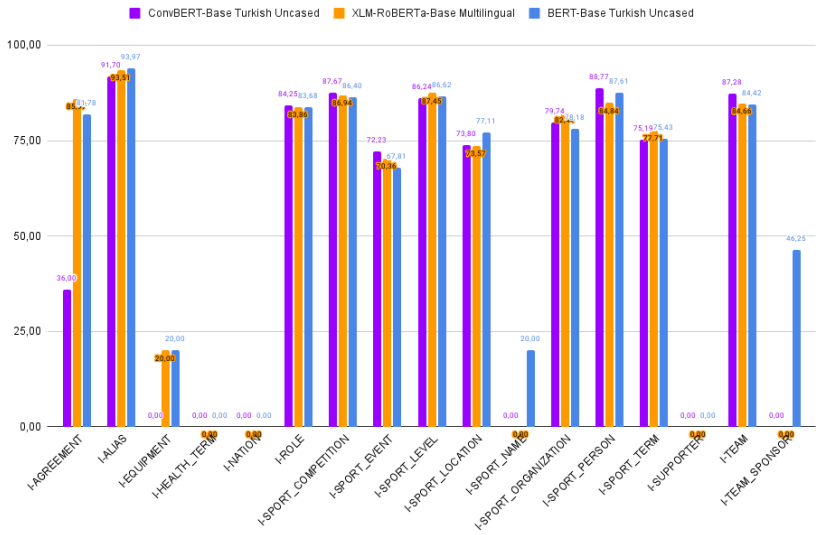


Figure 4.11: Label-wise Precision Values of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

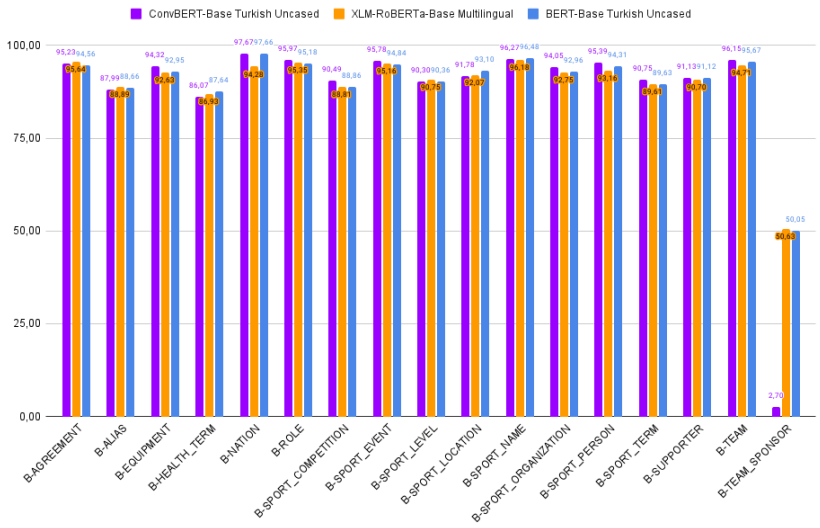


Figure 4.12: Label-wise Recall Values of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

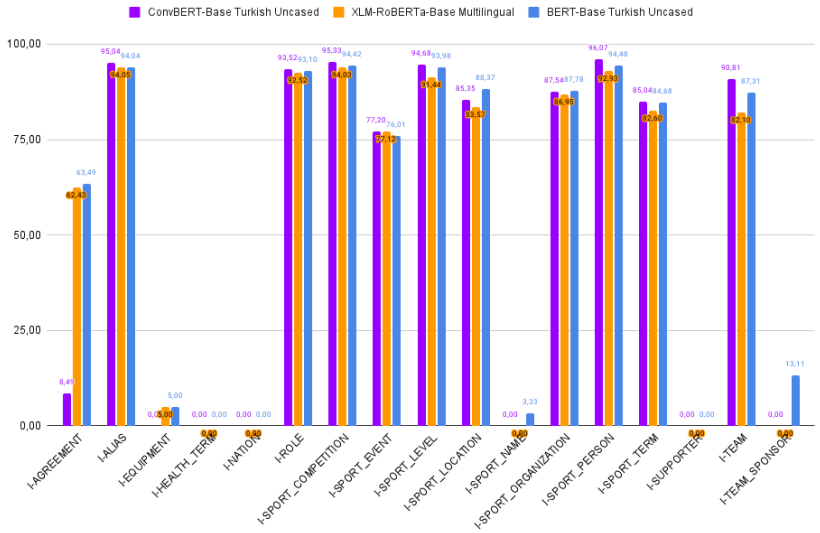


Figure 4.13: Label-wise Recall Values of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

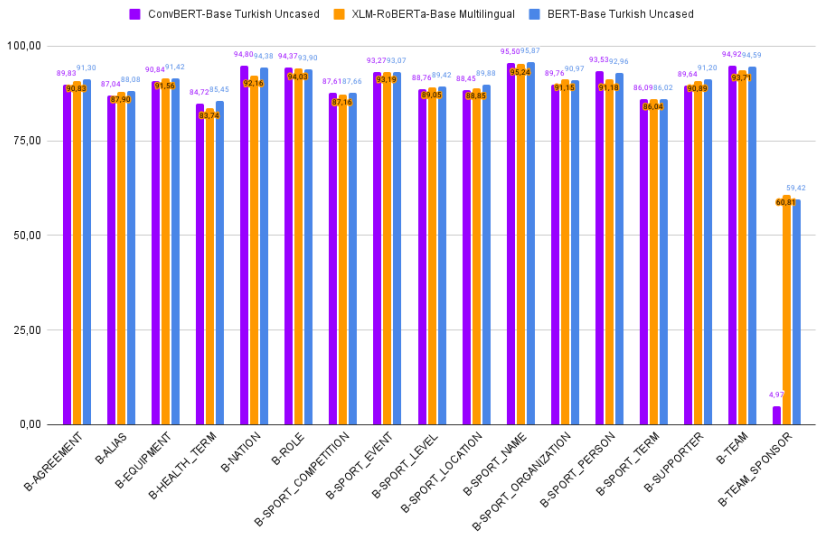


Figure 4.14: Label-wise F1-Scores of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

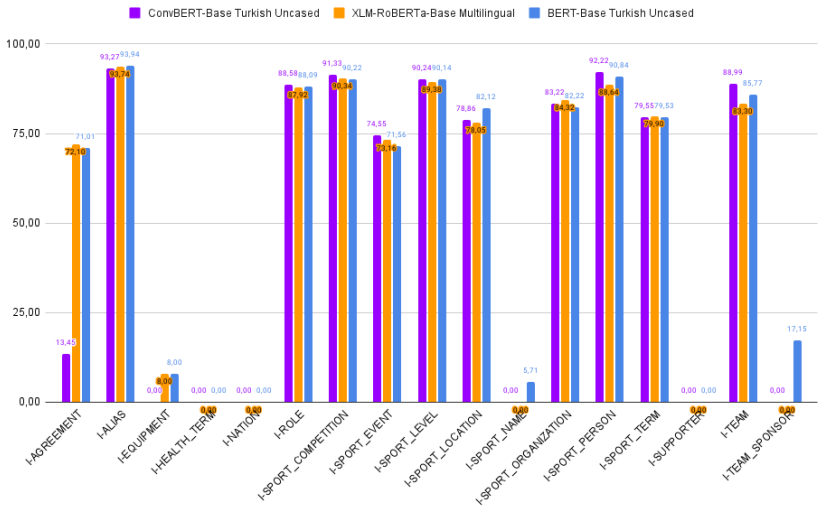


Figure 4.15: Label-wise F1-Scores of ConverBERT-Base Turkish Uncased, XLM-RoBERTa-Base Multilingual Model and BERT-Base Turkish Uncased on TurkishSportsNER-Parent-BIO data set

Table 4.74 represents the f1-scores of the evaluated models on the TurkishSportsNER-Child-NonBIO data set in partition k is equal to 4. Firstly, consistent proficiency is evident across models in recognizing the O (Other) named entity category, consistently surpassing a remarkable 97.16% f1-score. Distinguished by their highest f1-scores, the ConvBERT-Base Turkish Uncased and BERT-Base Turkish Uncased models demonstrate remarkable versatility across an array of categories. These models excel in designations like COACH, EQUIPMENT, FOOTBALL_TERM, NATION, ROLE, SPORT_COMPETITION, SPORT_EVENT, SPORT_LOCATION, SPORT_-ORGANIZATION, SPORT_PLAYER, TEAM, and TEAM_ALIAS. Such consistency accentuates their ability to discern and interpret intricate contextual nuances. Challenges arise in categories such as BASKETBALL_TERM, HORSE_RACING_TERM, RACE_TERM, TENNIS_TERM, and VOLLEYBALL_TERM, wherein models register lower f1-scores or even null performance. These entities seem to engender complexities that impede accurate classification. The XLM-RoBERTa-Base Multilingual model distinguishes itself by registering the highest f1-scores in the REFEREE and SPORT_LEVEL categories. This attests to its exceptional comprehension of these entities' contextual intricacies. However, pronounced variations emerge in the TEAM_SPONSOR category, indicating the intricacies in precisely identifying this entity. Remarkably, the XLM-RoBERTa-Base Multilingual model showcases the most pronounced capability here. While exhibiting commendable proficiency in recognizing entities, these models confront challenges in addressing intricate and less frequent categories. In addition, the models that were pre-trained on Turkish data sets are more successful compared to other models.

F1-scores of distinct BERT model variants applied to the TurkishSportsNER-Child-BIO data set in partition k=4 were represented in Table 4.75. Highly commendable performance is evident in the recognition of entity tagged as B-SPORT_NAME. Conversely, named entities like B-BASKETBALL_TERM, B-HORSE_RACING_TERM, and B-TENNIS_TERM yield consistently low scores across all models. This trend may signify the inherent challenge of differentiating sport-specific terminologies that are relatively less common. Entities characterized by B-EQUIPMENT are proficiently identified by most models. The BERT-Base Turkish Uncased model notably excels in this category, demonstrating its aptitude for recognizing sport-related

equipment references. Notably, B-MANAGER classification is challenging across all models, with the BERT-Base Turkish Uncased model leading in performance. This observation could reflect the intricacies involved in accurately identifying managerial people within the sports domain. Entities designated as B-PERSON_ALIAS pose challenges, particularly for models like ConvBERT-Base Turkish Uncased and DistilBERT-Base Turkish Cased which register comparatively lower f1-scores. This could be attributed to the complexity of recognizing aliases and pseudonyms within named entities. Furthermore, the models consistently excel in identifying O (Other) entities, reinforcing their adeptness in recognizing text segments that do not correspond to any named entity. While excelling in some categories, the models encounter challenges in accurately classifying less common and complex entities. The comparison discerns nuanced differences in their proficiency, further enriching the discourse on their practical utility in the domain of sports-related entity recognition.

The comparative assessment of various BERT model variants across different TurkishSportsNER data sets reveals distinct performance patterns. Analyzing the f1-scores achieved on all the data sets, the ConvBERT-Base Turkish Uncased model consistently stands out, excelling in multiple categories. Across various data sets, common trends emerge, with models generally proficient in identifying O (Other) entities. Additionally, models pre-trained on Turkish data exhibit better performance. However, challenges are evident in recognizing specialized terms such as sport-specific terminologies, and identifying entities that start with the "I-" prefix. The analyses underscore the complexity of sports-related entity recognition, shedding light on models' strengths and limitations, particularly in addressing nuanced contextual cues and less common categories.

In conclusion, evaluating BERT model variants across distinct TurkishSportsNER data sets, a discernible trend emerges in the hierarchy of model performance. Specifically, the ranking follows a pattern from the highest-performing data set, TurkishSportsNER-Parent-NonBIO, through TurkishSportsNER-Parent-BIO, TurkishSportsNER-Child-NonBIO, to the least-performing TurkishSportsNER-Child-BIO data set. This ranking signifies an intriguing correlation between data set complexity and model efficacy. It is observed that with an increase in data set complexity, there is a corresponding decrease in the performance of the models. This observation underscores

the inherent challenges presented by intricate data sets and offers valuable insights into the dynamic relationship between data complexity and model competence. The analysis of diverse BERT model variations across these data sets reveals distinct yet consistent performance trends. Notably, the ConvBERT-Base Turkish Uncased model emerges as a consistently high performer across multiple named entities. Models typically show proficiency in identifying 'O' (non-entity) categories, yet they encounter challenges when dealing with domain-specific terms and complex entities.

These issues are caused by an imbalance in the training data set. In cases where entities labeled with "I-" are significantly less represented compared to those labeled with "B-", the model often tends to incorrectly classify "I-" entities as "B-". Furthermore, the models face challenges in accurately identifying child entities of SPORT_TERM, as these entities have less number of samples in training data sets and often contain common words that the models struggle to contextualize appropriately. To mitigate the identified issues, a dual strategy is recommended. Firstly, to address the imbalance in "I-" labeled entities, it is crucial to augment the training data set with a greater number of such instances. This can be achieved either by collecting additional data or employing synthetic data generation methods. Secondly, for the child entities of SPORT_TERM, it is vital to enrich the data set with a broader array of context-specific examples. This entails incorporating diverse contextual representations of common sports terms, thereby aiding the model in differentiating between them more effectively. Implementing these approaches is anticipated to significantly enhance the model's proficiency in accurately classifying both "I-" prefixed entities and distinct SPORT_TERM entities, thereby improving its overall performance and reliability in sports-related named entity recognition tasks.

Table 4.72: F1-Scores on TurkishSportsNER-Parent-NonBIO in split k=4

	BERT-Base Multilingual Cased	**BERT-Base Turkish Uncased**	**ConvBERT-Base Turkish Uncased**	**DistilBERT-Base Turkish Uncased**	**XLM-RoBERTa-Base Multilingual**
O	98.35	98.67	**98.69**	97.21	98.51
AGREEMENT	90.99	**92.01**	91.71	89.18	91.54
ALIAS	92.50	94.33	**94.74**	84.30	93.28
EQUIPMENT	92.73	92.95	93.10	87.93	**94.27**
HEALTH_TERM	83.95	83.79	**84.43**	70.72	83.27
NATION	91.07	**95.35**	95.24	76.51	93.05
ROLE	95.40	95.59	**95.98**	91.19	95.59
SPORT_COMPETITION	89.52	90.37	**91.96**	72.84	90.24
SPORT_EVENT	94.66	94.97	94.91	90.24	**95.09**
SPORT_LEVEL	**93.41**	92.70	93.25	82.85	93.03
SPORT_LOCATION	91.98	**93.63**	92.43	74.03	93.42
SPORT_NAME	95.00	95.39	95.85	89.74	**95.87**
SPORT_ORGANIZATION	92.11	91.31	**92.58**	74.06	92.01
SPORT_PERSON	92.60	95.12	**95.20**	81.64	93.24
SPORT_TERM	91.12	92.02	**92.22**	84.31	91.75

Table 4.72 Continued

SUPPORTER	**90.73**	90.54	89.69	85.03	89.97
TEAM	93.34	95.26	**95.45**	82.64	94.47
TEAM_SPONSOR	58.82	78.43	**80.95**	23.08	78.26

Table 4.73: F1-Scores on TurkishSportsNER-Parent-BIO in split k=4

	BERT-Base Multilingual Cased	**BERT-Base Turkish Uncased**	**ConvBERT-Base Turkish Uncased**	**DistilBERT-Base Turkish Uncased**	**XLM-RoBERTa-Base Multilingual**
O	98.44	98.66	**98.68**	97.17	98.51
B-AGREEMENT	92.95	**93.71**	92.13	87.33	93.18
B-ALIAS	90.23	91.66	**92.14**	76.58	90.86
B-EQUIPMENT	93.30	93.65	93.10	88.03	**94.17**
B-HEALTH_TERM	87.39	**90.08**	86.42	61.08	85.38
B-NATION	92.14	95.24	**95.59**	72.34	93.01
B-ROLE	94.66	95.29	**95.81**	90.28	95.11
B-SPORT_COMPETITION	87.07	**88.75**	87.55	61.79	87.45
B-SPORT_EVENT	94.44	**94.77**	95.20	89.44	94.40
B-SPORT_LEVEL	89.09	**90.58**	89.81	78.68	89.85

Table 4.73 Continued

B-SPORT_LOCATION	**92.25**	92.23	90.88	74.61	91.62
B-SPORT_NAME	95.94	**96.47**	96.30	89.63	96.40
B-SPORT_ORGANIZATION	91.53	92.05	91.02	71.98	**92.34**
B-SPORT_PERSON	91.07	93.54	**93.84**	74.10	91.72
B-SPORT_TERM	90.80	91.51	**91.73**	82.87	91.56
B-SUPPORTER	**90.65**	**90.65**	89.87	84.34	90.59
B-TEAM	93.46	94.95	**95.50**	82.25	93.95
B-TEAM_SPONSOR	**78.57**	74.07	0.00	0.00	71.43
I-AGREEMENT	**71.43**	69.57	0.00	0.00	67.61
I-ALIAS	94.01	95.36	**96.73**	91.54	95.73
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00
I-ROLE	88.09	**90.39**	88.62	73.08	89.76
I-SPORT_COMPETITION	90.08	88.42	**91.40**	71.85	89.69
I-SPORT_EVENT	78.19	76.06	**80.45**	17.73	76.62
I-SPORT_LEVEL	91.33	91.03	**91.41**	79.67	90.49
I-SPORT_LOCATION	84.35	**87.73**	85.36	12.12	82.78

Table 4.73 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	83.61	83.56	84.44	52.67	**86.21**
I-SPORT_PERSON	87.77	90.86	**92.71**	66.27	88.88
I-SPORT_TERM	82.99	85.01	**85.03**	65.42	84.43
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00
I-TEAM	82.73	86.99	**89.31**	51.46	83.80
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00

Table 4.74: F1-Scores on TurkishSportsNER-Child-NonBIO in split k=4

	BERT-Base Multilingual Cased	**BERT-Base Turkish Uncased**	**ConvBERT-Base Turkish Uncased**	**DistilBERT-Base Turkish Uncased**	**XLM-RoBERTa-Base Multilingual**
O	98.41	98.65	**98.69**	97.16	98.54
AGREEMENT	91.53	91.84	**92.54**	88.52	91.62
BASKETBALL_TERM	0.00	**59.13**	0.00	0.00	0.00
COACH	87.47	90.52	**91.35**	59.84	87.94
EQUIPMENT	93.13	**94.45**	93.24	88.44	93.50
FOOTBALL_TERM	91.54	**93.07**	92.70	84.79	92.33

Table 4.74 Continued

HEALTH_TERM	85.25	89.54	**89.57**	70.00	85.49
HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00
MANAGER	82.75	**85.02**	83.74	62.76	79.39
NATION	92.19	**95.45**	94.91	72.43	93.95
PERSON_ALIAS	74.83	**79.35**	62.95	0.00	75.77
RACE_TERM	0.00	0.00	0.00	0.00	0.00
REFEREE	74.58	81.01	78.76	0.00	**82.37**
ROLE	95.33	95.59	**95.90**	90.72	95.62
SCORE_TERM	87.33	**88.37**	87.61	77.15	87.90
SPORT_COMPETITION	89.68	90.88	**91.46**	71.55	89.88
SPORT_EVENT	94.45	**95.14**	94.86	90.09	94.95
SPORT_LEVEL	92.83	93.06	92.13	82.17	**93.58**
SPORT_LOCATION	93.01	**93.34**	92.69	73.75	93.18
SPORT_NAME	95.25	95.67	95.67	90.14	**95.70**
SPORT_ORGANIZATION	90.94	92.28	**92.39**	72.73	91.69
SPORT_PLAYER	90.38	92.68	**94.06**	70.51	90.82
SUPPORTER	90.18	**90.68**	89.76	84.25	89.62
TEAM	93.47	95.15	**95.43**	81.21	94.85

Table 4.74 Continued

TEAM_ALIAS	93.41	**95.21**	92.00	81.47	93.24
TEAM_SPONSOR	54.05	78.43	56.25	0.00	**93.02**
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00

Table 4.75: F1-Scores on TurkishSportsNER-Child-BIO in split k=4

	BERT-Base Multilingual Cased	**BERT-Base Turkish Uncased**	**ConvBERT-Base Turkish Uncased**	**DistilBERT-Base Turkish Uncased**	**XLM-RoBERTa-Base Multilingual**
O	98.42	98.68	**98.70**	97.14	98.52
B-AGREEMENT	**93.99**	93.30	91.21	87.99	93.09
B-BASKETBALL_TERM	0.00	**3.92**	0.00	0.00	0.00
B-COACH	85.63	90.36	**91.58**	54.19	87.76
B-EQUIPMENT	92.35	**93.30**	93.01	88.13	92.65
B-FOOTBALL_TERM	91.63	92.32	**92.85**	84.73	91.93
B-HEALTH_TERM	88.24	**90.08**	85.14	68.89	85.38
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00
B-MANAGER	78.02	**85.19**	80.74	53.30	80.31

Table 4.75 Continued

B-NATION	90.58	95.31	**95.67**	72.81	94.25
B-PERSON_ALIAS	77.49	**79.61**	38.30	0.00	69.74
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00
B-REFEREE	70.08	**77.01**	73.91	5.46	76.16
B-ROLE	94.66	94.67	**95.56**	89.52	94.82
B-SCORE_TERM	86.52	**88.56**	86.16	77.67	86.93
B-SPORT_COMPETITION	87.04	**87.85**	87.20	60.53	87.60
B-SPORT_EVENT	94.32	94.68	94.73	89.19	**94.88**
B-SPORT_LEVEL	**89.88**	88.43	88.17	77.49	88.27
B-SPORT_LOCATION	91.32	91.94	**92.61**	73.27	92.48
B-SPORT_NAME	95.95	96.39	**96.64**	89.30	96.49
B-SPORT_ORGANIZATION	91.18	92.19	**92.37**	70.70	91.66
B-SPORT_PLAYER	87.84	91.53	**93.06**	65.35	87.98
B-SUPPORTER	89.89	**90.79**	88.85	84.68	89.84
B-TEAM	93.50	94.76	**95.52**	81.53	94.26
B-TEAM_ALIAS	91.88	**94.78**	89.50	80.22	91.85
B-TEAM_SPONSOR	74.07	**82.76**	0.00	0.00	74.07
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00

Table 4.75 Continued

B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	**69.57**	67.61	0.00	0.00	69.44
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00
I-COACH	84.81	**91.00**	90.59	63.08	86.88
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	83.36	85.02	**85.99**	67.89	83.96
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00
I-MANAGER	79.43	**83.30**	83.14	53.45	79.84
I-NATION	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	58.97	**85.15**	0.00	0.00	63.41
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00
I-REFEREE	75.00	**80.48**	79.07	0.00	80.00
I-ROLE	87.35	88.78	**89.95**	74.26	89.05
I-SCORE_TERM	51.72	**74.36**	14.29	0.00	50.00
I-SPORT_COMPETITION	89.93	88.79	**90.47**	71.63	89.53
I-SPORT_EVENT	75.35	**76.22**	75.90	21.26	74.16
I-SPORT_LEVEL	90.73	90.73	**91.23**	78.21	90.47

Table 4.75 Continued

I-SPORT_LOCATION	81.90	85.80	**88.18**	1.28	84.54
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	82.87	83.29	**84.85**	55.68	84.42
I-SPORT_PLAYER	82.36	88.44	**90.35**	48.97	83.09
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00
I-TEAM	82.50	86.62	**89.08**	47.02	83.69
I-TEAM_ALIAS	91.07	**96.79**	88.20	85.46	93.16
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00

4.9 Model and Named Entity Specific Explainability Results with LIME

In this section, explanations of the models by implementing LIME were given. These empirical investigations were conducted within the Google Colab platform, employing the Python programming language version 3.10. The pivotal tools employed encompassed the Transformers package, esteemed for its version 4.27.4, supplemented seamlessly by the indispensable Eli5 package. The number of samples was configured at 200.

In the experiment results that were given in the figures, green colors represent the token having a positive contribution to the predicted label, and red colors represent the token having a negative contribution.

4.9.1 Explanations with ConvBERT-Base Turkish Uncased Model

The model trained on the TurkishSportsNER-Parent-NonBIO data set, with a partitioning strategy employing a k value of 4, yielded the most successful outcomes. As a result, all experiments were consistently conducted using the meticulously fine-tuned ConvBERT-Base Turkish Uncased Model within this specific data set and partitioning scheme.

In reference to Figure 4.16, the intricate analysis pertains to the semantic dissection of the sentence *emre belezoğlu bugün fenerbahçe anlaştı imparator faith terim galatasaray anlaşma imzaladı*. The term *emre* is subject to meticulous examination within this context. The underlying ConvBERT-Base Turkish Uncased model aptly ascertained the label of SPORT_PERSON, thereby accurately capturing the context's nature. Notably, this determination was substantiated by a predictive probability of 22%. Delving into the contributing elements of this prediction, the terms *belezoğlu, fenerbahçe, fatih, terim, galatasaray*, and *imzaladı* collectively bolstered the model's accuracy by offering a positive influence on the predicted label. Conversely, the words *bugün, anlaştı, imparator, anlaşma* were identified as having a negative connotation within this predictive context, subsequently impacting the model's prediction. This analytical dissection provides a scholarly insight into the intricate dynamics of the BERT model's decision-making process, elucidating how certain words contribute

Figure 4.16: Explanation of the word *emre* in the sentence *emre belezoğlu bugün fenerbahçe anlaştı imparator faith terim galatasaray anlaşma imzaladı* by LIME.

positively or negatively to the outcome of the label prediction.

Figure 4.17 provides a visual representation of feature contributions in relation to specific named entity tags within the analyzed sentence. In Figure 4.17a, it becomes evident that terms associated with agreements, such as *anlaşma* and *anlaştı*, yield positive contributions toward the prediction of the AGREEMENT named entity. Interestingly, despite *imzaladı* is conceptually linked to agreement terms, it exhibits a negative contribution in predicting the AGREEMENT named entity. Moving to Figure 4.17b, discernible trends emerge as the names of sports personalities, including *emre, belezoğlu, fatih*, and *terim* alongside designations of sports clubs like *fenerbahçe* and *galatasaray*, collectively contribute positively to the prediction of the SPORT_PERSON named entity. Conversely, the alias *imparator* and agreement-

y=AGREEMENT top features	
Weight[?]	**Feature**
+0.931	[9] anlaşma
+0.916	[4] anlaştı
+0.528	[7] terim
+0.394	[2] bugün
+0.239	[5] imparator
-0.191	[3] fenerbahçe
-0.283	[10] imzaladı
-0.308	[8] galatasaray
-0.673	[1] belezoğlu
-1.421	[0] emre
-4.143	<BIAS>

(a) AGREEMENT

y=SPORT_PERSON top features	
Weight[?]	**Feature**
+1.003	[0] emre
+0.816	[1] belezoğlu
+0.397	[6] fatih
+0.122	[3] fenerbahçe
+0.105	[7] terim
+0.103	[8] galatasaray
+0.050	[10] imzaladı
-0.127	[2] bugün
-0.240	[5] imparator
-0.313	[4] anlaştı
-0.320	[9] anlaşma
-3.201	<BIAS>

(b) SPORT_PERSON

Figure 4.17: Contributions per Feature for Specific Named Entity Tag

associated terminology manifest adverse effects on the predictive accuracy of this classification. These observations reflect the nuanced interplay between language elements and named entity prediction, emphasizing the contextual sensitivity inherent in such endeavors.

In Figure 4.18, the term *galatasaray* within the sentence *tff bugün açıklanacak karar doğrultusunda fenerbahçe galatasaray beşiktaşın içinde bulunduğu üç büyüklere ceza verecek mi?* undergoes explication through the utilization of the LIME. The outcome of this explication demonstrates the model's inference of the named entity label *TEAM* with a corresponding probability of 81.5%. The contributing factors to this prediction are notable, with terms such as *tff, bugün, açıkalanacak, karar, doğrultusunda*, and *fenerbahçe* holding positive influences on the predictive outcome. It is paramount to mention that the second target label, in this context, is classified as *O*

Figure 4.18: Explanation of the word *galatasaray* in the sentence *tff bugün açıklanacak karar doğrultusunda fenerbahçe galatasaray beşiktaşın içinde bulunduğu üç büyüklere ceza verecek mi?* by LIME.

(Other), signifying those segments of the sentence that do not align with any named entity categorization.

Figure 4.19 showcases the interpretative elucidation of the term *antremanlar* embedded within the sentence *milli takım gittiği kampta başarılı antremanlar yaptı* facilitated by the LIME methodology. The predictive outcome of the model designates the named entity label SPORT_EVENT with an associated probability of 92.9%. Notably, the entirety of words constituting the sentence exhibits affirmative contributions towards this predictive inference. Such interpretative frameworks play a crucial role in unraveling the intricate nuances governing the model's decision-making, thereby enriching the comprehension of named entity recognition in the domain of sports-related text data.

In Figure 4.20, the term *aslanlar* within the sentence *cimbom bu maçta aslanlar gibi savaştı* is subject to explanation through the application of the LIME technique. How-

WORD TO EXPLAIN antremanlar

y=SPORT_EVENT (probability **0.929**, score **2.844**) top features

Contribution?	Feature
+12.456	Highlighted in text (sum)
-9.612	<BIAS>

milli takım gittiği kampta başarılı antremanlar yaptı

y=O (probability **0.034**, score **-3.327**) top features

Contribution?	Feature
+6.330	<BIAS>
-9.657	Highlighted in text (sum)

milli takım gittiği kampta başarılı antremanlar yaptı

y=SPORT_PERSON (probability **0.008**, score **-4.752**) top features

Contribution?	Feature
-0.621	Highlighted in text (sum)
-4.130	<BIAS>

milli takım gittiği kampta başarılı antremanlar yaptı

y=SPORT_TERM (probability **0.007**, score **-4.953**) top features

Contribution?	Feature
-0.209	Highlighted in text (sum)
-4.744	<BIAS>

milli takım gittiği kampta başarılı antremanlar yaptı

y=TEAM (probability **0.006**, score **-5.055**) top features

Contribution?	Feature
-0.147	Highlighted in text (sum)
-4.908	<BIAS>

milli takım gittiği kampta başarılı antremanlar yaptı

Figure 4.19: Explanation of the word *antremanlar* in the sentence *milli takım gittiği kampta başarılı antremanlar yaptı* by LIME.

ever, unlike the previously discussed examples, the model's prediction is erroneous, designating the term as an ALIAS with a probability of 85.8%. Notably, *aslanlar* denoting *lions* in English can indeed function as an alias for the sports team *galatasaray*.

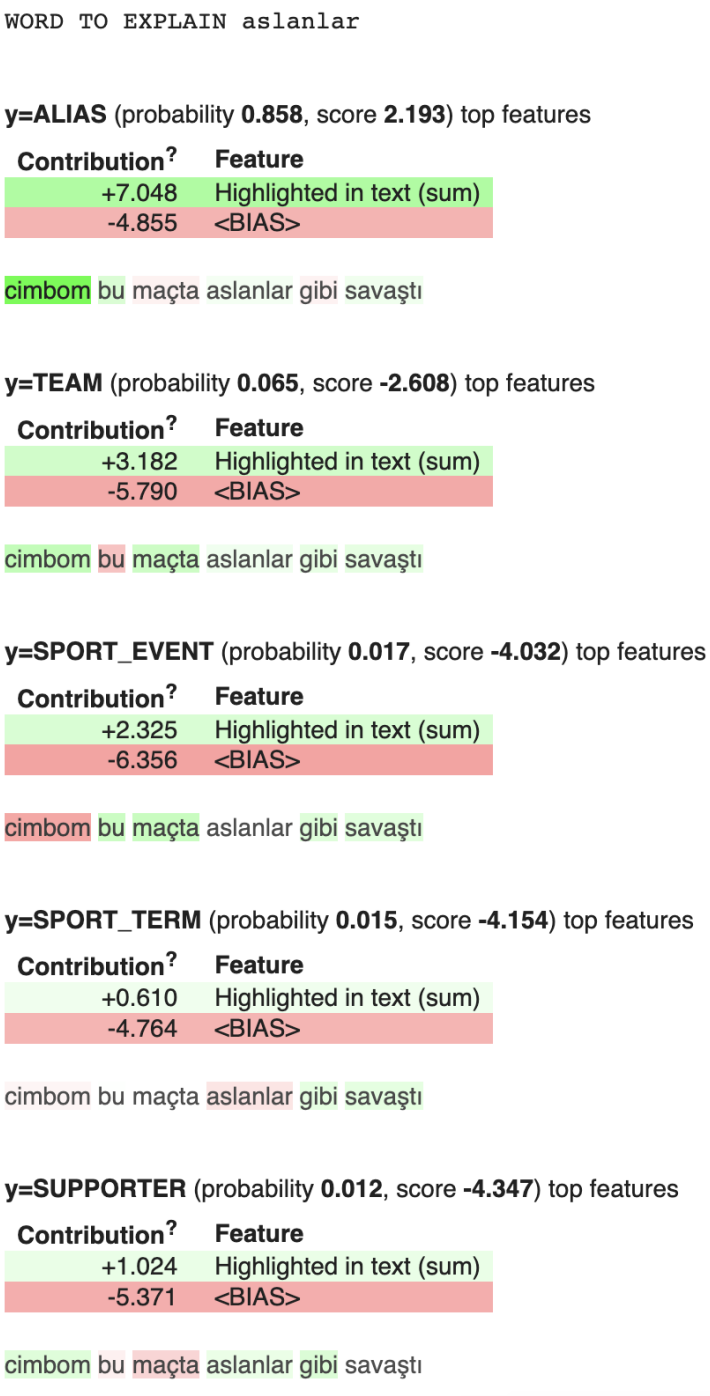


Figure 4.20: Explanation of the word *aslanlar* in the sentence *cimbom bu maçta aslanlar gibi savaştı* by LIME.

Nevertheless, in this particular context, the term *aslanlar* is employed adverbially to convey a challenging connotation. Hence, the model's misclassification can be attributed to its inability to discern the nuanced linguistic usage, resulting in an inaccu-

rate prediction.

4.9.2 Explanations with BERT-Base Multilingual Cased Model

Figure 4.21: Explanation of the word *galatasaray* in the sentence *beşiktaş galatasaray arasındaki maçta iki oyuncu kırmızı kart gördü* by LIME.

In Figure 4.21, the term *galatasaray* within the sentence *beşiktaş galatasaray arasındaki maçta iki oyuncu kırmızı kart gördü* was subjected to interpretation. Notably, the prediction for the named entity TEAM associated with *galatasaray* was accurately foreseen with a high confidence level of 97.9%. The entirety of words present in the sentence uniformly contributes positively towards this prediction outcome.

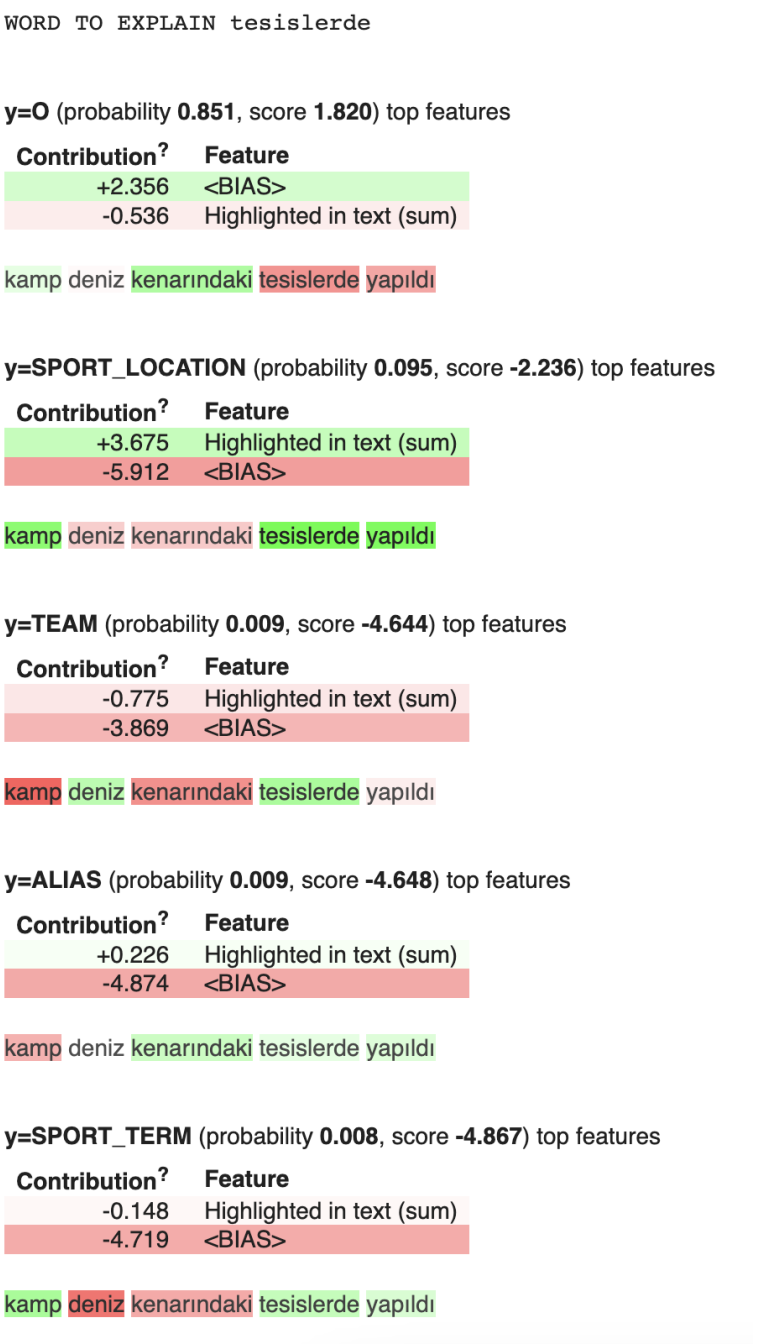


Figure 4.22: Explanation of the word *tesislerde* in the sentence *kamp deniz kenarındaki tesislerde yapıldı* by LIME.

In Figure 4.22, the explanation of the term *tesislerde* found within the sentence *kamp deniz kenarındaki tesislerde yapıldı* was given through the LIME methodology. Regrettably, the model's anticipation yielded an *O* label, accompanied by a probability estimation of 85.1%. However, it should be duly noted that the accurate pre-

diction should have been SPORT_LOCATION. Pertinently, for the secondary target SPORT_LOCATION, the terms *deniz* and *kenarındaki* manifested a counterproductive influence on the prediction. Despite their intent to convey a locational attribute, these terms paradoxically exhibited a negative contribution towards the classification. This occurrence highlights the nuanced challenges inherent in deciphering contextually intricate sport-related language constructs.

4.9.3 Explanations with BERT-Base Turkish Uncased Model

Figure 4.23: Explanation of the word *federasyonu* in the sentence *türkiye basketball federasyonu kurul toplantısından sonra anadolu efes fenerbahçe maçı başlayacak* by LIME.

Figure 4.23 presents the explanation of the term *federasyonu* within the sentence *türkiye basketball federasyonu kurul toplantısından sonra anadolu efes fenerbahçe maçı başlayacak* via the LIME methodology. The model's prediction, SPORT_ORGANIZATION, was accompanied by a high probability score of 98.1%. The predictive

Figure 4.24: Explanation of the word *kazandı* in the sentence *kartallar maçı hakemin oyunculara yanlış kartları göstermesi sonucunda kazandı* by LIME.

process entailed nuanced contributions from individual words within the sentence. For instance, *basketball, toplantısından, sonra, anadolu, efes*, and *maçı* were positively instrumental in shaping the outcome. Conversely, certain terms such as *kurul, fenerbahçe*, and *başlayacak* negatively influenced the prediction.

Illustrated in Figure 4.24, the term *kazandı* was explained through the application of LIME. The model's prediction, O, was subsequently deemed erroneous upon closer scrutiny. In this context, the veracious classification should have been SPORT_TERM, the second target label. Notably, the model assigned a high probability of 95.7% to the first target, further compounding the incorrect classification. This outcome un-

derscores a discernible failure of the model's predictive accuracy within the given context.

4.9.4 Explanations with XLM-RoBERTa-Base Multilingual Model

Figure 4.25: Explanation of the word *hakemin* in the sentence *dört büyüklerin oyuncuları maç içerisindeyken hakemin kararlarına ititraz ediyor.* by LIME.

Figure 4.25 presents the explanation of the term *hakemin* using the LIME methodology. In this context, the model proficiently and accurately forecasted the label as ROLE, a classification decision supported by a probability of 88.5%. Deconstructing the rationale behind this prediction, words such as *oyuncuları, kararlarına, itiraz*, and *ediyor* were identified to hold a positive contribution to the predicted label, substantiating their relevance in characterizing the term as a role within the sports context. This collective positive impact is measured through a cumulative score of 5.225, consolidating the model's alignment with the intended classification. Notably, the secondary target label, represented as O, assumes significance within the context of the data set,

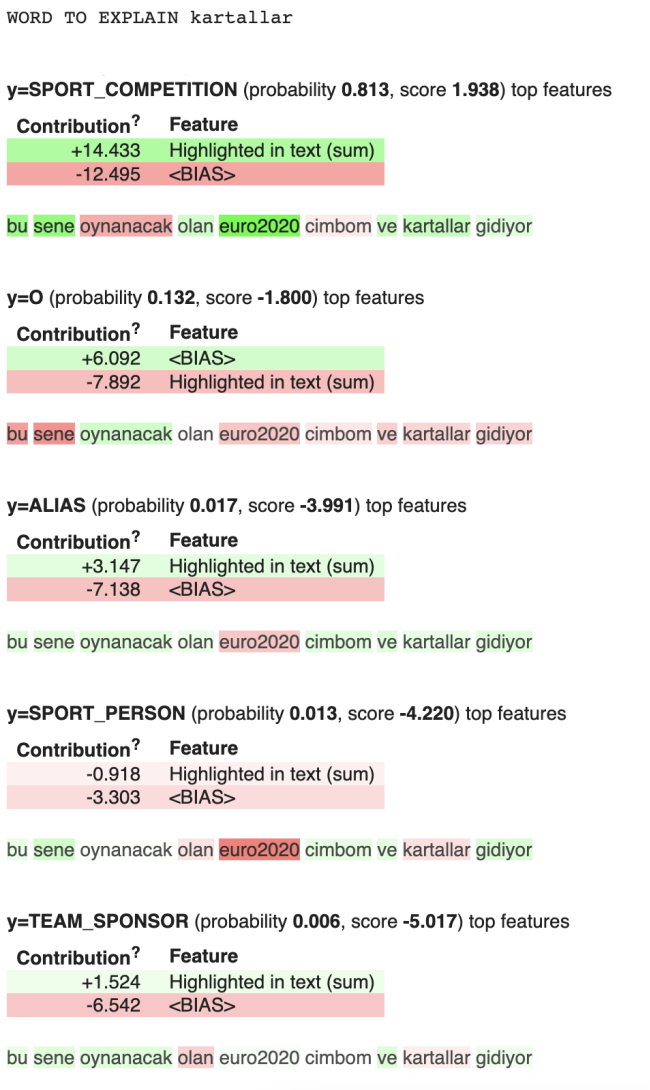


Figure 4.26: Explanation of the word *kartallar* in the sentence *bu sene oynanacak olan euro2020 cimbom ve kartallar gidiyor.* by LIME.

given the prevalence of entities categorized as O.

Figure 4.26 illustrates the interpretive analysis of the term *kartallar* through LIME. In the context of the figure, the model's primary prediction was the SPORT_COMPETITION label, accompanied by a probability of 81.3%. Additionally, the second and third target labels were O with a probability of 13.2% and ALIAS with a probability of 1.7%, respectively. Notably, the accurate prediction, which should have been ALIAS, was not achieved by the model. In the process of explaining this input, it is evident that the model attributed positive contributions to certain words such as *bu, sene, olan,*

euro2020, ve, and *gidiyor*, collectively influencing the SPORT_COMPETITION prediction. Conversely, for the third target, only the term *euro2020* demonstrated a negative impact. However, despite these considerations, the model's propensity to emphasize the contributions towards the first target over the third target led to an erroneous prediction of SPORT_COMPETITION instead of the intended ALIAS. This instance underscores the intricacies of prediction interpretation within the LIME framework and highlights the challenges in identifying nuanced contextual cues accurately.

4.9.5 Overview Analysis of Model Predictive Accuracy Using LIME

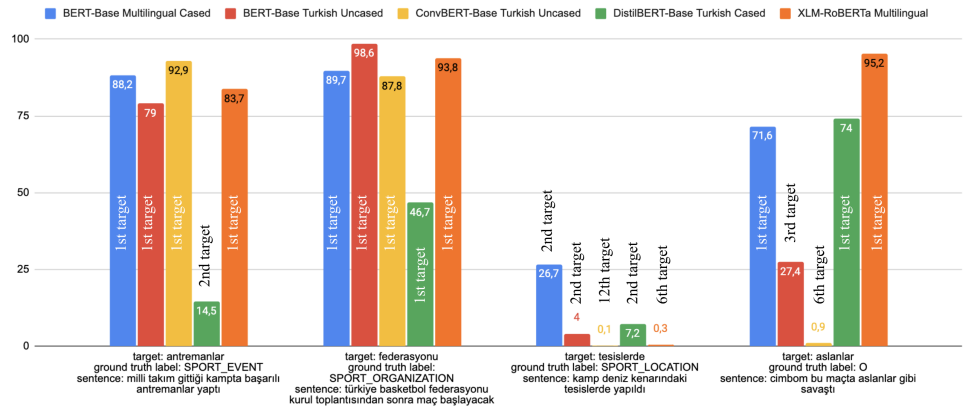


Figure 4.27: Prediction probability of ground truth annotation by LIME (In the x-axis, target word explained by LIME, ground truth label and sentence were given. The height of the bar indicates the probability assigned to correct label by the model according to LIME.)

Figure 4.27 illustrates the outcomes of a LIME analysis which evaluates the performance of various models on Turkish text data. Each bar represents a different model's ability to correctly predict a given label (ground truth), with the height of the bar indicating the probability assigned to the correct label by the model.

The target term "antremanlar" within the context "milli takım gittiği kampta başarılı antremanlar yaptı" has been accurately categorized as a "$SPORT\_EVENT$" entity by all models. Similarly, the term "federasyonu" in "türkiye basketbol federasyonu kurul toplantısından sonra maç başlayacak" has been correctly identified as a "$SPORT\_ORGANIZATION$". All models successfully identified the target as

the primary (first) entity, suggesting a higher degree of accuracy for these models in interpreting the context correctly.

Conversely, the models exhibit a decline in accuracy when predicting entities labeled as "$SPORT\_LOCATION$" and "O", diverging from the ground truth and assigning the correct label with diminished probability. This variance underscores the sensitivity of the LIME algorithm to the input example from which it generates perturbations. For instance, the first sentence contains multiple potential "$SPORT\_EVENT$" keywords, namely "kampta" and "antremanlar". The second sentence presents a trio of "$SPORT\_ORGANIZATION$" candidates: "türkiye", "basketbol", and "federasyonu". In contrast, the remaining sentences feature only a singular word corresponding to the ground truth label.

The implications of these outcomes are twofold. Firstly, they highlight the critical role of example selection in the LIME algorithm's performance, as the presence of multiple related terms within the input text can influence the interpretability outcome. Secondly, the results reflect the models' capabilities and limitations in discerning and prioritizing relevant entities within a given context.

4.10 Model Specific Explainability Results with SHAP

In this section, explanations of the models by implementing SHAP are represented. These experiments were conducted on Google Colab platform, employing the Python programming language version 3.10. Also, transformers package version 4.35.2 and shap version is 0.42.1 were used.

The experiments were conducted by using models that trained on the TurkishSportsNER-Parent-BIO data set, with the partition parameter k set to 4.

For the results shown in the figures, tokens with longer red bars have a stronger positive impact on the model's prediction score. These might be keywords that the model has learned are strongly associated with the prediction it is making. On the other hand, tokens with longer blue bars have a stronger negative contribution, meaning they are either detracting from the presence of a certain entity or decreasing the con-

fidence in a particular aspect of the model's prediction. This color coding helps in understanding how different parts of the data influence the prediction. Also, figures represent the explanation of the underscored named entity.

4.10.1 Explanations with BERT-Base Multilingual Cased Model

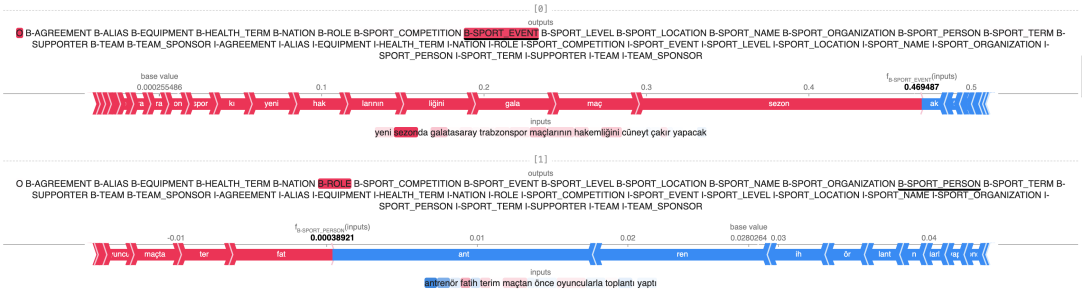


Figure 4.28: Explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* by SHAP.

Figure 4.28 represents the explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak*. It also interprets *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı*.

In the analysis of B-SPORT_EVENT for the first sentence, the term *sezon* is identified as having the highest positive contribution to the predicted label, followed by *maç*. This observation aligns with expectations, as these terms are annotated as B-SPORT_EVENT, indicating their strong relevance in the context of a sports event.

In the analysis of B-SPORT_PERSON for the second sentence, tokens *fat* and *ter*, components of the names *fatih* and *terim*, have the highest contributions to the predicted label. This underscores the model's capability to recognize name fragments as indicative of a sports person. Additionally, the tokens *ant* and *ren*' are observed to have the most negative impact on the prediction, perhaps due to their lesser association with the named entity.

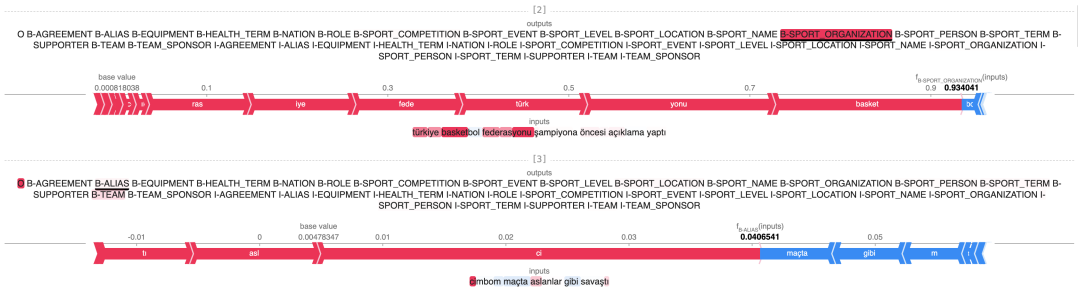


Figure 4.29: Explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* by SHAP.

Figure 4.29 represents the explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı*.

In the analysis of B-SPORT_ORGANIZATION for the first sentence, tokens *basket*, *yonu*, *türk*, *fede*, *iye*, *ras*are seen positively affecting the prediction, in that order. These tokens, forming part of *türkiye basketbol federasyonu*, underscore their alignment with the category of SPORT_ORGANIZATION.

In the analysis of B-ALIAS for the second sentence, tokens *ci* and *asl*, elements of *cimbom* and *aslanlar*, are the most influential for the predicted label. This reflects the model's effectiveness in tokenization parts of aliases for identification. Furthermore, the tokens *maçta* and *gibi*' are identified as having the most negative impact on the prediction.

4.10.2 Explanations with BERT-Base Turkish Uncased Model

Figure 4.30 illustrates the SHAP analysis of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak*. It also shows the interpretation for *B-SPORT_PERSON* in *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı*.

In the analysis of B-SPORT_EVENT for the first sentence, the term *sezon* was found to be the most positively influential term for the prediction, closely followed by *yeni*.

This suggests that these terms are key indicators of a sports event within the context. On the other hand, *yapacak* is noted as having the most significant negative impact, possibly because it's less directly related to identifying a sports event.

In the analysis of B-SPORT_PERSON for the second sentence, tokens *antren*, *toplantı* are identified as having the highest positive influence on the predicted label. In contrast, *fatih*, *ör*, *terim* are observed to negatively affect the prediction, which might be due to their tokenization.

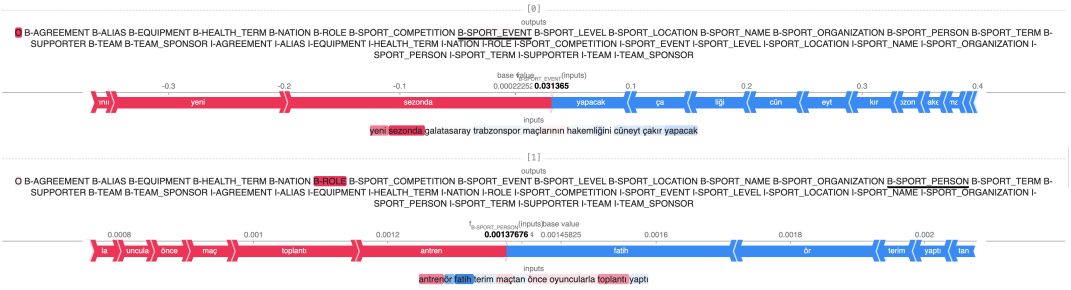


Figure 4.30: Explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* by SHAP.

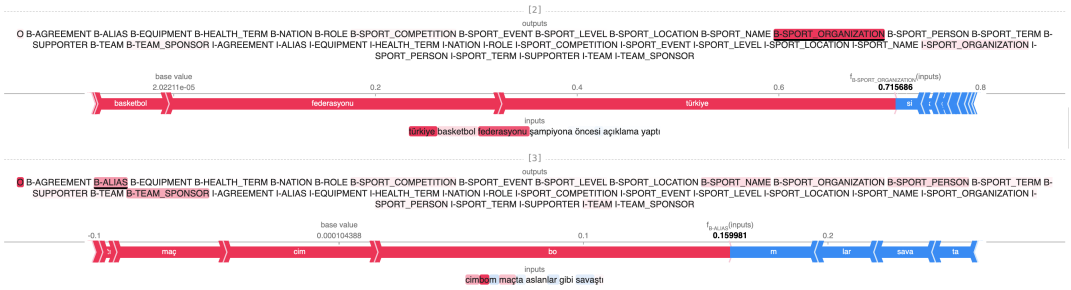


Figure 4.31: Explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* by SHAP.

Figure 4.31 provides the explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı*.

For B-SPORT_ORGANIZATION for the first sentence, tokens *türkiye*, *federasyonu*,

basketbol are found to have a positive effect on the prediction, in that order, aligning with their direct association with a sports organization.

Regarding B-ALIAS for the second sentence, tokens *bo*, *cim*, *maç* are identified as the most influential for the prediction. The tokens *bo* and *cim* forming part of *cim-bom* suggest the model's effectiveness in identifying relevant components of an alias within the context.

4.10.3 Explanations with ConvBERT-Base Turkish Uncased Model

Figure 4.32: Explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* by SHAP.

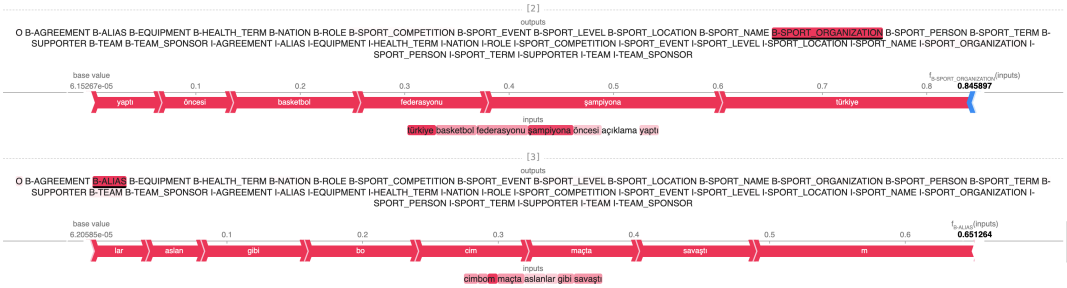
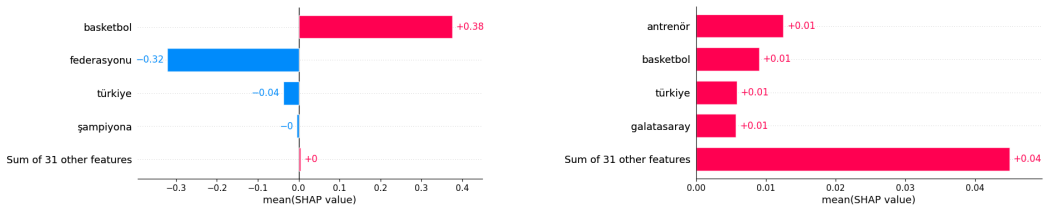


Figure 4.33: Explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* by SHAP.

Figure 4.32 illustrates the SHAP analysis for the *B-SPORT_EVENT* named entity for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt*



(a) B-SPORT_NAME

(b) B-ROLE

Figure 4.34: Contributions per Feature for Specific Named Entity Tag by SHAP (for 4 sentences)

çakır yapacak. The figure also includes an analysis of *B-SPORT_PERSON* in *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı*.

In the analysis of B-SPORT_EVENT for the first sentence, all the tokens in this sentence have a positive contribution to the prediction. This suggests a strong alignment of the entire sentence with the sports event context, highlighting the model's effective contextual understanding.

Regarding the B-SPORT_PERSON for the second sentence, tokens *antrenör*, *fatih*, and *terim* are identified as the most positively influential for the prediction. This aligns with their direct relevance to identifying a sports person. On the other hand, *maçtan* is noted to negatively affect the prediction.

The explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* were given in Figure 4.33.

Notably, every token in both sentences positively influences the prediction of their respective labels, indicating a comprehensive contextual coherence within the model's analysis.

In Figure 4.34, contributions per feature for specific named entity by SHAP are represented. Figure 4.34a illustrates the contributions per feature for B-SPORT_NAME. Notably, *basketbol* stands out as having the most significant positive effect on B-SPORT_NAME, which aligns with its frequent association as a B-SPORT_NAME in the training data set. Conversely, *federasyonu* exhibits the highest negative influence

on the label. This is understandable since *basketbol* within the term *türkiye basketbol federasyonu* is labeled as SPORT_ORGANIZATION (I-SPORT_ORGANIZATION), making *federasyonu* more likely to reduce the probability of correctly predicting B-SPORT_NAME. This indicates the model's nuanced understanding of contextual relevance and its impact on predictive accuracy.

In Figure 4.34b, the analysis of the contributions of each feature to the B-ROLE named entity. Notably, the token *antrenör* is observed to have the strongest positive influence on the predicted label. Additionally, tokens such as *basketbol*, *türkiye*, and *galatasaray* also contribute positively. This result is in line with expectations, as all these tokens have a direct association with sports-related terms. Their positive impact reflects the model's ability to identify and give weight to terms that are contextually relevant to sports roles, thereby enhancing the accuracy of its predictions for the B-ROLE entity.

4.10.4 Explanations with DistilBERT-Base Turkish Uncased Model

Figure 4.35: Explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* by SHAP.

The interpretations for *B-SPORT_EVENT* in *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* in *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* are provided in Figure 4.35.

In the analysis of B-SPORT_EVENT for the first sentence, *sezonda* and *maçlarını*are identified as the most positive contributors. This indicates their strong relevance in

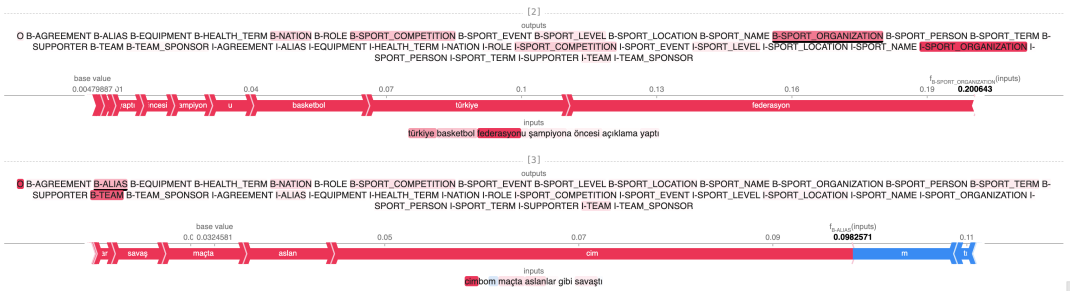


Figure 4.36: Explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* by SHAP.

contextually identifying a sports event.

Regarding the B-SPORT_PERSON for the second sentence, tokens *fat*, *ih*, and *terim* are identified as the most positively influential for the prediction, aligning with their contextual association with a sports person. On the other hand, *antrenör* and *maçtan* are noted to negatively impact the predicted label, which might be due to their less direct relation to SPORT_PERSON entity.

The explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* is presented in Figure 4.36.

Interestingly, in the case of both B-SPORT_ORGANIZATION in the first sentence and B-ALIAS in the second, nearly all tokens contribute positively to the prediction of their respective labels, with only one token in each case having a non-positive impact.

4.10.5 Explanations with XLM-RoBERTa-Base Multilingual Model

The SHAP analysis of the *B-SPORT_EVENT* named entity for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* in *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* is shown in Figure 4.37.

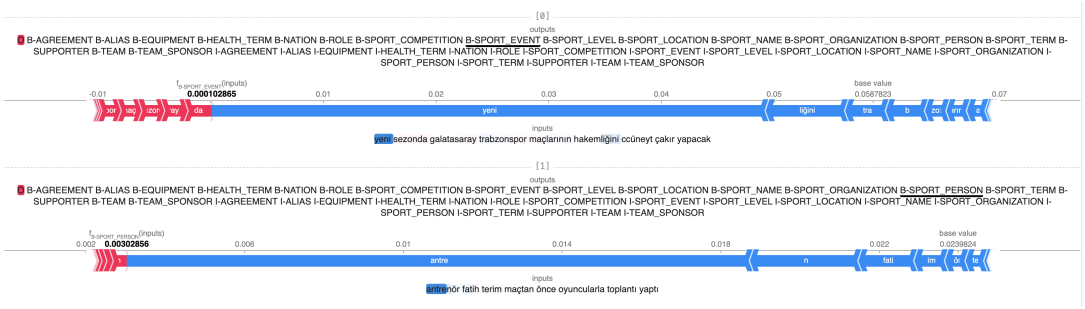


Figure 4.37: Explanation of the named entity *B-SPORT_EVENT* for the sentence *yeni sezonda galatasaray trabzonspor maçlarının hakemliğini cüneyt çakır yapacak* and *B-SPORT_PERSON* for the sentence *antrenör fatih terim maçtan önce oyuncularla toplantı yaptı* by SHAP.

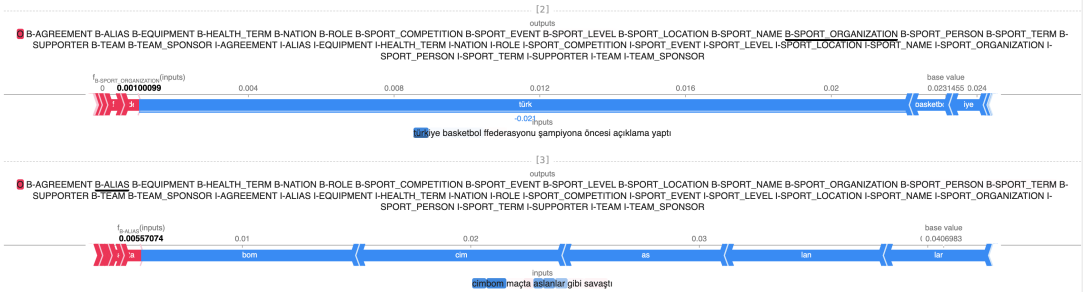


Figure 4.38: Explanation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı* by SHAP.

In the analysis, the token *yeni* is found to have the most significant negative impact on the B-SPORT_EVENT in the first sentence. This might be due to its general nature, not specifically contributing to the identification of a sports event.

For the B-SPORT_PERSON for the second sentence, tokens *antre*, *n*, and *fati* are identified as the most negatively affected tokens for the prediction.

Figure 4.38 provides the interpretation of the named entity *B-SPORT_ORGANIZATION* for the sentence *türkiye basketbol federasyonu şampiyona öncesi açıklama yaptı* and *B-ALIAS* for the sentence *cimbom maçta aslanlar gibi savaştı*.

In the case of SPORT_ORGANIZATION, *türk*, *basketbol*, *iye* are noted to have the most negative impact on the prediction.

Interestingly, for B-ALIAS in the second sentence, despite having a higher frequency in the training dataset as ALIAS, the tokens *bom*, *cim*, *as*, *lan*, and *lar* negatively influence the prediction. This could suggest a complexity in the model's ability to contextualize these tokens accurately within the given sentence structure.

4.11 Overview of the Results

In this section, an overview of the obtained results is presented. Key objectives are:

- ConvBERT-Base Turkish Uncased consistently demonstrated slightly higher f1-scores across all data set variations, indicating robust recognition performance.
- In contrast, DistilBERT-Base Turkish Cased exhibited relatively lower recognition efficiency compared to other models.
- BERT-Base Turkish Uncased ranked as the second-best performer since it was pre-trained on Turkish data sets.
- Multilingual models demonstrated relatively lower performance compared to Turkish models.
- XLM-RoBERTa-Base Multilingual model outperformed BERT-Base Multilingual Cased model since it has a larger pre-training data set. This emphasizes the significance of data volume in pre-training.
- A consistent trend emerged, wherein models excelled in recognition on the simpler TurkishSportsNER-Parent-NonBIO variant, while recognition performance lagged on the complex TurkishSportsNER-Child-BIO variant. This pattern highlights models' sensitivity to data set complexity and structural variations.
- LIME and SHAP explanations provide unique insights into the prediction mechanisms for NER.
- LIME offers localized interpretations by perturbing the input and observing the changes in model output. It provides insights into how individual features

(like words or tokens) influence the model's prediction for a specific instance. Conversely, SHAP, rooted in game theory, calculates the contribution of each feature to the prediction by considering all possible combinations of features. It provides a more global perspective, quantifying each feature's impact across various data points.

- In LIME, the explanations are often easier to interpret for a specific prediction. This is particularly useful for analyzing named entities in sentences. In contrast, SHAP can present more complexity in its interpretation, especially since it evaluates feature significance over the entire model rather than in isolated predictions.
- The method of text tokenization used by different models impacts their predictions.

CHAPTER 5

CONCLUSIONS

In summary, our study concentrated on Turkish Named Entity Recognition (NER) within the sports domain. Initially, sports-related textual data were collected from a newspaper website. Subsequent to this data collection, the named entities were determined based on the textual content. Following this, we introduced four distinct Turkish Named Entity Recognition (NER) data sets featuring hierarchical annotation levels (parent and child) and utilizing both BIO and non-BIO labeling schemes. These data sets were employed for comprehensive evaluations with a range of popular monolingual and multilingual transformer-based models. Through this rigorous experimentation, we not only facilitated a comparative analysis of NER formatting techniques but also investigated the influence of annotation intricacies on model performance.

The outcomes of our investigations revealed the ConvBERT-Base Turkish Uncased Model to exhibit notably high efficacy across our data sets. This heightened performance aligns with expectations, given the model's pre-training in Turkish. Remarkably, our analyses also unveiled the remarkable performance of the XLM-RoBERTa-Base Multilingual Model, despite its multilingual nature. This observation underscores the model's versatility and competence in the NER domain. Also, the DistilBERT-Base Turkish Cased model demonstrated relatively diminished performance.

The discernible patterns observed in the results underscore the intricate interplay between model efficiency and data set intricacies. Particularly, the models' adeptness in handling diverse data set variants emerged, with superior recognition on simpler structures and nuanced challenges on complex hierarchies. The introduced LIME and SHAP explanations offered a nuanced view of the models' prediction rationale,

bridging quantitative prediction scores and qualitative interpretability. LIME is particularly useful for understanding instance-level predictions and SHAP provides a comprehensive view of feature contributions at a global level.

In the context of future research endeavors, there exists the potential to augment the data sets by incorporating annotations characterized by a reduced instance count. If the data is balanced in terms of rare entities, it can be more suitable for improving detecting accuracy. Furthermore, an innovative real-time annotator could be devised, tailored specifically to the realm of sports. This novel tool would enable the dynamic examination of athletes' performance in real-time scenarios, contributing to a more nuanced understanding of their endeavors. Also, some other newly developed models such as GPT and techniques can be analyzed in these data sets. The results show that there is a need for explanation ground truth annotation to evaluate the success of the explanation.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018.

[2] Z. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, "Convbert: Improving bert with span-based dynamic convolution," 2021.

[3] "LIME - Local Interpretable Model-Agnostic Explanations)." `https://homes.cs.washington.edu/~marcotcr/blog/lime/`. Accessed 2023-07-30.

[4] R. Grishman and B. Sundheim, "Message Understanding Conference- 6: A brief history," in *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 1996.

[5] E. F. Tjong Kim Sang, "Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition," in *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*, 2002.

[6] S. Cucerzan and D. Yarowsky, "Language independent named entity recognition combining morphological and contextual evidence," in *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1999.

[7] G. Tur, D. Hakkani-Tur, and K. Oflazer, "A statistical information extraction system for turkish," *Natural Language Engineering*, vol. 9, pp. 181 – 210, 06 2003.

[8] D. Küçük and A. Yazıcı, "Named entity recognition experiments on turkish texts," in *Flexible Query Answering Systems*, pp. 524–535, 2009.

[9] D. Kucuk and A. Yazici, "A hybrid named entity recognizer for turkish," *Expert Syst. Appl.*, vol. 39, pp. 2733–2742, 02 2012.

[10] S. Tatar and I. Cicekli, "Automatic rule learning exploiting morphological features for named entity recognition in turkish," *J. Information Science*, vol. 37, pp. 137–151, 04 2011.

[11] R. Yeniterzi, "Exploiting morphology in Turkish named entity recognition system," in *Proceedings of the ACL 2011 Student Session*, (Portland, OR, USA), pp. 105–110, Association for Computational Linguistics, June 2011.

[12] S. Ozkaya and B. Diri, "Named entity recognition by conditional random fields from turkish informal texts," *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)*, pp. 662–665, 2011.

[13] G. A. Şeker and G. Eryiğit, "Initial explorations on using CRFs for Turkish named entity recognition," in *Proceedings of COLING 2012*, (Mumbai, India), pp. 2459–2474, The COLING 2012 Organizing Committee, Dec. 2012.

[14] G. A. Seker and G. Eryigit, "Extending a CRF-based named entity recognition model for turkish well formed text and user generated content," *Semantic Web*, vol. 8, pp. 625–642, 2017.

[15] A. Gunes and A. C. Tantug, "Turkish named entity recognition with deep learning," *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, 2018.

[16] R. F. Cekinel, M. Agriman, P. Karagoz, and B. Yilmaz, "Named entity recognition with conditional random fields on turkish news dataset: Revisiting the features," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, 2019.

[17] Y. P. Kilic, D. Dinc, and P. Karagoz, "Named entity recognition on morphologically rich language: Exploring the performance of bert with varying training levels," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4613–4619, 2020.

[18] E. Kağan Akkaya and B. Can, "Transfer learning for turkish named entity recognition on noisy text," *Natural Language Engineering*, vol. 27, no. 1, p. 35–64, 2021.

[19] B. Çarık and R. Yeniterzi, "A Twitter corpus for named entity recognition in Turkish," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, (Marseille, France), pp. 4546–4551, European Language Resources Association, June 2022.

[20] "Turkish BERT model." `https://huggingface.co/dbmdz/bert-base-turkish-uncased`. Accessed 2023-07-30.

[21] C. Çetindağ, B. Yazıcıoğlu, and A. Koç, "Named-entity recognition in turkish legal texts," *Natural Language Engineering*, vol. 29, no. 3, p. 615–642, 2023.

[22] T. Yao, W. Ding, and G. Erbach, "CHINERS: A Chinese named entity recognition system for the sports domain," in *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, (Sapporo, Japan), pp. 55–62, Association for Computational Linguistics, July 2003.

[23] D. Küçük, "Joint named entity recognition and stance detection in tweets," 2017.

[24] D. Küçük, "Stance detection in turkish tweets," 2017.

[25] D. Küçük and R. Steinberger, "Experiments to improve named entity recognition on Turkish tweets," in *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, (Gothenburg, Sweden), pp. 71–78, Association for Computational Linguistics, Apr. 2014.

[26] D. Küçük and F. Can, "A tweet dataset annotated for named entity recognition and stance detection," 2019.

[27] X. Seti, A. Wumaier, T. Yibulayin, D. Paerhati, L. Wang, and A. Saimaiti, "Named-entity recognition in sports field based on a character-level graph convolutional network," *Information*, vol. 11, p. 30, 01 2020.

[28] W. Wijesinghe and M. Tissera, "Sinhala named entity recognition model: Domain-specific classes in sports," in *2022 4th International Conference on Advancements in Computing (ICAC)*, pp. 138–143, 2022.

[29] J. Yang, J. Luo, W. Guo, D. Niu, and Y. Xu, "Exploiting hierarchically structured categories in fine-grained Chinese named entity recognition," in *Findings*

of the Association for Computational Linguistics: ACL 2023, (Toronto, Canada), pp. 3407–3421, Association for Computational Linguistics, July 2023.

[30] Z. Liu, Y. Leng, M. Wang, and C. Lin, "Named entity recognition and named entity on esports contents," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, pp. 189–192, 2020.

[31] E. Kokalj, B. Škrlj, N. Lavrač, S. Pollak, and M. Robnik-Šikonja, "BERT meets shapley: Extending SHAP explanations to transformer-based classifiers," in *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*, (Online), pp. 16–21, Association for Computational Linguistics, Apr. 2021.

[32] M. Z. Hossain, A. H. Z. So, S. Silwal, H. A. Gonzalez Gongora, A. M. Samin, J. A. Junaed, A. Mazumder, S. Saha, and S. Tahsin Soha, "garNER at SemEval-2023: Simplified knowledge augmentation for multilingual complex named entity recognition," in *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, (Toronto, Canada), pp. 823–835, Association for Computational Linguistics, July 2023.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[34] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," *CoRR*, vol. abs/1609.08144, 2016.

[35] "BERT multilingual base model (cased)." `https://huggingface.co/bert-base-multilingual-cased`. Accessed 2023-07-30.

[36] "bert." `https://github.com/google-research/bert`. Accessed 2023-07-30.

[37] "Turkish ConvBERT model." https://huggingface.co/dbmdz/convbert-base-turkish-mc4-uncased. Accessed 2023-07-30.

[38] "The C4 Multilingual Dataset." https://github.com/allenai/allennlp/discussions/5265. Accessed 2023-07-30.

[39] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," 2020.

[40] "Distilled Turkish BERT model." https://huggingface.co/dbmdz/distilbert-base-turkish-cased. Accessed 2023-07-30.

[41] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," 2020.

[42] "Unsupervised Cross-lingual Representation Learning at Scale (XLM-RoBERTa)." https://github.com/facebookresearch/fairseq/tree/main/examples/xlmr. Accessed 2023-07-30.

[43] G. Lample and A. Conneau, "Cross-lingual language model pretraining," 2019.

[44] "XLM-RoBERTa (base-sized model)." https://huggingface.co/xlm-roberta-base. Accessed 2023-07-30.

[45] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," 2016.

[46] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[47] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018.

[48] "Open Super-large Crawled Aggregated coRpus." https://oscar-project.org/. Accessed 2023-07-30.

[49] "OPUS." https://opus.nlpl.eu/. Accessed 2023-07-30.

[50] "BERT base model (uncased)." https://huggingface.co/bert-base-uncased. Accessed 2023-07-30.

APPENDIX A

DETAILED RESULTS OF 5-FOLD CROSS-VALIDATION EXPERIMENTS

Overall results of 5-fold cross-validation experiments were given in Chapter 4 and detailed results for each fold and each model per named entity are shown in the tables below.

Table A.1: Precision Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.64	98.66	98.63	98.67	98.60	98.64
AGREEMENT	73.05	92.00	86.80	87.90	92.58	86.47
ALIAS	85.45	92.83	84.62	94.03	85.60	88.50
EQUIPMENT	83.11	94.63	89.23	90.64	91.88	89.90
HEALTH_TERM	77.46	86.03	86.54	78.46	86.36	82.97
NATION	89.86	84.99	87.68	94.25	95.88	90.53
ROLE	90.41	93.09	92.08	92.87	93.74	92.44
SPORT_COMPETITION	84.84	91.82	89.78	86.61	87.61	88.13
SPORT_EVENT	82.62	92.15	93.31	93.30	93.36	90.95
SPORT_LEVEL	84.67	92.78	92.09	91.57	91.87	90.60
SPORT_LOCATION	74.53	85.03	91.90	90.72	90.53	86.54
SPORT_NAME	91.87	95.15	91.18	96.72	94.22	93.83
SPORT_ORGANIZATION	88.70	83.58	90.24	91.40	86.63	88.11
SPORT_PERSON	88.20	91.00	91.55	89.89	86.73	89.47
SPORT_TERM	64.88	83.09	87.76	90.91	90.61	83.45
SUPPORTER	85.68	96.57	91.10	94.04	93.94	92.27
TEAM	89.50	91.60	91.96	95.56	92.63	92.25
TEAM_SPONSOR	52.70	97.06	60.00	90.91	95.65	79.26

Table A.2: Recall Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.84	97.54	97.91	98.03	97.92	97.45
AGREEMENT	97.14	93.83	94.87	94.32	93.09	94.65
ALIAS	86.48	91.35	85.80	91.02	93.86	89.70
EQUIPMENT	88.73	85.45	93.16	94.92	94.52	91.36
HEALTH_TERM	68.37	84.17	92.47	90.27	89.62	84.98

Table A.2 Continued

NATION	92.39	92.27	90.88	88.10	79.71	88.67
ROLE	93.61	96.66	97.50	98.07	96.45	96.46
SPORT_COMPETITION	89.09	91.90	91.91	92.65	95.86	92.28
SPORT_EVENT	94.90	95.79	93.49	96.07	95.94	95.24
SPORT_LEVEL	90.82	95.42	93.72	95.32	93.49	93.75
SPORT_LOCATION	93.09	96.34	87.98	93.27	92.08	92.55
SPORT_NAME	94.11	96.01	93.48	93.34	95.51	94.49
SPORT_ORGANIZATION	86.53	91.06	93.27	92.82	92.43	91.22
SPORT_PERSON	94.28	94.07	95.56	95.49	92.16	94.31
SPORT_TERM	87.72	88.83	86.67	91.34	88.79	88.67
SUPPORTER	84.78	91.04	95.17	87.65	92.81	90.29
TEAM	93.30	95.18	95.10	91.21	92.70	93.50
TEAM_SPONSOR	76.47	70.21	54.55	43.48	40.00	56.94

Table A.3: Precision Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.61	98.68	98.62	98.67	98.63	98.64
B-AGREEMENT	72.87	93.44	85.89	89.79	93.08	87.02
B-ALIAS	79.83	91.88	83.23	92.33	79.38	85.33
B-EQUIPMENT	80.70	95.21	90.98	91.58	92.33	90.16
B-HEALTH_TERM	74.16	85.83	89.47	83.20	86.67	83.86
B-NATION	88.00	85.05	86.92	94.49	96.13	90.12
B-ROLE	88.41	93.28	92.51	92.78	93.96	92.19
B-SPORT_COMPETITION	82.33	88.60	87.04	84.60	86.30	85.77
B-SPORT_EVENT	82.48	92.42	94.38	93.89	94.56	91.55
B-SPORT_LEVEL	79.87	90.00	87.89	87.75	89.84	87.07
B-SPORT_LOCATION	73.28	86.46	89.21	91.68	88.91	85.91
B-SPORT_NAME	91.37	94.33	93.03	97.03	94.74	94.10

Table A.3 Continued

B-SPORT_ORGANIZATION	82.42	88.68	91.86	91.63	91.47	89.21
B-SPORT_PERSON	87.37	90.51	90.27	89.33	86.14	88.72
B-SPORT_TERM	64.78	82.00	88.17	91.85	89.98	83.36
B-SUPPORTER	83.16	95.28	90.29	94.30	93.66	91.33
B-TEAM	88.14	89.80	93.01	95.85	93.80	92.12
B-TEAM_SPONSOR	87.50	94.44	60.00	100.00	100.00	88.39
I-AGREEMENT	84.62	80.95	100.00	86.21	83.33	87.02
I-ALIAS	93.72	95.09	84.34	96.91	93.38	92.69
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	81.76	79.81	80.37	81.70	86.27	81.98
I-SPORT_COMPETITION	83.23	89.34	87.30	87.52	89.80	87.44
I-SPORT_EVENT	57.75	70.98	74.29	74.19	69.54	69.35
I-SPORT_LEVEL	82.94	88.16	92.70	88.57	86.67	87.81
I-SPORT_LOCATION	62.77	68.25	78.70	83.54	73.47	73.35
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	78.28	74.74	80.46	79.69	69.44	76.52
I-SPORT_PERSON	85.69	85.04	84.59	84.39	79.30	83.80
I-SPORT_TERM	53.38	83.33	76.85	78.60	82.50	74.93
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	78.36	91.94	87.31	81.20	78.92	83.55
I-TEAM_SPONSOR	100.00	0.00	0.00	0.00	0.00	20.00

Table A.4: Recall Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.88	97.64	98.08	98.20	98.02	97.57
B-AGREEMENT	97.30	92.33	95.56	96.35	95.65	95.44

Table A.4 Continued

B-ALIAS	84.82	87.65	82.19	88.22	92.58	87.09
B-EQUIPMENT	88.79	88.54	91.82	95.08	94.48	91.74
B-HEALTH_TERM	77.65	82.58	93.15	92.04	87.50	86.58
B-NATION	94.02	92.70	92.00	89.90	80.35	89.79
B-ROLE	92.38	95.63	95.44	96.62	96.34	95.28
B-SPORT_COMPETITION	84.33	90.80	84.26	89.68	92.86	88.39
B-SPORT_EVENT	95.19	94.35	93.11	94.99	94.30	94.39
B-SPORT_LEVEL	87.72	91.04	92.31	90.46	89.56	90.22
B-SPORT_LOCATION	93.76	93.40	87.52	92.83	90.40	91.58
B-SPORT_NAME	95.28	97.44	94.57	94.88	95.48	95.53
B-SPORT_ORGANIZATION	86.81	89.15	92.35	91.43	91.67	90.28
B-SPORT_PERSON	92.56	91.91	93.09	92.88	87.81	91.65
B-SPORT_TERM	87.57	90.26	86.78	89.78	90.24	88.93
B-SUPPORTER	84.51	91.67	94.72	87.27	93.09	90.25
B-TEAM	94.66	95.97	94.94	91.19	92.65	93.88
B-TEAM_SPONSOR	45.16	69.86	37.50	64.71	46.15	52.68
I-AGREEMENT	73.33	56.67	63.16	60.98	57.14	62.26
I-ALIAS	86.72	96.44	92.78	91.27	92.21	91.89
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.91	91.51	92.94	95.56	93.49	91.68
I-SPORT_COMPETITION	92.58	92.83	94.13	92.80	94.95	93.46
I-SPORT_EVENT	56.25	82.04	64.20	82.63	85.21	74.07
I-SPORT_LEVEL	87.82	93.38	89.75	94.26	92.86	91.61
I-SPORT_LOCATION	69.41	90.53	60.28	85.16	87.80	78.64
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	72.08	80.11	89.74	87.93	89.29	83.83
I-SPORT_PERSON	88.46	90.35	91.86	91.43	90.65	90.55
I-SPORT_TERM	73.83	79.89	82.82	87.90	79.20	80.73

Table A.4 Continued

I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	71.57	85.84	81.07	84.32	81.52	80.86
I-TEAM_SPONSOR	5.00	0.00	0.00	0.00	0.00	1.00

Table A.5: Precision Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.53	98.57	98.65	98.61	98.58	98.59
AGREEMENT	72.04	93.11	87.50	89.34	92.85	86.97
BASKETBALL_TERM	0.00	0.00	100.00	0.00	58.33	31.67
COACH	84.23	89.60	76.47	88.62	85.47	84.88
EQUIPMENT	83.38	94.26	89.88	91.58	93.48	90.52
FOOTBALL_TERM	71.75	81.71	87.35	89.30	90.23	84.07
HEALTH_TERM	80.00	88.10	87.26	79.39	81.03	83.16
HORSE_RACING_TERM	26.76	0.00	0.00	0.00	0.00	5.35
MANAGER	85.13	73.67	71.02	78.48	73.82	76.43
NATION	89.13	83.89	84.40	95.10	95.56	89.62
PERSON_ALIAS	26.35	77.87	75.22	81.88	78.24	67.91
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	74.38	76.85	79.71	71.97	66.94	73.97
ROLE	89.85	92.93	91.78	92.86	93.50	92.19
SCORE_TERM	53.89	77.87	84.60	91.45	86.66	78.89
SPORT_COMPETITION	84.03	91.11	89.75	86.64	89.63	88.23
SPORT_EVENT	82.65	92.45	94.30	93.36	94.43	91.44
SPORT_LEVEL	84.63	92.72	92.62	91.13	92.52	90.72
SPORT_LOCATION	73.70	85.34	91.67	91.80	90.11	86.53
SPORT_NAME	91.96	94.00	90.88	97.06	94.58	93.70
SPORT_ORGANIZATION	82.11	82.48	89.98	91.88	87.67	86.82
SPORT_PLAYER	82.80	89.82	88.74	88.57	84.40	86.87

Table A.5 Continued

SUPPORTER	85.79	96.10	92.36	94.28	94.77	92.66
TEAM	88.85	91.91	92.16	95.64	93.23	92.36
TEAM_ALIAS	89.45	92.80	83.08	93.32	86.80	89.09
TEAM_SPONSOR	81.25	100.00	46.15	71.43	71.05	73.98
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.6: Recall Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.11	97.63	97.95	98.21	98.01	97.58
AGREEMENT	96.36	91.95	94.23	93.83	93.23	93.92
BASKETBALL_TERM	0.00	0.00	23.08	0.00	17.07	8.03
COACH	85.08	89.85	83.38	86.35	80.40	85.01
EQUIPMENT	86.20	84.55	93.16	94.74	95.10	90.75
FOOTBALL_TERM	89.24	91.66	92.49	93.90	92.01	91.86
HEALTH_TERM	65.31	79.86	93.84	92.04	88.68	83.94
HORSE_RACING_TERM	50.00	0.00	0.00	0.00	0.00	10.00
MANAGER	79.66	82.94	87.57	87.50	86.73	84.88
NATION	92.39	90.61	92.10	89.44	78.97	88.70
PERSON_ALIAS	51.32	64.63	60.71	68.90	87.11	66.53
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	88.56	87.37	71.74	77.39	80.60	81.13
ROLE	93.09	96.49	97.15	97.93	96.57	96.25
SCORE_TERM	83.13	90.37	81.87	83.56	88.81	85.55
SPORT_COMPETITION	89.64	94.72	91.61	92.93	95.40	92.86
SPORT_EVENT	94.69	95.84	93.56	95.56	95.04	94.94
SPORT_LEVEL	91.59	94.63	92.91	94.59	93.66	93.48
SPORT_LOCATION	92.73	95.38	89.53	94.25	94.99	93.37

Table A.6 Continued

SPORT_NAME	93.68	96.56	93.04	93.51	95.51	94.46
SPORT_ORGANIZATION	88.03	92.67	94.02	90.02	92.61	91.47
SPORT_PLAYER	90.78	89.49	90.52	92.27	87.84	90.18
SUPPORTER	83.99	92.04	94.42	86.42	92.22	89.82
TEAM	93.05	95.39	95.18	91.38	92.49	93.50
TEAM_ALIAS	77.84	91.96	91.39	93.51	91.96	89.33
TEAM_SPONSOR	25.49	54.26	54.55	43.48	49.09	45.37
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.7: Precision Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.49	98.56	98.63	98.64	98.59	98.58
B-AGREEMENT	74.13	94.10	87.50	91.58	91.83	87.83
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	100.00	20.00
B-COACH	80.35	89.09	80.00	86.82	84.88	84.23
B-EQUIPMENT	80.66	94.56	89.92	91.28	93.43	89.97
B-FOOTBALL_TERM	71.28	79.38	87.33	90.62	90.13	83.75
B-HEALTH_TERM	75.29	86.40	85.99	84.00	84.26	83.19
B-HORSE_RACING_TERM	54.55	0.00	0.00	0.00	0.00	10.91
B-MANAGER	79.98	74.16	63.08	74.44	66.47	71.62
B-NATION	88.17	85.64	88.46	93.58	93.60	89.89
B-PERSON_ALIAS	31.53	76.04	79.59	90.24	71.35	69.75
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	75.59	64.62	76.78	67.36	62.40	69.35
B-ROLE	88.65	93.34	92.25	93.01	94.07	92.26
B-SCORE_TERM	52.34	77.35	82.39	88.14	84.71	76.99
B-SPORT_COMPETITION	80.88	85.58	84.55	85.92	86.48	84.68

Table A.7 Continued

B-SPORT_EVENT	83.14	91.99	93.70	93.84	94.94	91.52
B-SPORT_LEVEL	79.44	91.30	88.60	89.31	90.06	87.74
B-SPORT_LOCATION	73.58	85.65	88.28	91.23	88.98	85.55
B-SPORT_NAME	88.13	95.79	93.99	96.87	93.81	93.72
B-SPORT_ORGANIZATION	80.70	83.95	87.59	91.58	89.46	86.66
B-SPORT_PLAYER	79.13	87.45	88.16	85.98	83.93	84.93
B-SUPPORTER	83.83	95.23	91.51	93.02	95.68	91.85
B-TEAM	88.39	90.55	92.32	95.88	93.98	92.22
B-TEAM_ALIAS	88.75	91.53	85.07	92.57	83.28	88.24
B-TEAM_SPONSOR	90.91	100.00	0.00	100.00	100.00	78.18
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	75.00	90.00	100.00	85.71	83.33	86.81
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	80.18	82.45	69.17	86.56	87.96	81.27
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	57.79	86.17	78.20	76.44	82.94	76.31
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	87.85	75.69	73.53	73.58	69.76	76.09
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	13.51	73.08	77.78	100.00	87.50	70.37
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	73.10	66.10	81.97	72.88	61.48	71.11
I-ROLE	83.16	77.68	79.33	80.44	86.08	81.34
I-SCORE_TERM	51.16	75.00	76.47	78.95	83.33	72.98
I-SPORT_COMPETITION	83.44	88.77	85.79	87.07	89.52	86.92
I-SPORT_EVENT	60.98	75.43	74.10	70.10	66.48	69.42
I-SPORT_LEVEL	84.68	88.33	92.06	87.46	88.51	88.21
I-SPORT_LOCATION	57.58	71.07	83.84	80.63	75.44	73.71

Table A.7 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	75.29	74.87	79.53	79.79	71.03	76.10
I-SPORT_PLAYER	75.30	78.16	74.47	79.18	74.55	76.33
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	79.74	91.42	85.28	79.45	78.86	82.95
I-TEAM_ALIAS	92.02	91.30	82.54	87.45	89.47	88.56
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.8: Recall Values of BERT-Base Multilingual Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.04	97.69	98.01	98.20	98.03	97.59
B-AGREEMENT	97.57	91.36	95.56	96.52	96.23	95.45
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	2.86	0.57
B-COACH	81.09	87.50	82.21	84.47	77.98	82.65
B-EQUIPMENT	86.14	88.54	92.28	93.44	95.06	91.09
B-FOOTBALL_TERM	90.90	92.56	92.91	92.67	91.99	92.21
B-HEALTH_TERM	77.06	81.82	92.47	92.92	87.50	86.35
B-HORSE_RACING_TERM	19.35	0.00	0.00	0.00	0.00	3.87
B-MANAGER	77.41	79.67	81.90	81.97	80.08	80.21
B-NATION	93.45	90.45	92.00	87.77	80.10	88.75
B-PERSON_ALIAS	57.38	64.04	61.90	67.89	88.81	68.00
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	82.99	80.77	64.29	73.03	69.64	74.14
B-ROLE	92.41	94.72	95.71	96.38	96.13	95.07
B-SCORE_TERM	82.81	90.43	81.31	84.96	89.91	85.88
B-SPORT_COMPETITION	84.99	86.40	82.87	88.18	92.86	87.06

Table A.8 Continued

B-SPORT_EVENT	94.82	95.07	94.30	94.80	94.99	94.79
B-SPORT_LEVEL	89.47	91.04	92.01	90.46	88.92	90.38
B-SPORT_LOCATION	92.26	94.34	89.26	91.40	88.38	91.13
B-SPORT_NAME	95.28	95.79	94.42	95.05	95.28	95.16
B-SPORT_ORGANIZATION	88.09	89.95	91.29	90.79	91.23	90.27
B-SPORT_PLAYER	88.96	87.60	89.02	89.79	83.00	87.67
B-SUPPORTER	84.51	90.66	93.58	86.96	93.09	89.76
B-TEAM	94.23	95.66	95.14	91.25	92.30	93.72
B-TEAM_ALIAS	77.45	90.30	88.14	91.21	94.04	88.23
B-TEAM_SPONSOR	32.26	61.64	0.00	58.82	46.15	39.78
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	80.00	60.00	63.16	58.54	57.14	63.77
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	83.61	88.10	83.84	83.13	79.50	83.63
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	76.34	85.99	88.21	91.64	85.71	85.58
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	79.04	84.50	87.50	86.28	89.07	85.28
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	33.33	57.58	50.00	41.82	82.35	53.02
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	85.03	90.70	72.12	77.25	84.27	81.87
I-ROLE	84.04	90.87	94.17	95.56	93.49	91.63
I-SCORE_TERM	64.71	66.67	44.07	38.46	37.50	50.28
I-SPORT_COMPETITION	91.69	91.37	94.52	93.00	94.04	92.92
I-SPORT_EVENT	52.08	79.04	63.58	81.44	85.21	72.27
I-SPORT_LEVEL	88.24	92.33	90.11	94.26	91.67	91.32
I-SPORT_LOCATION	67.06	90.53	58.87	83.23	78.66	75.67

Table A.8 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	74.72	82.32	87.18	86.21	91.96	84.48
I-SPORT_PLAYER	79.20	80.63	78.53	85.81	87.85	82.40
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	73.01	86.43	80.71	85.80	80.25	81.24
I-TEAM_ALIAS	76.55	93.15	93.98	95.00	94.44	90.62
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.9: Precision Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.83	99.17	98.95	98.99	98.98	98.98
AGREEMENT	75.30	93.82	88.55	91.49	91.93	88.22
ALIAS	87.22	92.50	89.21	94.92	90.44	90.86
EQUIPMENT	84.29	94.16	90.20	89.30	94.81	90.55
HEALTH_TERM	78.16	87.22	86.71	75.71	79.83	81.53
NATION	91.71	89.87	86.41	94.36	95.93	91.66
ROLE	90.63	93.48	93.16	93.55	92.83	92.73
SPORT_COMPETITION	85.04	92.86	91.20	87.28	89.17	89.11
SPORT_EVENT	83.35	92.91	92.55	94.58	93.75	91.43
SPORT_LEVEL	83.22	92.31	90.73	92.57	90.69	89.90
SPORT_LOCATION	76.05	85.69	93.54	91.55	90.89	87.54
SPORT_NAME	94.06	94.23	94.83	97.90	95.34	95.27
SPORT_ORGANIZATION	88.72	86.18	92.74	89.27	87.08	88.80
SPORT_PERSON	92.37	94.66	94.80	93.26	88.38	92.69
SPORT_TERM	59.95	84.97	87.83	91.18	91.39	83.06
SUPPORTER	85.64	94.42	92.09	92.58	92.04	91.35

Table A.9 Continued

TEAM	91.86	92.80	94.01	96.39	94.39	93.89
TEAM_SPONSOR	65.00	84.38	50.00	71.43	79.55	70.07

Table A.10: Recall Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.92	97.96	98.16	98.34	98.06	97.69
AGREEMENT	96.62	93.72	94.23	92.53	91.30	93.68
ALIAS	88.73	92.36	87.60	93.76	97.52	92.00
EQUIPMENT	90.70	87.88	95.14	96.91	94.81	93.09
HEALTH_TERM	69.39	83.45	93.84	93.81	89.62	86.02
NATION	96.62	95.58	96.66	96.35	98.04	96.65
ROLE	94.37	97.01	96.72	97.72	96.93	96.55
SPORT_COMPETITION	90.53	93.93	94.18	93.70	96.55	93.78
SPORT_EVENT	94.60	95.88	93.90	95.37	95.77	95.10
SPORT_LEVEL	91.01	96.68	94.52	92.84	92.61	93.53
SPORT_LOCATION	95.82	95.76	91.60	95.79	96.04	95.00
SPORT_NAME	95.55	97.64	95.11	93.01	95.90	95.44
SPORT_ORGANIZATION	85.58	94.81	95.51	93.45	94.89	92.85
SPORT_PERSON	95.22	97.56	96.96	97.05	95.25	96.41
SPORT_TERM	88.63	91.18	89.97	92.88	90.44	90.62
SUPPORTER	86.09	92.54	95.17	88.58	93.41	91.16
TEAM	95.13	97.33	97.17	94.15	93.20	95.40
TEAM_SPONSOR	76.47	86.17	18.18	86.96	63.64	66.28

Table A.11: Precision Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.75	99.08	98.97	98.98	99.03	98.96
B-AGREEMENT	75.26	94.29	87.34	92.83	93.31	88.61
B-ALIAS	81.18	90.71	86.27	92.76	87.14	87.61
B-EQUIPMENT	81.48	94.54	90.29	90.77	93.37	90.09
B-HEALTH_TERM	80.12	86.26	87.58	84.50	79.31	83.55
B-NATION	90.24	89.41	87.12	93.46	96.59	91.36
B-ROLE	89.63	93.69	92.79	94.00	93.14	92.65
B-SPORT_COMPETITION	82.62	87.86	88.47	87.30	86.32	86.51
B-SPORT_EVENT	82.87	92.75	92.12	94.65	95.02	91.48
B-SPORT_LEVEL	79.49	92.66	89.83	90.70	90.30	88.59
B-SPORT_LOCATION	74.91	85.74	92.68	90.94	91.71	87.20
B-SPORT_NAME	93.42	95.23	95.61	97.39	94.79	95.29
B-SPORT_ORGANIZATION	85.25	85.38	94.12	91.18	89.81	89.15
B-SPORT_PERSON	90.59	93.93	93.60	92.27	87.85	91.65
B-SPORT_TERM	62.23	83.20	88.18	92.05	91.53	83.44
B-SUPPORTER	83.51	96.04	91.58	92.56	92.96	91.33
B-TEAM	90.89	92.07	94.00	96.45	94.60	93.60
B-TEAM_SPONSOR	85.00	98.51	0.00	100.00	94.74	75.65
I-AGREEMENT	68.75	78.26	92.86	85.71	83.33	81.78
I-ALIAS	93.04	97.42	89.12	97.35	92.94	93.97
I-EQUIPMENT	100.00	0.00	0.00	0.00	0.00	20.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	83.45	83.52	82.21	85.55	83.67	83.68
I-SPORT_COMPETITION	82.76	89.10	88.06	83.80	88.31	86.40
I-SPORT_EVENT	57.92	71.43	67.05	71.81	70.86	67.81
I-SPORT_LEVEL	83.86	87.90	87.54	88.01	85.77	86.62
I-SPORT_LOCATION	65.49	75.42	80.88	83.63	80.11	77.11

Table A.11 Continued

I-SPORT_NAME	0.00	100.00	0.00	0.00	0.00	20.00
I-SPORT_ORGANIZATION	83.77	79.80	81.56	78.68	67.09	78.18
I-SPORT_PERSON	89.75	91.07	88.89	87.38	80.96	87.61
I-SPORT_TERM	53.58	85.38	77.82	79.00	81.34	75.43
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	83.08	88.32	86.46	82.49	81.73	84.42
I-TEAM_SPONSOR	56.25	75.00	0.00	0.00	100.00	46.25

Table A.12: Recall Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.12	97.90	98.12	98.35	98.05	97.71
B-AGREEMENT	96.22	94.60	94.20	94.61	93.18	94.56
B-ALIAS	86.01	89.83	82.50	90.58	94.38	88.66
B-EQUIPMENT	90.86	88.22	94.75	96.72	94.19	92.95
B-HEALTH_TERM	75.88	85.61	91.78	96.46	88.46	87.64
B-NATION	97.44	97.19	97.85	97.09	98.76	97.66
B-ROLE	92.97	95.75	94.91	96.62	95.66	95.18
B-SPORT_COMPETITION	84.99	87.36	88.65	90.24	93.09	88.86
B-SPORT_EVENT	95.10	95.43	93.98	94.88	94.80	94.84
B-SPORT_LEVEL	87.02	94.80	94.08	90.46	85.44	90.36
B-SPORT_LOCATION	93.76	95.05	90.05	93.55	93.10	93.10
B-SPORT_NAME	96.31	98.72	95.32	95.56	96.46	96.48
B-SPORT_ORGANIZATION	88.51	95.77	92.88	92.93	94.74	92.96
B-SPORT_PERSON	93.61	96.12	95.24	94.84	91.72	94.31
B-SPORT_TERM	88.96	90.54	88.70	90.97	88.96	89.63
B-SUPPORTER	85.33	91.92	94.34	88.82	95.20	91.12
B-TEAM	96.30	97.47	97.18	93.49	93.91	95.67
B-TEAM_SPONSOR	54.84	90.41	0.00	58.82	46.15	50.05

Table A.12 Continued

I-AGREEMENT	73.33	60.00	68.42	58.54	57.14	63.49
I-ALIAS	88.80	93.95	95.56	93.45	98.44	94.04
I-EQUIPMENT	25.00	0.00	0.00	0.00	0.00	5.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.91	94.69	93.56	95.82	96.51	93.10
I-SPORT_COMPETITION	91.69	93.16	96.67	93.58	97.02	94.42
I-SPORT_EVENT	55.21	83.83	72.84	80.84	87.32	76.01
I-SPORT_LEVEL	89.50	96.17	94.35	94.26	95.63	93.98
I-SPORT_LOCATION	87.06	93.68	78.01	92.26	90.85	88.37
I-SPORT_NAME	0.00	16.67	0.00	0.00	0.00	3.33
I-SPORT_ORGANIZATION	72.08	89.50	93.59	89.08	94.64	87.78
I-SPORT_PERSON	90.22	94.99	95.61	94.62	96.98	94.48
I-SPORT_TERM	75.51	83.46	85.23	92.02	87.20	84.68
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	78.07	91.45	88.93	92.01	86.08	87.31
I-TEAM_SPONSOR	45.00	14.29	0.00	0.00	6.25	13.11

Table A.13: Precision Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.73	99.07	98.92	98.94	98.98	98.93
AGREEMENT	74.90	93.70	90.12	91.47	92.63	88.56
BASKETBALL_TERM	60.00	80.00	56.76	70.83	53.66	64.25
COACH	88.51	92.29	78.27	89.29	88.86	87.44
EQUIPMENT	85.52	96.25	89.31	91.78	92.88	91.15
FOOTBALL_TERM	75.80	82.57	88.25	91.09	92.50	86.04
HEALTH_TERM	79.39	91.94	85.71	84.92	81.58	84.71
HORSE_RACING_TERM	8.55	0.00	100.00	0.00	0.00	21.71

Table A.13 Continued

MANAGER	86.91	85.09	77.68	81.23	68.00	79.78
NATION	91.49	91.27	88.12	94.21	96.41	92.30
PERSON_ALIAS	28.77	75.18	87.00	84.25	80.18	71.08
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	80.66	79.81	82.81	78.69	76.65	79.72
ROLE	90.63	93.51	93.41	93.52	92.87	92.79
SCORE_TERM	54.01	82.98	82.89	90.38	89.12	79.88
SPORT_COMPETITION	85.91	92.52	90.16	87.73	88.98	89.06
SPORT_EVENT	83.00	92.95	92.32	94.72	94.02	91.40
SPORT_LEVEL	84.28	93.05	89.83	92.00	91.19	90.07
SPORT_LOCATION	75.22	84.93	92.62	91.40	90.30	86.89
SPORT_NAME	93.34	93.28	91.94	97.58	94.77	94.18
SPORT_ORGANIZATION	88.56	83.99	92.03	91.30	85.56	88.29
SPORT_PLAYER	89.97	93.18	92.89	91.55	87.48	91.01
SUPPORTER	84.75	94.37	90.14	92.88	93.47	91.12
TEAM	91.36	92.55	93.57	96.27	94.74	93.70
TEAM_ALIAS	94.10	95.32	85.28	95.70	93.19	92.72
TEAM_SPONSOR	61.82	97.14	0.00	71.43	84.38	62.95
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.14: Recall Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.31	98.06	98.14	98.36	98.07	97.79
AGREEMENT	96.10	93.31	93.59	92.21	91.99	93.44
BASKETBALL_TERM	54.93	17.70	53.85	50.75	53.66	46.18
COACH	89.53	92.81	86.85	91.78	85.60	89.31
EQUIPMENT	89.86	85.45	92.71	97.28	93.95	91.85

Table A.14 Continued

FOOTBALL_TERM	89.94	93.55	94.23	95.13	91.27	92.83
HEALTH_TERM	66.84	82.01	90.41	94.69	87.74	84.34
HORSE_RACING_TERM	94.74	0.00	6.97	0.00	0.00	20.34
MANAGER	81.38	87.63	88.54	89.19	91.30	87.61
NATION	96.90	95.30	96.96	96.74	98.53	96.89
PERSON_ALIAS	55.26	72.11	62.14	75.00	89.69	70.84
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	92.96	87.37	80.65	83.48	86.57	86.21
ROLE	94.05	97.04	96.72	97.75	96.69	96.45
SCORE_TERM	85.57	90.98	84.09	86.44	89.36	87.29
SPORT_COMPETITION	88.98	92.61	94.08	94.27	96.55	93.30
SPORT_EVENT	94.56	95.88	93.97	95.56	96.29	95.25
SPORT_LEVEL	91.20	97.31	93.88	94.15	92.96	93.90
SPORT_LOCATION	94.91	95.57	90.83	95.37	94.59	94.25
SPORT_NAME	94.68	98.01	94.67	93.84	95.51	95.34
SPORT_ORGANIZATION	85.31	95.71	94.95	93.29	95.95	93.04
SPORT_PLAYER	93.04	94.43	92.41	93.84	89.94	92.73
SUPPORTER	86.09	91.79	95.17	88.58	94.31	91.19
TEAM	95.79	97.08	97.17	94.05	93.65	95.55
TEAM_ALIAS	79.64	93.05	93.33	94.73	98.08	91.77
TEAM_SPONSOR	66.67	72.34	0.00	86.96	49.09	55.01
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.15: Precision Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.74	99.06	98.97	98.95	98.99	98.94
B-AGREEMENT	75.05	93.95	89.03	92.19	91.50	88.34

Table A.15 Continued

B-BASKETBALL_TERM	80.65	72.73	23.08	33.33	58.62	53.68
B-COACH	85.80	91.65	81.01	89.22	88.15	87.17
B-EQUIPMENT	81.96	94.24	90.63	90.43	92.26	89.90
B-FOOTBALL_TERM	74.65	81.56	88.80	91.84	92.24	85.82
B-HEALTH_TERM	75.88	89.68	86.27	84.50	81.25	83.52
B-HORSE_RACING_TERM	6.68	0.00	0.00	0.00	0.00	1.34
B-MANAGER	85.50	83.76	77.55	82.56	66.45	79.16
B-NATION	90.21	89.15	87.60	93.96	95.65	91.32
B-PERSON_ALIAS	32.69	74.04	86.46	84.54	73.29	70.20
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	79.07	79.63	78.60	73.47	72.03	76.56
B-ROLE	88.67	94.32	92.03	93.58	92.85	92.29
B-SCORE_TERM	53.32	80.74	82.19	90.05	85.51	78.36
B-SPORT_COMPETITION	83.59	88.46	87.35	86.26	86.32	86.40
B-SPORT_EVENT	83.12	92.25	92.86	94.68	94.81	91.54
B-SPORT_LEVEL	78.23	93.39	88.58	89.24	90.24	87.94
B-SPORT_LOCATION	74.19	85.96	92.19	90.89	89.20	86.49
B-SPORT_NAME	92.59	94.57	95.15	97.22	94.22	94.75
B-SPORT_ORGANIZATION	86.43	87.38	94.13	92.09	89.33	89.87
B-SPORT_PLAYER	87.69	93.13	93.08	91.16	87.20	90.45
B-SUPPORTER	82.98	94.56	90.68	92.86	91.47	90.51
B-TEAM	91.67	91.49	93.90	96.62	95.31	93.80
B-TEAM_ALIAS	94.44	93.90	86.43	96.58	91.85	92.64
B-TEAM_SPONSOR	86.36	100.00	50.00	100.00	88.89	85.05
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	73.33	85.71	87.50	80.00	82.61	81.83
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	86.73	90.71	70.68	90.44	88.71	85.45
I-EQUIPMENT	100.00	0.00	0.00	0.00	0.00	20.00

Table A.15 Continued

I-FOOTBALL_TERM	67.90	81.25	78.58	76.77	79.17	76.73
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	88.41	82.31	78.35	77.99	66.33	78.68
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	15.63	75.00	77.78	93.48	90.38	70.45
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	75.56	77.08	80.66	80.72	77.67	78.34
I-ROLE	83.92	81.42	80.21	83.30	81.66	82.10
I-SCORE_TERM	38.89	80.00	67.19	74.36	76.67	67.42
I-SPORT_COMPETITION	83.40	89.51	87.06	84.15	88.40	86.50
I-SPORT_EVENT	63.79	74.03	67.47	73.08	67.93	69.26
I-SPORT_LEVEL	82.63	89.07	86.97	87.46	84.81	86.19
I-SPORT_LOCATION	63.64	72.13	78.99	80.68	78.77	74.84
I-SPORT_NAME	0.00	100.00	0.00	0.00	0.00	20.00
I-SPORT_ORGANIZATION	80.08	86.41	79.46	77.34	66.03	77.86
I-SPORT_PLAYER	80.80	85.71	82.15	84.89	79.06	82.52
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	82.82	89.16	84.80	81.14	83.17	84.22
I-TEAM_ALIAS	94.79	96.30	85.79	97.69	92.07	93.33
I-TEAM_SPONSOR	75.00	100.00	0.00	0.00	66.67	48.33
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.16: Recall Values of BERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.20	98.02	98.20	98.42	98.01	97.77
B-AGREEMENT	96.76	93.84	94.20	94.43	93.76	94.60

Table A.16 Continued

B-BASKETBALL_TERM	39.06	8.99	30.00	2.08	48.57	25.74
B-COACH	89.41	93.48	86.59	91.53	84.44	89.09
B-EQUIPMENT	91.15	88.54	93.98	96.36	93.60	92.73
B-FOOTBALL_TERM	91.40	92.67	93.33	92.80	88.93	91.83
B-HEALTH_TERM	75.88	85.61	90.41	96.46	87.50	87.17
B-HORSE_RACING_TERM	96.77	0.00	0.00	0.00	0.00	19.35
B-MANAGER	79.90	83.30	84.44	87.98	87.01	84.53
B-NATION	97.15	96.91	97.85	96.70	98.51	97.42
B-PERSON_ALIAS	55.74	67.54	65.87	75.23	82.52	69.38
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	87.63	82.69	75.79	80.90	75.89	80.58
B-ROLE	93.42	95.66	94.91	95.78	95.04	94.96
B-SCORE_TERM	84.66	90.27	82.98	87.12	88.01	86.61
B-SPORT_COMPETITION	85.43	88.12	88.05	89.49	93.09	88.84
B-SPORT_EVENT	95.38	95.34	93.70	94.68	93.98	94.62
B-SPORT_LEVEL	87.02	93.93	94.08	87.63	84.81	89.49
B-SPORT_LOCATION	93.33	93.87	87.68	93.01	90.40	91.66
B-SPORT_NAME	95.87	98.90	94.72	95.56	96.07	96.23
B-SPORT_ORGANIZATION	88.09	95.24	93.14	92.29	93.64	92.48
B-SPORT_PLAYER	92.08	93.31	92.41	91.90	87.24	91.39
B-SUPPORTER	86.14	92.17	95.47	88.82	93.39	91.20
B-TEAM	96.48	97.75	97.24	92.98	93.50	95.59
B-TEAM_ALIAS	80.36	92.64	88.66	93.04	97.02	90.35
B-TEAM_SPONSOR	61.29	87.67	12.50	70.59	61.54	58.72
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	73.33	60.00	73.68	58.54	54.29	63.97
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	86.94	91.52	86.87	91.56	90.38	89.45
I-EQUIPMENT	25.00	0.00	0.00	0.00	0.00	5.00

Table A.16 Continued

I-FOOTBALL_TERM	75.72	88.32	91.59	95.24	92.40	88.66
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	82.19	86.56	90.50	89.38	93.02	88.33
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	33.33	63.64	50.00	78.18	92.16	63.46
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	92.52	86.05	82.21	80.24	89.89	86.18
I-ROLE	83.33	94.90	93.25	95.04	96.28	92.56
I-SCORE_TERM	61.76	74.07	72.88	74.36	57.50	68.12
I-SPORT_COMPETITION	91.46	95.93	96.09	93.97	96.10	94.71
I-SPORT_EVENT	57.81	80.24	69.14	79.64	88.03	74.97
I-SPORT_LEVEL	89.92	96.52	94.35	94.26	95.24	94.05
I-SPORT_LOCATION	82.35	92.63	77.30	91.61	85.98	85.98
I-SPORT_NAME	0.00	16.67	0.00	0.00	0.00	3.33
I-SPORT_ORGANIZATION	75.85	87.85	94.23	90.23	91.96	88.02
I-SPORT_PLAYER	85.63	89.53	83.51	92.29	95.24	89.24
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	77.83	92.18	89.64	92.90	87.59	88.03
I-TEAM_ALIAS	80.53	94.35	94.58	95.91	98.89	92.85
I-TEAM_SPONSOR	30.00	33.33	0.00	0.00	12.50	15.17
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.17: Precision Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.94	99.31	98.97	99.03	99.13	99.08
AGREEMENT	72.66	92.57	89.54	90.13	91.51	87.28

Table A.17 Continued

ALIAS	86.99	92.40	86.89	96.38	90.76	90.68
EQUIPMENT	81.82	94.70	90.98	89.73	92.35	89.92
HEALTH_TERM	80.37	89.60	86.84	78.63	86.79	84.45
NATION	90.74	91.38	86.83	94.52	96.87	92.07
ROLE	91.09	94.12	93.77	94.36	93.94	93.46
SPORT_COMPETITION	85.13	92.99	92.12	89.51	89.23	89.80
SPORT_EVENT	81.45	92.08	93.31	94.05	93.09	90.80
SPORT_LEVEL	84.52	91.35	90.37	91.67	89.55	89.49
SPORT_LOCATION	75.29	84.01	92.82	91.60	87.38	86.22
SPORT_NAME	93.15	95.40	91.75	97.59	95.19	94.61
SPORT_ORGANIZATION	86.59	80.76	93.81	91.73	88.51	88.28
SPORT_PERSON	93.05	93.81	95.04	93.10	87.87	92.57
SPORT_TERM	58.08	83.47	87.26	90.96	90.07	81.97
SUPPORTER	83.04	93.55	88.05	90.82	88.67	88.83
TEAM	91.06	92.83	94.38	96.76	95.01	94.01
TEAM_SPONSOR	94.12	0.00	83.33	89.47	100.00	73.38

Table A.18: Recall Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.71	97.75	98.13	98.36	97.98	97.58
AGREEMENT	96.62	92.47	93.27	93.34	93.78	93.90
ALIAS	88.04	92.80	84.80	93.15	94.91	90.74
EQUIPMENT	91.27	86.67	91.95	96.73	93.95	92.11
HEALTH_TERM	66.84	80.58	90.41	91.15	86.79	83.15
NATION	96.62	96.69	98.18	95.97	98.29	97.15
ROLE	93.52	96.99	96.80	97.65	97.11	96.41
SPORT_COMPETITION	90.53	95.69	94.67	94.56	96.21	94.33
SPORT_EVENT	94.99	96.85	94.49	95.78	96.55	95.73

Table A.18 Continued

SPORT_LEVEL	90.82	98.42	93.72	94.88	95.07	94.58
SPORT_LOCATION	94.18	96.15	91.86	93.27	97.76	94.64
SPORT_NAME	95.69	97.64	95.56	94.18	96.68	95.95
SPORT_ORGANIZATION	86.12	94.63	93.46	93.45	96.30	92.79
SPORT_PERSON	96.30	98.34	97.14	97.40	95.52	96.94
SPORT_TERM	88.63	92.43	91.20	93.51	91.94	91.54
SUPPORTER	86.09	93.78	95.91	88.58	96.11	92.09
TEAM	96.72	98.41	96.80	94.17	93.88	96.00
TEAM_SPONSOR	31.37	0.00	45.45	73.91	38.18	37.78

Table A.19: Precision Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	99.01	99.35	99.00	99.08	99.14	99.12
B-AGREEMENT	72.14	90.15	86.75	88.85	88.80	85.34
B-ALIAS	80.56	86.73	84.09	93.77	85.95	86.22
B-EQUIPMENT	77.15	92.09	89.31	89.43	90.83	87.76
B-HEALTH_TERM	80.00	85.16	86.93	80.77	85.19	83.61
B-NATION	90.96	90.36	89.66	94.33	95.39	92.14
B-ROLE	89.19	93.41	93.69	94.33	93.54	92.83
B-SPORT_COMPETITION	80.58	86.64	87.96	85.51	84.02	84.94
B-SPORT_EVENT	80.97	92.17	93.29	94.70	94.01	91.03
B-SPORT_LEVEL	80.19	91.85	87.01	89.92	87.82	87.36
B-SPORT_LOCATION	74.96	85.03	90.66	89.84	87.54	85.61
B-SPORT_NAME	93.01	94.50	94.64	97.05	94.58	94.76
B-SPORT_ORGANIZATION	81.75	78.25	95.20	88.80	86.39	86.08
B-SPORT_PERSON	91.12	93.63	94.44	92.24	87.36	91.76
B-SPORT_TERM	61.18	82.19	87.85	91.21	90.48	82.58
B-SUPPORTER	80.87	91.81	88.57	91.61	88.45	88.26

Table A.19 Continued

B-TEAM	90.18	91.51	94.59	96.74	96.06	93.81
B-TEAM_SPONSOR	100.00	0.00	0.00	0.00	100.00	40.00
I-AGREEMENT	80.00	0.00	100.00	0.00	0.00	36.00
I-ALIAS	92.64	91.61	84.65	96.73	92.88	91.70
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.89	87.23	82.13	81.86	85.15	84.25
I-SPORT_COMPETITION	82.59	90.92	89.60	87.95	87.27	87.67
I-SPORT_EVENT	56.52	73.40	77.78	75.39	78.06	72.23
I-SPORT_LEVEL	85.08	86.42	86.17	87.85	85.66	86.24
I-SPORT_LOCATION	62.62	66.14	80.60	82.53	77.11	73.80
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	80.16	82.99	82.39	81.72	71.43	79.74
I-SPORT_PERSON	92.04	89.79	90.40	89.71	81.91	88.77
I-SPORT_TERM	50.26	84.17	79.46	79.22	82.84	75.19
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	86.95	89.59	87.21	87.29	85.37	87.28
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.20: Recall Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.88	97.55	98.15	98.27	97.90	97.55
B-AGREEMENT	97.30	93.84	93.86	95.65	95.50	95.23
B-ALIAS	86.31	88.62	80.94	90.58	93.48	87.99
B-EQUIPMENT	92.63	92.68	94.14	97.09	95.06	94.32
B-HEALTH_TERM	75.29	82.58	91.10	92.92	88.46	86.07
B-NATION	97.44	97.47	98.77	96.89	97.76	97.67

Table A.20 Continued

B-ROLE	93.63	96.76	95.53	97.33	96.58	95.97
B-SPORT_COMPETITION	86.09	91.95	90.24	89.68	94.47	90.49
B-SPORT_EVENT	95.56	95.79	95.88	95.70	95.95	95.78
B-SPORT_LEVEL	89.47	94.51	91.12	89.69	86.71	90.30
B-SPORT_LOCATION	93.33	92.45	88.94	91.94	92.26	91.78
B-SPORT_NAME	96.17	97.62	95.93	95.56	96.07	96.27
B-SPORT_ORGANIZATION	89.57	97.09	94.20	93.36	96.05	94.05
B-SPORT_PERSON	95.69	97.62	95.79	95.50	92.34	95.39
B-SPORT_TERM	89.13	91.98	89.16	92.27	91.23	90.75
B-SUPPORTER	86.14	93.43	93.58	88.20	94.29	91.13
B-TEAM	97.16	98.58	96.59	94.29	94.14	96.15
B-TEAM_SPONSOR	3.23	0.00	0.00	0.00	10.26	2.70
I-AGREEMENT	26.67	0.00	15.79	0.00	0.00	8.49
I-ALIAS	88.80	97.15	95.00	96.73	97.51	95.04
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.74	95.75	94.48	96.61	96.05	93.52
I-SPORT_COMPETITION	91.69	96.25	96.09	95.14	97.48	95.33
I-SPORT_EVENT	54.17	82.63	77.78	86.23	85.21	77.20
I-SPORT_LEVEL	88.66	97.56	94.70	95.27	97.22	94.68
I-SPORT_LOCATION	78.82	88.42	76.60	88.39	94.51	85.35
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	74.72	88.95	92.95	87.36	93.75	87.54
I-SPORT_PERSON	93.40	97.30	96.71	95.92	97.04	96.07
I-SPORT_TERM	72.90	84.96	86.76	91.76	88.80	85.04
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	85.06	96.46	92.50	91.42	88.61	90.81
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.21: Precision Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.83	99.25	98.98	98.96	99.10	99.03
AGREEMENT	71.32	92.25	88.72	91.44	92.34	87.21
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	90.02	92.60	76.43	90.39	89.97	87.88
EQUIPMENT	81.86	92.65	91.58	90.31	90.48	89.38
FOOTBALL_TERM	74.05	81.57	86.79	90.45	90.46	84.66
HEALTH_TERM	83.66	87.30	86.84	88.03	84.26	86.02
HORSE_RACING_TERM	4.55	0.00	0.00	0.00	0.00	0.91
MANAGER	87.08	81.45	76.16	77.63	67.93	78.05
NATION	90.08	90.26	85.64	93.32	96.17	91.09
PERSON_ALIAS	26.47	77.89	83.33	90.80	75.17	70.74
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	57.01	62.86	79.61	73.86	59.04	66.48
ROLE	90.69	93.60	93.07	94.08	92.70	92.83
SCORE_TERM	52.64	78.74	82.23	87.86	86.68	77.63
SPORT_COMPETITION	83.18	93.23	90.99	88.48	88.17	88.81
SPORT_EVENT	81.74	91.72	93.62	94.21	93.64	90.99
SPORT_LEVEL	82.01	91.32	89.30	91.01	89.81	88.69
SPORT_LOCATION	74.15	83.61	92.55	91.18	89.89	86.28
SPORT_NAME	90.50	95.58	92.23	97.74	94.79	94.17
SPORT_ORGANIZATION	85.52	78.24	92.00	90.19	84.67	86.12
SPORT_PLAYER	89.47	94.60	94.39	93.73	88.88	92.21
SUPPORTER	85.26	91.00	88.85	91.64	86.67	88.68
TEAM	91.82	93.05	94.03	96.85	95.56	94.26
TEAM_ALIAS	92.95	89.55	84.46	87.68	89.92	88.91
TEAM_SPONSOR	73.33	0.00	0.00	100.00	100.00	54.67
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.22: Recall Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.45	97.80	98.07	98.42	97.96	97.74
AGREEMENT	96.88	93.41	93.27	93.67	93.23	94.09
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	89.96	95.25	87.95	92.34	89.67	91.03
EQUIPMENT	91.55	87.88	90.88	96.37	93.08	91.95
FOOTBALL_TERM	89.11	93.96	96.17	95.07	93.54	93.57
HEALTH_TERM	65.31	79.14	90.41	91.15	85.85	82.37
HORSE_RACING_TERM	2.63	0.00	0.00	0.00	0.00	0.53
MANAGER	81.31	89.87	89.32	90.88	91.56	88.59
NATION	97.18	97.24	99.70	96.55	98.29	97.79
PERSON_ALIAS	47.37	50.34	57.14	48.17	56.19	51.84
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	92.96	92.63	80.65	84.35	86.07	87.33
ROLE	93.93	97.01	96.96	97.79	97.02	96.54
SCORE_TERM	84.87	91.74	86.99	87.36	88.95	87.98
SPORT_COMPETITION	90.31	94.54	95.76	94.65	95.98	94.25
SPORT_EVENT	94.56	96.38	93.97	95.52	95.99	95.28
SPORT_LEVEL	90.63	98.10	94.04	93.27	93.13	93.84
SPORT_LOCATION	94.91	96.34	91.47	94.25	98.55	95.10
SPORT_NAME	95.83	98.01	96.74	93.68	95.90	96.03
SPORT_ORGANIZATION	85.99	95.17	94.58	94.70	96.30	93.35
SPORT_PLAYER	94.29	96.01	92.16	94.38	90.72	93.51
SUPPORTER	85.04	93.03	94.80	87.96	93.41	90.85
TEAM	96.09	98.41	96.56	94.05	94.25	95.87
TEAM_ALIAS	81.64	95.61	93.61	96.75	98.25	93.17
TEAM_SPONSOR	21.57	0.00	0.00	39.13	32.73	18.69
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.23: Precision Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	k=1	k=2	k=3	k=4	k=5	Avg.
O	98.92	99.29	98.97	99.08	99.10	99.07
B-AGREEMENT	71.51	91.33	86.15	87.16	87.79	84.79
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	84.76	93.53	78.34	91.05	88.67	87.27
B-EQUIPMENT	77.69	91.35	91.60	89.41	86.99	87.41
B-FOOTBALL_TERM	72.95	80.86	87.32	91.81	90.66	84.72
B-HEALTH_TERM	79.63	86.51	86.36	77.94	81.25	82.34
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	81.37	74.00	68.16	73.27	65.70	72.50
B-NATION	90.16	89.95	86.06	94.85	93.81	90.96
B-PERSON_ALIAS	29.89	82.50	93.88	84.38	91.84	76.49
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	74.48	66.67	84.92	71.58	66.41	72.81
B-ROLE	88.88	93.91	93.32	93.93	92.98	92.60
B-SCORE_TERM	50.22	75.85	79.89	85.09	82.45	74.70
B-SPORT_COMPETITION	77.14	83.89	87.70	85.03	84.07	83.56
B-SPORT_EVENT	81.24	91.58	93.06	94.20	94.50	90.92
B-SPORT_LEVEL	75.38	90.86	89.66	87.95	87.38	86.24
B-SPORT_LOCATION	71.99	82.02	90.46	92.04	87.52	84.81
B-SPORT_NAME	91.44	95.37	94.05	97.57	93.51	94.39
B-SPORT_ORGANIZATION	83.47	86.43	94.23	90.20	89.78	88.82
B-SPORT_PLAYER	89.65	94.41	93.43	92.52	88.33	91.67
B-SUPPORTER	80.76	89.66	84.72	89.84	79.65	84.93
B-TEAM	90.73	91.45	94.24	96.39	95.58	93.68
B-TEAM_ALIAS	86.21	83.82	82.41	84.42	82.78	83.92
B-TEAM_SPONSOR	100.00	0.00	0.00	0.00	0.00	20.00
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.23 Continued

I-AGREEMENT	100.00	0.00	0.00	0.00	0.00	20.00
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	85.59	91.48	74.24	90.37	90.95	86.52
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	67.65	82.83	76.96	78.78	79.95	77.23
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	87.77	75.11	68.03	74.65	67.74	74.66
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	72.11	63.49	81.09	76.84	60.58	70.82
I-ROLE	85.48	87.72	79.84	84.76	82.53	84.07
I-SCORE_TERM	50.00	0.00	0.00	100.00	100.00	50.00
I-SPORT_COMPETITION	77.52	90.95	88.45	86.24	88.42	86.32
I-SPORT_EVENT	51.50	68.50	76.76	70.62	75.16	68.51
I-SPORT_LEVEL	79.32	88.99	89.26	87.81	88.19	86.72
I-SPORT_LOCATION	59.46	67.52	81.20	87.34	79.59	75.02
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	80.99	86.02	81.98	81.48	72.54	80.60
I-SPORT_PLAYER	89.01	90.54	85.49	87.29	83.20	87.11
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	81.76	90.62	86.05	86.59	85.75	86.15
I-TEAM_ALIAS	88.09	84.27	76.42	80.99	83.18	82.59
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.24: Recall Values of ConvBERT-Base Turkish Uncased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.41	97.72	98.07	98.33	97.93	97.69
B-AGREEMENT	97.03	93.30	95.56	95.65	94.92	95.29
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	91.47	94.30	86.59	92.12	87.28	90.35
B-EQUIPMENT	91.45	90.76	90.90	96.90	93.31	92.66
B-FOOTBALL_TERM	90.79	94.79	94.92	93.92	92.64	93.41
B-HEALTH_TERM	75.88	82.58	91.10	93.81	87.50	86.17
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	83.59	87.30	86.98	89.89	86.58	86.87
B-NATION	96.58	98.03	98.77	96.50	98.01	97.58
B-PERSON_ALIAS	42.62	28.95	36.51	24.77	31.47	32.86
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	73.71	84.62	60.32	76.40	77.68	74.55
B-ROLE	93.74	96.48	95.49	97.25	96.48	95.89
B-SCORE_TERM	84.84	92.66	87.50	87.24	88.60	88.17
B-SPORT_COMPETITION	85.65	91.76	89.44	89.49	92.40	89.75
B-SPORT_EVENT	95.00	95.48	95.09	95.27	95.54	95.28
B-SPORT_LEVEL	88.07	94.80	92.31	88.40	87.66	90.25
B-SPORT_LOCATION	93.98	93.63	89.89	93.19	93.27	92.79
B-SPORT_NAME	96.17	98.17	95.32	95.73	96.27	96.33
B-SPORT_ORGANIZATION	89.15	96.03	94.72	94.65	96.27	94.16
B-SPORT_PLAYER	94.12	96.09	92.94	93.60	87.24	92.80
B-SUPPORTER	86.68	94.19	96.23	87.89	96.40	92.28
B-TEAM	96.39	98.78	96.74	94.66	94.38	96.19
B-TEAM_ALIAS	81.82	96.99	91.75	95.24	98.68	92.89
B-TEAM_SPONSOR	9.68	0.00	0.00	0.00	0.00	1.94
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.24 Continued

I-AGREEMENT	6.67	0.00	0.00	0.00	0.00	1.33
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	90.26	92.71	90.24	90.82	92.47	91.30
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	75.31	88.11	92.60	94.67	93.31	88.80
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	84.88	90.44	91.50	93.81	92.79	90.68
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	93.20	93.02	78.37	81.44	93.26	87.86
I-ROLE	83.68	95.54	94.79	95.82	95.58	93.08
I-SCORE_TERM	20.59	0.00	0.00	7.69	5.00	6.66
I-SPORT_COMPETITION	91.46	96.58	95.89	95.14	96.33	95.08
I-SPORT_EVENT	53.65	82.04	67.28	82.04	85.21	74.04
I-SPORT_LEVEL	88.66	98.61	93.99	94.93	94.84	94.21
I-SPORT_LOCATION	77.65	83.16	76.60	89.03	95.12	84.31
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	73.96	88.40	90.38	88.51	91.96	86.64
I-SPORT_PLAYER	87.47	92.69	84.58	93.63	94.85	90.64
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	85.30	96.90	92.50	91.72	88.35	90.95
I-TEAM_ALIAS	91.59	97.18	97.59	96.82	98.89	96.41
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.25: Precision Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.44	97.60	97.47	97.54	97.88	97.59
AGREEMENT	76.84	91.28	87.38	88.75	90.80	87.01
ALIAS	85.54	89.61	87.16	90.56	84.59	87.49
EQUIPMENT	81.04	86.62	91.32	89.33	89.22	87.51
HEALTH_TERM	92.73	95.10	88.24	94.12	90.00	92.04
NATION	90.28	79.10	81.63	87.22	87.06	85.06
ROLE	85.67	88.79	89.01	88.67	89.37	88.30
SPORT_COMPETITION	69.12	76.87	73.38	69.76	72.85	72.39
SPORT_EVENT	81.35	88.59	91.63	90.39	89.93	88.38
SPORT_LEVEL	75.98	79.57	84.01	86.12	83.90	81.92
SPORT_LOCATION	74.55	80.00	87.13	84.66	89.46	83.16
SPORT_NAME	86.24	86.20	89.62	89.29	88.18	87.91
SPORT_ORGANIZATION	71.33	73.67	70.12	72.67	73.52	72.26
SPORT_PERSON	75.38	77.56	75.52	74.39	74.89	75.55
SPORT_TERM	64.85	78.79	82.14	86.19	83.42	79.08
SUPPORTER	88.63	97.40	95.36	96.11	95.77	94.65
TEAM	77.67	79.54	82.56	85.46	80.93	81.23
TEAM_SPONSOR	75.00	0.00	0.00	100.00	100.00	55.00

Table A.26: Recall Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	94.99	96.08	96.56	96.88	96.82	96.27
AGREEMENT	89.61	89.75	84.29	89.61	87.29	88.11
ALIAS	74.87	77.09	63.80	78.84	75.98	74.12
EQUIPMENT	74.65	78.48	76.75	86.57	85.88	80.47
HEALTH_TERM	52.04	69.78	61.64	56.64	67.92	61.61

Table A.26 Continued

NATION	73.24	67.96	70.21	68.14	60.88	68.09
ROLE	90.07	93.10	92.59	93.85	93.77	92.67
SPORT_COMPETITION	73.27	75.18	80.26	76.22	79.89	76.96
SPORT_EVENT	90.66	92.43	89.28	90.09	91.71	90.83
SPORT_LEVEL	78.01	75.67	79.55	79.82	73.42	77.29
SPORT_LOCATION	75.64	74.76	63.82	65.78	68.34	69.67
SPORT_NAME	84.63	92.75	84.44	90.18	94.73	89.35
SPORT_ORGANIZATION	67.35	76.57	78.50	75.51	81.16	75.82
SPORT_PERSON	87.80	88.22	89.89	90.45	88.37	88.94
SPORT_TERM	79.51	79.91	81.15	82.50	81.06	80.83
SUPPORTER	79.79	83.83	84.01	76.23	81.44	81.06
TEAM	82.77	86.11	84.85	80.00	85.12	83.77
TEAM_SPONSOR	5.88	0.00	0.00	13.04	16.36	7.06

Table A.27: Precision Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.56	97.18	97.37	97.57	97.74	97.49
B-AGREEMENT	71.91	86.62	81.50	83.84	85.26	81.83
B-ALIAS	82.26	85.26	84.15	87.54	82.99	84.44
B-EQUIPMENT	74.02	83.02	86.61	87.01	86.51	83.43
B-HEALTH_TERM	92.11	89.09	89.55	94.44	89.23	90.88
B-NATION	80.42	72.51	74.24	81.16	79.86	77.64
B-ROLE	83.94	89.19	86.80	87.90	88.14	87.19
B-SPORT_COMPETITION	62.44	68.74	69.88	66.38	67.22	66.93
B-SPORT_EVENT	77.37	85.82	88.77	87.22	88.30	85.49
B-SPORT_LEVEL	73.48	74.73	75.31	84.37	79.03	77.39
B-SPORT_LOCATION	71.37	76.59	80.39	81.97	83.82	78.83
B-SPORT_NAME	83.92	84.89	87.50	87.87	87.77	86.39

Table A.27 Continued

B-SPORT_ORGANIZATION	64.10	72.55	64.30	68.00	54.00	64.59
B-SPORT_PERSON	69.38	71.91	72.08	65.84	67.64	69.37
B-SPORT_TERM	61.20	74.82	79.16	82.55	83.60	76.26
B-SUPPORTER	88.62	95.04	94.71	94.59	95.67	93.73
B-TEAM	76.35	78.75	81.94	84.02	81.68	80.55
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-ALIAS	96.55	96.81	95.39	97.14	95.62	96.30
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	74.12	72.69	71.21	72.80	77.56	73.68
I-SPORT_COMPETITION	60.56	71.99	68.70	63.53	65.19	65.99
I-SPORT_EVENT	71.43	50.00	75.00	50.00	61.70	61.63
I-SPORT_LEVEL	71.94	75.35	80.97	76.80	78.63	76.74
I-SPORT_LOCATION	57.14	42.86	100.00	100.00	75.00	75.00
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	79.66	78.89	60.00	78.41	66.23	72.64
I-SPORT_PERSON	66.25	62.90	60.85	63.49	58.37	62.37
I-SPORT_TERM	55.26	71.35	63.77	71.21	65.41	65.40
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	56.70	76.97	48.83	56.79	56.08	59.07
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.28: Recall Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	94.95	96.20	96.62	96.77	96.86	96.28
B-AGREEMENT	91.35	90.17	88.74	91.13	89.84	90.25

Table A.28 Continued

B-ALIAS	64.88	64.41	48.13	68.06	63.60	61.81
B-EQUIPMENT	83.19	84.08	78.86	89.07	85.76	84.19
B-HEALTH_TERM	41.18	37.12	41.10	45.13	55.77	44.06
B-NATION	76.07	67.42	67.38	65.24	58.21	66.86
B-ROLE	89.46	91.13	89.95	92.79	92.27	91.12
B-SPORT_COMPETITION	53.20	55.17	57.77	57.79	64.75	57.73
B-SPORT_EVENT	93.56	92.22	92.04	91.79	91.91	92.30
B-SPORT_LEVEL	68.07	60.69	71.30	73.71	62.03	67.16
B-SPORT_LOCATION	77.20	70.99	65.40	68.46	67.17	69.85
B-SPORT_NAME	87.76	94.69	85.52	91.47	94.50	90.79
B-SPORT_ORGANIZATION	72.55	78.31	74.14	76.45	82.89	76.87
B-SPORT_PERSON	83.05	83.70	83.98	84.73	79.60	83.01
B-SPORT_TERM	81.74	81.96	82.29	83.19	82.22	82.28
B-SUPPORTER	78.26	82.32	81.13	76.09	79.58	79.48
B-TEAM	85.87	84.38	85.19	80.54	84.19	84.03
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-ALIAS	81.33	86.48	80.56	86.55	81.62	83.31
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	73.86	81.95	70.55	73.37	84.42	76.83
I-SPORT_COMPETITION	72.81	71.99	79.45	82.68	84.63	78.31
I-SPORT_EVENT	15.63	15.57	12.96	10.78	20.42	15.07
I-SPORT_LEVEL	76.47	75.61	76.68	82.77	77.38	77.78
I-SPORT_LOCATION	9.41	6.32	6.38	6.45	1.83	6.08
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	35.47	39.23	48.08	39.66	45.54	41.59
I-SPORT_PERSON	72.70	71.14	70.28	69.31	71.78	71.04
I-SPORT_TERM	54.02	50.56	61.82	60.51	55.47	56.47

Table A.28 Continued

I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	39.76	57.67	44.64	47.04	42.03	46.23
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.29: Precision Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.41	97.45	97.39	97.63	97.91	97.56
AGREEMENT	78.03	92.38	87.30	89.40	90.86	87.59
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	54.50	62.09	69.97	62.81	60.79	62.03
EQUIPMENT	80.24	86.49	89.49	88.69	87.61	86.50
FOOTBALL_TERM	69.45	77.43	79.81	84.83	82.88	78.88
HEALTH_TERM	94.90	95.88	86.96	94.03	89.04	92.16
HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
MANAGER	63.55	52.16	47.05	59.10	53.51	55.08
NATION	85.39	69.60	72.99	82.92	81.40	78.46
PERSON_ALIAS	15.79	0.00	0.00	0.00	9.09	4.98
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	100.00	100.00	100.00	0.00	100.00	80.00
ROLE	85.89	87.21	86.30	87.32	86.54	86.65
SCORE_TERM	51.03	72.12	85.15	81.38	82.89	74.51
SPORT_COMPETITION	60.07	68.01	67.83	66.21	68.02	66.03
SPORT_EVENT	77.70	86.67	89.76	89.09	89.42	86.53
SPORT_LEVEL	72.92	76.80	81.56	81.47	80.62	78.67
SPORT_LOCATION	72.11	82.59	84.08	84.85	90.04	82.73
SPORT_NAME	86.85	87.95	90.35	91.30	89.48	89.19
SPORT_ORGANIZATION	70.18	76.42	66.08	73.72	63.64	70.00
SPORT_PLAYER	59.23	66.97	63.60	60.40	60.67	62.17

Table A.29 Continued

SUPPORTER	90.52	95.42	94.47	94.62	92.47	93.50
TEAM	74.45	76.58	78.95	81.75	78.64	78.07
TEAM_ALIAS	91.18	84.98	85.99	80.04	79.62	84.36
TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.30: Recall Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	94.95	95.92	96.47	96.69	96.68	96.14
AGREEMENT	90.39	88.81	85.90	87.66	87.85	88.12
BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
COACH	57.32	66.59	45.11	57.14	46.40	54.51
EQUIPMENT	76.62	77.58	75.08	88.20	83.57	80.21
FOOTBALL_TERM	79.97	79.22	86.77	84.76	80.27	82.20
HEALTH_TERM	47.45	66.91	54.79	55.75	61.32	57.24
HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
MANAGER	73.36	71.96	65.05	66.89	79.61	71.38
NATION	74.08	67.68	69.00	64.30	59.90	66.99
PERSON_ALIAS	3.95	0.00	0.00	0.00	0.52	0.89
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	0.29	9.47	1.96	0.00	2.49	2.84
ROLE	90.13	94.14	92.86	94.40	94.54	93.21
SCORE_TERM	77.74	79.51	64.40	73.33	80.94	75.19
SPORT_COMPETITION	73.39	77.46	82.43	77.84	83.10	78.84
SPORT_EVENT	91.09	91.63	89.06	91.12	91.20	90.82
SPORT_LEVEL	77.25	74.25	79.07	82.89	78.35	78.36
SPORT_LOCATION	73.82	74.95	60.72	65.22	64.38	67.82

Table A.30 Continued

SPORT_NAME	84.48	92.57	83.26	89.02	94.73	88.81
SPORT_ORGANIZATION	65.31	72.45	76.82	71.76	77.64	72.80
SPORT_PLAYER	74.05	72.27	84.61	84.69	78.33	78.79
SUPPORTER	77.69	82.84	82.53	75.93	80.84	79.96
TEAM	83.67	85.59	86.41	80.69	85.22	84.32
TEAM_ALIAS	74.25	81.72	75.00	82.96	81.29	79.05
TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.31: Precision Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	97.13	96.89	96.93	97.40	97.54	97.18
B-AGREEMENT	74.83	90.63	84.39	85.06	90.40	85.06
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	59.70	70.33	66.40	62.42	58.25	63.42
B-EQUIPMENT	75.97	84.92	89.35	87.03	86.89	84.83
B-FOOTBALL_TERM	67.98	75.81	78.77	83.52	84.69	78.15
B-HEALTH_TERM	93.88	93.33	87.10	92.54	90.28	91.42
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	63.68	52.75	36.78	50.61	49.09	50.58
B-NATION	81.50	73.75	78.89	81.15	77.02	78.46
B-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	100.00	100.00	100.00	100.00	100.00	100.00
B-ROLE	82.69	88.55	85.19	85.86	86.45	85.75
B-SCORE_TERM	49.09	72.69	80.00	81.87	82.04	73.14
B-SPORT_COMPETITION	59.41	61.67	63.17	63.32	65.00	62.51

Table A.31 Continued

B-SPORT_EVENT	75.95	85.37	88.19	85.99	86.10	84.32
B-SPORT_LEVEL	68.21	73.26	73.56	82.32	78.69	75.21
B-SPORT_LOCATION	74.89	76.73	80.92	76.56	81.43	78.11
B-SPORT_NAME	85.17	85.48	89.24	87.54	86.43	86.77
B-SPORT_ORGANIZATION	62.69	71.68	62.96	67.78	66.67	66.36
B-SPORT_PLAYER	56.30	61.93	62.37	54.94	56.30	58.37
B-SUPPORTER	89.97	95.06	93.07	94.98	96.04	93.82
B-TEAM	74.90	70.44	77.95	82.24	77.95	76.70
B-TEAM_ALIAS	88.94	84.31	85.28	79.50	76.92	82.99
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	59.06	72.50	72.84	76.14	71.94	70.50
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	54.75	68.78	65.01	69.43	64.74	64.54
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	58.31	46.34	30.53	43.79	41.91	44.17
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	76.25	72.41	69.44	76.31	76.32	74.15
I-SCORE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_COMPETITION	63.07	73.54	67.75	64.49	66.01	66.97
I-SPORT_EVENT	66.67	53.19	84.00	55.00	59.26	63.62
I-SPORT_LEVEL	71.78	70.43	79.71	74.39	74.32	74.13
I-SPORT_LOCATION	100.00	100.00	0.00	100.00	0.00	60.00

Table A.31 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	78.51	76.60	56.33	76.77	58.51	69.34
I-SPORT_PLAYER	40.21	45.72	43.36	46.31	43.25	43.77
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	68.35	82.81	59.43	57.76	57.79	65.23
I-TEAM_ALIAS	95.85	88.98	91.72	82.91	82.09	88.31
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.32: Recall Values of DistilBERT-Base Turkish Cased Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	95.47	96.61	96.81	96.89	97.15	96.59
B-AGREEMENT	90.00	88.77	86.69	91.13	88.82	89.08
B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	48.41	55.36	41.10	47.88	34.54	45.46
B-EQUIPMENT	81.12	82.48	73.77	89.25	82.85	81.89
B-FOOTBALL_TERM	84.76	83.06	86.34	85.98	81.10	84.25
B-HEALTH_TERM	54.12	63.64	55.48	54.87	62.50	58.12
B-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-MANAGER	60.88	55.72	46.35	56.28	60.88	56.02
B-NATION	74.07	66.29	70.15	66.02	59.20	67.15
B-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	0.52	3.85	0.40	2.81	3.57	2.23
B-ROLE	89.74	91.73	90.97	93.51	93.43	91.88
B-SCORE_TERM	75.05	80.22	64.29	73.89	80.12	74.71
B-SPORT_COMPETITION	53.64	56.70	61.16	57.97	59.91	57.88

Table A.32 Continued

B-SPORT_EVENT	93.79	92.13	92.00	92.65	93.10	92.73
B-SPORT_LEVEL	67.02	60.98	71.60	73.20	60.76	66.71
B-SPORT_LOCATION	73.76	70.75	63.67	70.25	67.17	69.12
B-SPORT_NAME	85.55	93.77	85.07	91.13	95.09	90.12
B-SPORT_ORGANIZATION	71.49	74.34	71.77	73.88	78.95	74.08
B-SPORT_PLAYER	74.61	72.86	78.78	80.61	74.09	76.19
B-SUPPORTER	77.99	82.58	81.13	76.40	80.18	79.65
B-TEAM	86.73	86.38	86.73	80.84	86.65	85.46
B-TEAM_ALIAS	73.09	77.26	71.65	80.95	76.16	75.82
B-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	47.98	56.10	39.73	53.85	38.08	47.15
I-EQUIPMENT	0.00	0.00	0.00	0.00	0.00	0.00
I-FOOTBALL_TERM	59.26	55.20	68.76	66.43	64.74	62.88
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	64.07	58.91	58.00	68.58	74.65	64.84
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	0.00	0.00	0.00	0.00	0.00	0.00
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	72.11	83.01	69.02	72.32	83.95	76.08
I-SCORE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_COMPETITION	71.01	71.99	81.41	80.54	85.09	78.01
I-SPORT_EVENT	14.58	14.97	12.96	13.17	11.27	13.39
I-SPORT_LEVEL	72.69	73.87	77.74	82.43	75.79	76.50
I-SPORT_LOCATION	1.18	1.05	0.00	0.65	0.00	0.57

Table A.32 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	35.85	39.78	57.05	43.68	49.11	45.09
I-SPORT_PLAYER	48.85	45.95	52.67	51.96	54.16	50.72
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	39.04	54.72	37.14	39.64	35.70	41.25
I-TEAM_ALIAS	81.86	84.68	80.12	88.18	81.48	83.26
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.33: Precision Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.82	99.01	98.75	98.92	98.88	98.88
AGREEMENT	74.46	91.85	87.83	89.21	91.32	86.93
ALIAS	85.98	91.69	88.98	95.83	86.77	89.85
EQUIPMENT	83.81	94.24	91.09	91.61	91.41	90.43
HEALTH_TERM	79.65	83.33	86.54	74.31	86.92	82.15
NATION	90.11	88.83	87.83	93.59	95.00	91.07
ROLE	90.51	94.45	93.60	93.15	93.75	93.09
SPORT_COMPETITION	85.71	92.13	88.97	87.52	89.20	88.71
SPORT_EVENT	83.12	92.17	93.36	94.26	93.51	91.29
SPORT_LEVEL	85.54	92.38	93.06	91.51	90.41	90.58
SPORT_LOCATION	76.29	83.00	93.78	91.30	91.78	87.23
SPORT_NAME	93.37	92.59	93.27	97.10	94.76	94.22
SPORT_ORGANIZATION	88.02	87.42	92.51	94.27	88.98	90.24
SPORT_PERSON	89.79	93.25	92.53	89.97	86.79	90.47
SPORT_TERM	63.71	84.16	88.47	91.37	91.51	83.84
SUPPORTER	86.42	93.28	87.88	91.40	95.72	90.94

Table A.33 Continued

TEAM	90.31	92.13	92.94	95.70	94.93	93.20
TEAM_SPONSOR	49.21	90.67	0.00	78.26	76.67	58.96

Table A.34: Recall Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.01	97.76	98.11	98.10	98.03	97.60
AGREEMENT	97.66	93.10	94.87	93.99	91.57	94.24
ALIAS	87.18	92.22	85.60	90.87	96.74	90.52
EQUIPMENT	90.42	84.24	91.64	97.10	95.10	91.70
HEALTH_TERM	69.90	82.73	92.47	94.69	87.74	85.50
NATION	94.93	96.69	92.10	92.51	92.91	93.83
ROLE	93.87	96.99	96.45	98.17	97.11	96.52
SPORT_COMPETITION	88.86	92.69	93.19	93.12	95.86	92.75
SPORT_EVENT	94.73	96.47	95.27	95.93	95.81	95.64
SPORT_LEVEL	91.59	95.73	92.91	94.59	92.96	93.56
SPORT_LOCATION	91.27	96.92	87.73	95.65	95.78	93.47
SPORT_NAME	95.11	97.28	94.37	94.68	95.31	95.35
SPORT_ORGANIZATION	85.99	94.45	92.34	89.86	95.25	91.58
SPORT_PERSON	95.11	96.41	95.78	96.76	94.07	95.62
SPORT_TERM	89.20	90.22	89.12	92.14	89.94	90.12
SUPPORTER	86.88	93.28	97.03	88.58	93.71	91.90
TEAM	93.60	96.61	95.64	93.27	92.83	94.39
TEAM_SPONSOR	60.78	72.34	0.00	78.26	41.82	50.64

Table A.35: Precision Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	k=1	k=2	k=3	k=4	k=5	Avg.
O	98.85	99.01	98.77	98.89	98.84	98.87
B-AGREEMENT	73.82	92.55	85.76	90.36	91.78	86.85
B-ALIAS	80.83	89.66	86.08	91.96	86.63	87.03
B-EQUIPMENT	83.15	94.14	92.07	91.42	92.02	90.56
B-HEALTH_TERM	80.63	83.58	83.75	77.14	80.36	81.09
B-NATION	89.54	88.30	86.72	93.01	93.62	90.24
B-ROLE	88.56	94.47	93.65	93.53	93.50	92.74
B-SPORT_COMPETITION	83.05	88.13	85.63	84.83	86.45	85.62
B-SPORT_EVENT	82.96	92.26	93.70	93.89	94.30	91.42
B-SPORT_LEVEL	77.51	90.14	90.49	88.50	91.12	87.55
B-SPORT_LOCATION	75.00	83.44	92.11	91.13	89.22	86.18
B-SPORT_NAME	92.88	91.61	96.19	96.74	94.37	94.36
B-SPORT_ORGANIZATION	82.63	86.94	94.82	93.04	91.19	89.73
B-SPORT_PERSON	87.21	91.43	91.60	89.31	86.95	89.30
B-SPORT_TERM	63.88	82.41	87.05	92.26	91.29	83.38
B-SUPPORTER	83.64	94.82	90.22	93.11	93.92	91.14
B-TEAM	90.47	90.95	92.60	95.33	94.64	92.80
B-TEAM_SPONSOR	85.00	95.16	20.00	90.91	100.00	78.21
I-AGREEMENT	84.62	78.26	100.00	80.00	86.96	85.97
I-ALIAS	92.61	93.81	90.76	97.73	92.65	93.51
I-EQUIPMENT	100.00	0.00	0.00	0.00	0.00	20.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.71	81.32	83.06	84.21	86.02	83.86
I-SPORT_COMPETITION	85.20	90.39	86.38	85.79	86.96	86.94
I-SPORT_EVENT	60.89	72.45	80.14	72.34	65.98	70.36
I-SPORT_LEVEL	82.94	88.40	91.58	87.90	86.45	87.45
I-SPORT_LOCATION	60.64	69.29	81.88	77.84	78.19	73.57

Table A.35 Continued

I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	80.59	88.83	83.15	86.21	73.43	82.44
I-SPORT_PERSON	87.64	87.38	86.15	84.41	78.61	84.84
I-SPORT_TERM	57.99	84.04	79.02	81.46	86.03	77.71
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	83.24	92.03	83.27	82.71	82.06	84.66
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.36: Recall Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Parent-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.09	97.66	98.04	98.13	98.04	97.59
B-AGREEMENT	97.57	93.84	96.59	96.17	94.05	95.64
B-ALIAS	86.61	90.31	83.13	89.79	94.61	88.89
B-EQUIPMENT	88.79	92.04	91.36	97.09	93.90	92.63
B-HEALTH_TERM	75.88	84.85	91.78	95.58	86.54	86.93
B-NATION	95.16	97.47	94.46	93.01	91.29	94.28
B-ROLE	93.49	95.63	94.95	96.74	95.96	95.35
B-SPORT_COMPETITION	86.53	86.78	87.85	90.24	92.63	88.81
B-SPORT_EVENT	96.13	94.84	94.77	94.92	95.13	95.16
B-SPORT_LEVEL	89.47	92.49	92.90	91.24	87.66	90.75
B-SPORT_LOCATION	91.61	96.23	88.47	92.11	91.92	92.07
B-SPORT_NAME	96.17	97.99	95.17	96.08	95.48	96.18
B-SPORT_ORGANIZATION	88.09	96.83	91.82	91.65	95.39	92.75
B-SPORT_PERSON	94.37	94.44	93.43	94.27	89.30	93.16
B-SPORT_TERM	89.13	90.18	88.54	90.87	89.33	89.61
B-SUPPORTER	86.14	92.42	93.96	88.20	92.79	90.70
B-TEAM	94.66	97.50	95.72	92.61	93.06	94.71
B-TEAM_SPONSOR	54.84	80.82	12.50	58.82	46.15	50.63

Table A.36 Continued

I-AGREEMENT	73.33	60.00	63.16	58.54	57.14	62.43
I-ALIAS	88.38	97.15	92.78	93.82	98.13	94.05
I-EQUIPMENT	25.00	0.00	0.00	0.00	0.00	5.00
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-ROLE	84.56	94.27	93.25	96.08	94.42	92.52
I-SPORT_COMPETITION	90.56	94.95	94.32	93.97	96.33	94.03
I-SPORT_EVENT	56.77	85.03	72.22	81.44	90.14	77.12
I-SPORT_LEVEL	87.82	90.24	92.23	93.24	93.65	91.44
I-SPORT_LOCATION	67.06	92.63	80.14	88.39	89.63	83.57
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	72.08	87.85	94.87	86.21	93.75	86.95
I-SPORT_PERSON	90.55	92.72	92.83	93.85	94.70	92.93
I-SPORT_TERM	74.58	82.14	84.90	87.63	83.73	82.60
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	73.01	90.27	83.57	84.91	78.73	82.10
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00

Table A.37: Precision Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.73	98.96	98.65	98.94	98.90	98.84
AGREEMENT	74.06	91.43	84.81	88.50	91.23	86.01
BASKETBALL_TERM	0.00	0.00	100.00	0.00	100.00	40.00
COACH	86.85	90.67	74.55	90.12	88.19	86.08
EQUIPMENT	84.29	95.19	90.66	90.63	91.90	90.54
FOOTBALL_TERM	70.30	82.15	88.75	90.79	91.57	84.71
HEALTH_TERM	76.54	83.09	86.62	76.76	86.54	81.91
HORSE_RACING_TERM	33.02	0.00	0.00	0.00	0.00	6.60

Table A.37 Continued

MANAGER	85.24	80.82	61.76	72.65	65.59	73.21
NATION	90.79	89.43	87.86	94.04	95.66	91.56
PERSON_ALIAS	27.70	73.57	82.69	86.05	75.23	69.05
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	83.61	59.86	82.43	82.13	66.09	74.83
ROLE	90.38	94.03	93.70	93.38	93.02	92.90
SCORE_TERM	52.31	79.87	83.05	87.30	88.55	78.22
SPORT_COMPETITION	84.92	92.53	88.34	87.10	88.83	88.34
SPORT_EVENT	82.84	93.05	92.59	93.71	93.53	91.14
SPORT_LEVEL	84.98	91.67	93.06	91.36	91.83	90.58
SPORT_LOCATION	76.36	82.19	92.05	91.60	89.16	86.27
SPORT_NAME	92.18	93.86	93.70	97.09	93.87	94.14
SPORT_ORGANIZATION	88.94	87.28	93.74	92.13	87.52	89.92
SPORT_PLAYER	81.13	91.04	91.36	87.82	85.09	87.29
SUPPORTER	85.71	94.12	87.71	91.35	94.80	90.74
TEAM	90.17	91.98	93.50	95.69	95.61	93.39
TEAM_ALIAS	90.29	93.50	85.56	92.77	91.17	90.66
TEAM_SPONSOR	95.83	97.14	0.00	100.00	54.55	69.50
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.38: Recall Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-NonBIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.02	97.76	98.12	98.14	97.98	97.60
AGREEMENT	97.14	93.72	94.87	94.97	91.99	94.54
BASKETBALL_TERM	0.00	0.00	5.13	0.00	4.88	2.00
COACH	87.16	93.19	83.74	85.87	82.13	86.42
EQUIPMENT	90.70	90.00	91.49	96.55	94.81	92.71

Table A.38 Continued

FOOTBALL_TERM	89.59	92.34	92.92	93.92	90.53	91.86
HEALTH_TERM	69.90	81.29	93.15	96.46	84.91	85.14
HORSE_RACING_TERM	92.11	0.00	0.00	0.00	0.00	18.42
MANAGER	83.04	84.01	88.74	87.50	92.44	87.14
NATION	94.37	95.86	92.40	93.86	91.69	93.63
PERSON_ALIAS	53.95	70.07	61.43	67.68	82.99	67.22
RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
REFEREE	89.74	89.47	66.30	82.61	76.62	80.95
ROLE	94.37	96.58	96.26	97.96	97.44	96.52
SCORE_TERM	86.78	92.81	87.76	88.51	88.67	88.91
SPORT_COMPETITION	89.64	93.75	93.48	92.84	95.98	93.14
SPORT_EVENT	94.52	96.22	95.35	96.22	96.72	95.80
SPORT_LEVEL	91.97	95.58	92.91	95.91	92.96	93.86
SPORT_LOCATION	92.18	96.92	88.24	94.81	95.51	93.53
SPORT_NAME	96.55	96.92	94.81	94.34	95.70	95.67
SPORT_ORGANIZATION	84.22	95.71	95.14	91.26	95.07	92.28
SPORT_PLAYER	92.91	92.40	89.34	94.02	89.33	91.60
SUPPORTER	86.61	91.54	95.54	87.96	92.81	90.89
TEAM	93.11	96.82	94.78	94.03	92.70	94.29
TEAM_ALIAS	79.84	94.70	90.56	93.71	95.63	90.89
TEAM_SPONSOR	45.10	72.34	0.00	86.96	43.64	49.61
TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.39: Precision Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	98.77	98.99	98.71	98.91	98.81	98.84
B-AGREEMENT	73.13	92.69	85.37	90.34	91.70	86.65

Table A.39 Continued

B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	84.66	89.31	78.41	89.99	88.61	86.20
B-EQUIPMENT	81.60	94.93	90.51	91.18	90.93	89.83
B-FOOTBALL_TERM	72.05	80.04	88.78	90.78	92.04	84.74
B-HEALTH_TERM	77.58	83.97	86.08	77.14	84.40	81.83
B-HORSE_RACING_TERM	21.36	0.00	0.00	0.00	0.00	4.27
B-MANAGER	82.25	78.97	65.61	75.91	63.84	73.32
B-NATION	89.78	86.07	87.46	93.01	95.26	90.32
B-PERSON_ALIAS	28.70	74.55	83.67	79.07	73.49	67.90
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	83.58	64.00	78.90	78.92	56.69	72.42
B-ROLE	88.72	94.72	93.01	93.77	93.39	92.72
B-SCORE_TERM	51.45	78.97	81.97	86.62	85.53	76.91
B-SPORT_COMPETITION	83.51	88.34	84.93	85.79	86.05	85.72
B-SPORT_EVENT	83.15	91.97	93.45	94.39	94.27	91.44
B-SPORT_LEVEL	78.33	92.49	89.24	87.37	88.82	87.25
B-SPORT_LOCATION	73.96	83.13	90.97	91.27	88.30	85.53
B-SPORT_NAME	90.29	90.96	94.74	96.74	94.89	93.52
B-SPORT_ORGANIZATION	84.48	87.32	93.32	90.42	90.23	89.15
B-SPORT_PLAYER	82.16	88.96	89.20	85.01	84.55	85.98
B-SUPPORTER	84.27	93.80	88.69	91.88	94.21	90.57
B-TEAM	89.62	90.46	92.44	95.71	94.51	92.55
B-TEAM_ALIAS	89.71	92.41	88.83	92.88	91.96	91.16
B-TEAM_SPONSOR	84.62	100.00	0.00	100.00	100.00	76.92
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	73.33	75.00	100.00	80.65	80.00	81.80
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	82.54	85.35	66.67	89.27	90.00	82.77
I-EQUIPMENT	100.00	0.00	0.00	0.00	0.00	20.00

Table A.39 Continued

I-FOOTBALL_TERM	57.84	83.91	78.35	77.47	83.19	76.15
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	86.00	74.84	67.03	74.62	65.35	73.57
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	14.29	76.00	77.78	96.30	92.50	71.37
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	84.38	66.10	76.50	82.28	65.18	74.89
I-ROLE	85.51	82.28	82.92	83.37	85.86	83.99
I-SCORE_TERM	42.86	60.71	74.47	64.00	75.00	63.41
I-SPORT_COMPETITION	83.71	90.56	85.92	85.49	86.83	86.50
I-SPORT_EVENT	56.32	71.50	75.66	69.84	67.37	68.14
I-SPORT_LEVEL	84.13	88.85	90.63	86.69	86.94	87.45
I-SPORT_LOCATION	56.86	73.50	80.88	82.72	78.31	74.45
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	80.25	80.90	80.77	83.24	71.72	79.38
I-SPORT_PLAYER	75.41	82.34	75.12	77.13	73.43	76.69
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	78.92	90.84	82.72	81.69	82.11	83.26
I-TEAM_ALIAS	91.19	92.16	88.82	90.56	92.14	90.97
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00

Table A.40: Recall Values of XLM-RoBERTa-Base Multilingual Model Trained on TurkishSportsNER-Child-BIO Data Set

Labels	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**Avg.**
O	96.17	97.66	98.08	98.14	97.98	97.60
B-AGREEMENT	97.84	94.49	95.56	96.00	93.03	95.39

Table A.40 Continued

B-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-COACH	86.23	91.21	81.45	85.65	83.76	85.66
B-EQUIPMENT	90.27	89.49	91.20	94.17	93.31	91.69
B-FOOTBALL_TERM	91.23	92.27	92.67	93.11	89.22	91.70
B-HEALTH_TERM	75.29	83.33	93.15	95.58	88.46	87.16
B-HORSE_RACING_TERM	70.97	0.00	0.00	0.00	0.00	14.19
B-MANAGER	79.90	83.12	85.40	85.25	81.78	83.09
B-NATION	95.16	97.19	92.31	95.53	90.05	94.05
B-PERSON_ALIAS	54.10	71.93	65.08	62.39	85.31	67.76
B-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-REFEREE	86.60	92.31	68.25	73.60	64.29	77.01
B-ROLE	93.32	95.94	95.13	95.90	96.17	95.29
B-SCORE_TERM	85.03	92.82	85.00	87.24	88.16	87.65
B-SPORT_COMPETITION	87.20	88.51	86.45	89.49	92.40	88.81
B-SPORT_EVENT	95.14	95.25	94.85	95.39	95.26	95.18
B-SPORT_LEVEL	88.77	92.49	93.20	89.18	87.97	90.32
B-SPORT_LOCATION	91.61	95.28	89.10	93.73	92.76	92.50
B-SPORT_NAME	96.02	97.62	95.02	96.25	94.89	95.96
B-SPORT_ORGANIZATION	89.15	94.71	92.08	92.93	95.18	92.81
B-SPORT_PLAYER	91.83	90.95	89.85	91.16	85.10	89.78
B-SUPPORTER	85.87	91.67	94.72	87.89	92.79	90.59
B-TEAM	94.54	97.05	94.70	92.85	92.24	94.28
B-TEAM_ALIAS	79.27	93.65	86.08	90.84	94.70	88.91
B-TEAM_SPONSOR	35.48	68.49	0.00	58.82	51.28	42.82
B-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
B-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-AGREEMENT	73.33	60.00	73.68	60.98	57.14	65.03
I-BASKETBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-COACH	86.46	90.18	78.79	84.62	88.49	85.71
I-EQUIPMENT	18.75	0.00	0.00	0.00	0.00	3.75

Table A.40 Continued

I-FOOTBALL_TERM	75.10	87.47	90.34	91.64	85.71	86.05
I-HEALTH_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-HORSE_RACING_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-MANAGER	83.68	89.15	92.50	85.84	92.56	88.75
I-NATION	0.00	0.00	0.00	0.00	0.00	0.00
I-PERSON_ALIAS	33.33	57.58	50.00	47.27	72.55	52.15
I-RACE_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-REFEREE	91.84	90.70	79.81	77.84	82.02	84.44
I-ROLE	83.86	93.63	92.33	95.56	94.65	92.01
I-SCORE_TERM	61.76	62.96	59.32	41.03	45.00	54.02
I-SPORT_COMPETITION	92.36	95.28	94.32	93.97	96.79	94.54
I-SPORT_EVENT	55.73	82.63	70.99	79.04	90.14	75.71
I-SPORT_LEVEL	89.08	91.64	92.23	94.59	92.46	92.00
I-SPORT_LOCATION	68.24	90.53	78.01	86.45	90.24	82.69
I-SPORT_NAME	0.00	0.00	0.00	0.00	0.00	0.00
I-SPORT_ORGANIZATION	73.58	88.95	94.23	85.63	92.86	87.05
I-SPORT_PLAYER	80.00	82.02	75.21	90.06	91.28	83.71
I-SUPPORTER	0.00	0.00	0.00	0.00	0.00	0.00
I-TEAM	73.98	89.23	80.36	85.80	78.99	81.67
I-TEAM_ALIAS	77.88	94.76	90.96	95.91	95.56	91.01
I-TEAM_SPONSOR	0.00	0.00	0.00	0.00	0.00	0.00
I-TENNIS_TERM	0.00	0.00	0.00	0.00	0.00	0.00
I-VOLLEYBALL_TERM	0.00	0.00	0.00	0.00	0.00	0.00