Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 2939–2969
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2023.26802.3413

Prediction of football match results by using artificial intelligence-based methods and proposal of hybrid methods

İsmail Hakkı Kinalioğlu[a,*], Coşkun Kuş[b]

[a] *Department of Actuerial Science, Selcuk University, 42250 Konya, Turkey*

[b] *Department of Statistics, Selcuk University, 42250 Konya, Turkey*

(Communicated by Javad Damirchi)

Abstract

In this study, hybrid classification methods are proposed, and they are used to predict the results of future football matches. Our hybrid classification methods are introduced by using clustering and classification algorithms together. By developing a web scraping tool, data on 6396 football matches played in European leagues are collected. Unlike similar studies, the data includes fans' opinions gathered from social media platforms in addition to statistical information about the teams and players. The raw data is transformed into suitable datasets through a software developed by authors, and the processed data is used in the classification analysis. The match result variable (dependent variable) is considered as three types denoted by MR-1, MR-2 and MR-3, respectively: The first one has three classes with Home, Draw and Away, the second one has two classes with Home and Draw-Away, the last one has also two classes Home-Draw and Away. The performances of the proposed hybrid methods are compared with the classification algorithms frequently used in the literature. As a result, our hybrid methods are more successful than classical classification algorithms. The prediction successes are 65.46% in the case of MR-1, 81.76% in the case of MR-2, and 77.8% in the case of MR-3.

Keywords: Clustering Analysis, Data Mining, Prediction, Football Statistics, Hybrid Classifier, Machine Learning, Multiple Classification, Sentiment Analysis, Sport Analytics
2020 MSC: Primary 68T05; Secondary 68T07, 68T09, 68T10

1 Introduction

Among the sports branches that have become universal values, football is the most popular compared to the others [11, 14, 11, 22, 30, 23]. Thanks to this popularity, the cultural and economic value of football is reached very high levels. According to a recent study, the market size of European football reached 28.4 billion euros in the 2017-18 season [28]. Various relationships are established between football and disciplines such as training science, economics, law, management science, psychology, statistics and engineering in order to maintain and expand the cultural and economic greatness of football.

One of the important relationships is between football and statistics. For many years, various statistical data are collected from football matches. The diversity of these data collected with the impact of technological innovations

*Corresponding author
Email addresses: ismailhakkikinali@gmail.com (İsmail Hakkı Kinalioğlu), coskun@selcuk.edu.tr (Coşkun Kuş)

Received: April 2022 *Accepted:* January 2023

increases exponentially day by day. These developments lead to an increase in the number of data-based studies in football [5, 8, 15, 18, 33]. One of the most common problems in these studies is to predict the results of football matches. Prediction plays a significant role in sports analytics studies such as evaluating performance data, explaining the relationship between the collected data and the match result, determining the team and player strength, and calculating the market values of the players. In addition to its scientific merit, the match result prediction is also crucial for the betting industry. The sports-based betting industry is growing day by day with the effect of the increase in online platforms. The industry size, which was 104.31 billion dollars in 2017, is expected to reach 155.49 billion dollars in 2024 [38]. A successful forecast model is crucial for both bettors and organizers as it will increase profitability. While the prediction is used to determine the odds by organizers, it is also used by players to get more profit from the bet.

There are two types of studies used to predict the result of football matches. The first one is to predict the outcome of the match as home win, draw and away win [6, 24, 26, 34, 35]. The second one is to forecast the number of goals scored by the teams [9, 19, 25, 46]. Some studies related to type one are reviewed as in the following.

Joseph et al. [29] compared some other machine learning methods (MC4, decision tree, naive bayes, and K nearest neighbor) to predict match results. Their average success rate is 59.21%.

Timmaraju et al. [43] predicted the results of the matches played in the 2011-2012 and 2012-2013 seasons in the English Premier League using the polynomial logistic regression and support vector machine. In the model testing phase they developed, it achieved a success of 52.1% for the 2011-2012 season and 48.15% for the 2012-2013 season.

Igiri [27] investigated the performance of the support vector machine in predicting the results of football matches. He created a dataset that covered 38 features for each match. The success rate (8/15) achieved as a result of the test operation is 53.3%.

Amadin and Obi [4] predicted match results using the adaptive-network based fuzzy inference systems (ANFIS) approach. The datasets contain five attributes. These are the result of the last two matches of teams, their places on the scoreboard, the popularity of teams, and the home team's advantage. As a result of the tests they carried out, they reached the success rate of 71% by predicting five of the seven matches correctly.

Robertson et al. [39] measured the ability of Australian football team performance indicators to explain the outcome of the match. They measured the relationship of the attributes with each other by Spearmen's correlation coefficient. Using a one-way analysis of variance, they observed the change of match result (win, loss) according to the attributes. They developed logistic regression models using attributes that showed a statistically significant relationship with the match result. They tested the models they developed with the data of matches whose results are known, and they achieved prediction successes of 87.1% and 85.8%.

Martins et al. [34] proposed a new approach that predicts the results of the matches using the live data obtained by the scouts during the matches played in England, Spain, and Brazil leagues. In order to increase the prediction success, they handled the dependent variable in types consisting of 2 classes. They developed a multi-term classification method for the prediction process and compared this method with some machine learning algorithms. The method they developed is reached 96% prediction success. This rate is higher in all the machine learning methods they use to compare.

Baboota and Kaur [7] predicted the match results using gaussian naive bayes, random forest, support vector machine, and gradient boosting methods. The dataset consists of the matches played in the English Premier League 2014-2015 and 2015-2016 seasons. There are 33 attributes in the dataset that provide information about the past performance of the teams. As a result of tests, they reached the most successful prediction rate, with 57% with the gradient boosting method.

Rahman [37] predicted the group stage matches in the 2018 World Cup. The dataset is prepared by evaluating the performances of the teams in international matches between 1872-2018. He proposes a deep neural network model using LSTM and reaches 63.3% prediction success.

The main purpose of the study is to predict the outcome of a football game with the highest possible success rate. In this context, we offer a tool to collect the football data and to develop some hybrid prediction methods. The data used in similar studies generally consists of a few attributes that provide information about team performance and strength. In this study, in addition to the existing attributes that show team performance and the opinions of fans are also included. In order to obtain the attributes that indicate the opinions of the fans, the shares of the fans are collected from the social media platforms, and the sentiment analysis for these shares is carried out using text mining methods. Our data contains 191 attributes that contain information on the previous matches performance of the teams, on team and player values, on betting odds, on injured and suspended players and on fans' opinions. Using this rich data, we use the well-known classifiers to predict the match results along with the K-means and fuzzy C-means cluster analyzes.

The rest of the paper is organized as follows: Section 2 introduces the process of data collection, preparation and reduction. In addition, the proposed hybrid methods are described in detail. In Section 3, new methods and existing classifiers are performed to predict the match results based on the prepared data, and the results are compared and discussed. Some concluding remarks and suggestions are provided in Section 4.

2 Materials and Methods

2.1 Data Collection Process

The scope of the data is 6396 football matches played in the european leagues in the seasons of 2014-15, 2015-16, 2016-17. The top-level football leagues of England, Spain, Germany, Italy, France, and Turkey are included in this scope. A large number of qualitative and quantitative information is obtained from various web platforms for these football matches. These web platforms are whoscored.com, mackolik.com, transfermarkt.com, injuriesandsuspensions.com, and twitter.com. Data collection tools are developed to collect and archive match information from these platforms.

First, a database table named *"fixture"* is created, which includes the team names, date, and score information of 6396 football matches. Then, web scraping is done for each match in the fixture table. The information obtained as a result of this process is edited and recorded in the database tables *"match-day-info"* and *"match-stats"*. The match-day-info table contains information that can be obtained before playing the match for each match. In the match-stats table, there are the performance statistics of the teams in match.

2.2 Dataset Structure and Preparation Process

2.2.1 Structure of the Dataset

After the data collection process, some filtering, querying, and calculations are performed on the data in the database tables. These procedures are carried out using tools developed for the study. After completing the procedures, alternative datasets are obtained to be used in the study.

The attributes in these datasets are divided into three parts, as given in Fig. 1. These; matchday data, performance statistics from previous matches and are the produced attributes. The result of the match used as the target variable is in three different types. These are MR-1, MR-2, and MR-3.

MR-1 (H, D, A): The target variable has three class labels. These are "Home - 1", "Draw - 0" and "Away - 2". MR-2 (HD, A): Home and Draw labels are combined, and the number of class labels of the target variable is reduced to two. These are "Home or Draw - 1" and "Away - 0". MR-3 (H, DA): In the last case, "Draw" and "Away" classes are combined. In this case, the target variable has two class labels. These are "Home - 1", "Draw or Away - 0".

- **Matchday Data**

Matchday data consists of 58 attributes and is divided into three parts: these, bet odds, squad status, and fan opinions.

Betting odds consist of 36 attributes in different types of bets determined by the Spor Toto Organization. These features obtained from the website "mackolik.com" are presented in Table 1.

Team line-up status on match day consists of 16 attributes. These attributes include information about team squad, player market values, injured, and suspensions. These features obtained from the websites "mackolik.com", "whoscored.com", "transfermarkt.com" and "injuriesandsuspensions.com" are shown in Table 2.

Fan opinions data consists of two parts. The first piece is the survey results collected on various websites and fan forums. The second piece is the data obtained as a result of the sentiment analysis performed on the posts about the matches on social media platforms. Fan surveys are posted on the "mackolik.com" and "whoscored.com" websites and are active until the match time. Tens of thousands of fans vote in these surveys. Social media data are opinions shared via Twitter.com, starting from 48 hours before the match time until the match time. Sentiment analysis is carried out to make sense of these opinions and add them to the dataset. As a result of the analysis, the tweets are divided into three groups as positive, negative, and neutral.

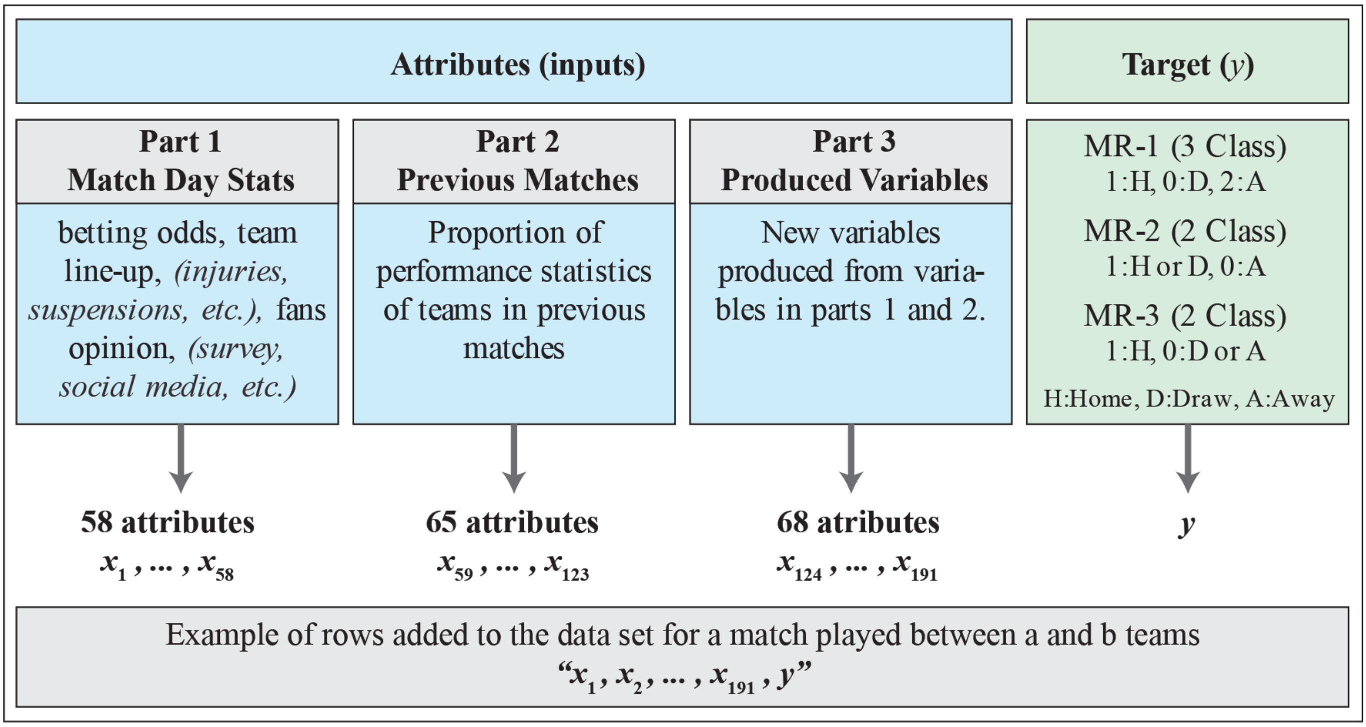

Figure 1: Structure of the dataset

Table 1: Betting odds

Attributes name			
Home team win	Under 1.5 goals	Half time - draw	Half time / full time – H/H
Draw	Over 1.5 goals	Half time - away team win	Half time / full time – H/D
Away team win	Under 2.5 goals	Home or Draw	Half time / full time – H/A
Handicap team	Over 2.5 goals	Home or Away	Half time / full time – D/H
Handicap home team win	Under 3.5 goals	Draw or Away	Half time / full time – D/D
Handicap draw	Over 3.5 goals	0-1 goals	Half time / full time – D/A
Handicap away team win	Both teams score	2-3 Goals	Half time / full time – A/H
Half time under 1.5 goals	None teams score	4-6 Goals	Half time / full time – A/D
Half time over 1.5 goals	Half time - home team win	+7 Goals	Half time / full time – A/A

In the sentiment analysis, 3000 positive, 3000 negative, and 3000 neutral tweets are marked by the authors. The most frequently used words are determined by examining these marked tweets. By making some additions to these words, three different dictionaries, namely "positive", "negative" and "neutral", are created. Afterward, all tweets are analyzed according to these dictionaries.

The stages of collecting and analyzing the tweets for the match between A and B teams are shown below.

Step 1. Up to five names and nicknames that characterize the teams are determined to be used as keywords in searches. Let and be the set of word groups that describe the teams.

$$W_A = \{W_A^1, W_A^2, \ldots, W_A^{n_A}\}, \qquad W_B = \{W_B^1, W_B^2, \ldots, W_B^{n_B}\},$$

where n_A and n_B denote the number of nicknames of team A and team B, respectively.

Step 2. Twit scraping is performed according to all elements of W_A and W_B clusters by using applications developed for the study. This process is done for 48 hours retrospectively from the match time.

Step 3. As a result of twit scraping, two data files are created with shares of A and B teams. Then spam, advertisements, etc. cleaning operations are performed.

Step 4. Classification is performed.

Step 5. Positive, negative, and neutral tweets count are determined for A and B teams.

Table 2: Team line-up (injuries, suspensions, etc)

Attributes name			
Home's 11 market value	Home's 11 average performance rating	Home's injuries, suspensions players playing time	Home's 11 contribution to the score
Away's 11 market value	Away's 11 average performance rating	Away's injuries, suspensions players playing time	Away's 11 contribution to the score
Home's 11 playing times	Home's injuries, suspensions players performance rating	Home's injuries, suspensions players contribution to the score	Home's injuries, suspensions players market value
Away's 11 playing times	Away's injuries, suspensions players performance rating	Away's injuries, suspensions players contribution to the score	Away's injuries, suspensions players market value

Step 6. The three attributes that will be added to the dataset for this match are calculated as follows.

$$Number\,of\,home\,win\,tweets = A_{\text{win\_tweets}} = A_{\text{the\_number\_of\_positive\_tweets}} + B_{\text{the\_number\_of\_negative\_tweets}}$$
$$Number\,of\,away\,win\,tweets = B_{\text{win\_tweets}} = B_{\text{the\_number\_of\_positive\_tweets}} + A_{\text{the\_number\_of\_negative\_tweets}}$$
$$Draw\_Tweets = A_{\text{the\_number\_of\_neutral\_tweets}} + B_{\text{the\_number\_of\_neutral\_tweets}}$$

Table 3: Fan opinion attributes

Attributes name	
Home win voting rate	The Number of home win tweets
Draw voting rate	The Number of away win tweets
Away win voting rate	The Number of draw tweets

- **Historical Performance Statistics from Previous Matches**

These statistics are obtained from the website "whoscored.com". The dataset contains 65 attributes that provide information about performance statistics. These attributes are the average of the teams' statistics in previous games. The number of matches to be included in the average is optional. Various tries are carried out throughout the study. After the average calculation, a value vector of 1x65 dimension is obtained for both teams. These vectors show the field performance statistics of the teams. Then these vectors are compared. After comparison, two vectors are combined, and 65 attributes are added to the dataset. The attributes showing the performances of the teams in previous matches are given in Table 4 divided into general, pass, and shot categories.

Table 4: Performance statistics attributes

General Performance		Shots Performance		Pass Performance	
Sub Cat.	**Attribute Name**	**Sub Cat.**	**Attribute Name**	**Sub Cat.**	**Attribute Name**
Performance	Rating	Number	Total Shots	Number	Pass success %
	Possession %	Results	Goals		Total passes
	Touches		Woodworks		Accurate passes
Dribbles	Dribbles attempted		Shots on target	Pass type	Key passes
	Dribbles won		Shots off target		Cross
	Dribble success		Blocked		Freekick
Tackles	Successful tackles		Own		Through ball
	Tackles attempted	Zones	6-yard box		Throw-in
	Was dribbled		Penalty Area	Length	Long
	Tackle success		Outside of Penalty Area		Short
	Clearances	Situation	Open play	Height	Chipped
	Interceptions		Fastbreak		Ground
Aerials	Aerials win		Set pieces	Body parts	Head
	Aerials win %		Penalty		Feet
	Offensive aerials	Body parts	Right foot	Direction	Forward
	Defensive aerials		Left foot		Backward
Corners	Corners		Head		Left
	Corners accuracy		Other body parts		Right
Goalkeeper	Saves			Target zone	Defensive third
	Claims				Mid third
	Punches				Final third
Others	Offside				
	Faul				
	Loss of possession				
	Errors				
	Blocks				

- **Produced Attributes**

This section of the database contains 68 new attributes produced using matchday data and performance statistics. Expert opinion is used in determining these attributes. In the calculation phase of these attributes, the tools developed within the scope of the study are used. Produced attributes are presented in Table 5.

Table 5: Produced attributes

Attributes name		
Home team total wins	Away team total wins	Home team number of one different wins
Home team total draws	Away team total draws	Home team number of handicap wins
Home team total losses	Away team total losses	Home team number of one different losses
Home team wins of last M matches*	Away team wins of last M matches*	Home team number of handicap losses
Home team draws of last M matches*	Away team draws of last M matches*	Away team number of one different wins
Home team losses of last M matches*	Away team losses of last M matches*	Away team number of handicap wins
Home team total points	Away team total points	Away team number of one different losses
Home team points of last M matches*	Away team points of last M matches*	Away team number of handicap losses
Home team win percentage	Away team win percentage	Home team previous 1 match result
Home team draw percentage	Away team draw percentage	Home team previous 2 match result
Home team loss percentage	Away team loss percentage	Home team previous 3 match result
Home team wins at home	Away team wins at away	Home team previous 4 match result
Home team win percentage at home	Away team win percentage at away	Home team previous 5 match result
Home team draws at home	Away team draws at away	Away team previous 1 match result
Home team draw percentage at home	Away team draw percentage at away	Away team previous 2 match result
Home team losses at home	Away team losses at away	Away team previous 3 match result
Home team loss percentage at home	Away team loss percentage at away	Away team previous 4 match result
Distance of home team to last win	Distance of away team to last win	Away team previous 5 match result
Distance of home team to last draw	Distance of away team to last draw	Total Points Difference
Distance of home team to last loss	Distance of away team to last loss	Is it the target for the home team?
Home team longest win streak	Away team longest win streak	Is it the target for the away team?
Home team longest draw streak	Away team longest draw streak	Market Value Ratio of Teams 11
Home team longest loss streak	Away team longest loss streak	

* M: The limit of number of matches set in the dataset preparation software

2.2.2 Dataset Preparation Software

Dataset preparation software is developed to prepare alternative datasets to be used in the study. For this, the software performs a number of operations on data collected from various sources. These operations are carried out in two stages. In the first stage, a series of calculations are performed on the database tables. In the second stage, datasets are prepared by using the values obtained as a result of calculations. Alternative datasets can be prepared by changing various parameters in both phases. These parameters are used to calculate the performance values of teams and to export the dataset. They are divided into five titles according to their location in the software interface. The interface of the software is given by Fig. 2.

For the calculation phase,

- **Interval of Previous Matches:** it is determined which matches will be taken into account when calculating the past performances of the teams.
- **Team Data Comparison:** In this field, when comparing the performance values of the teams, which method is used is determined.

For the preparation phase of the dataset,

- **Dataset Contents:** In this field, the attributes to be included in the dataset are determined.
- **Train and Test File Options:** In this field, the structure of the training and test datasets are determined, and a filter can be added to the dataset. According to Fig. 3., there are two filter options as league and season.
- **File Saving Options:** In this field, the type and structure of the files to be exported are determined.

The calculation phase is carried out in two steps. In the first step, the performance statistics of the teams are calculated and compared. All details of this step are described in the heading "Calculation of the Performance Values of the Teams". In the second step of the calculation phase, the operations described under the heading "produced attributes" are performed. In the dataset preparation phase, datasets are prepared according to the determined parameters. These datasets can be saved in "xlsx" or "csv" formats for use in the study.

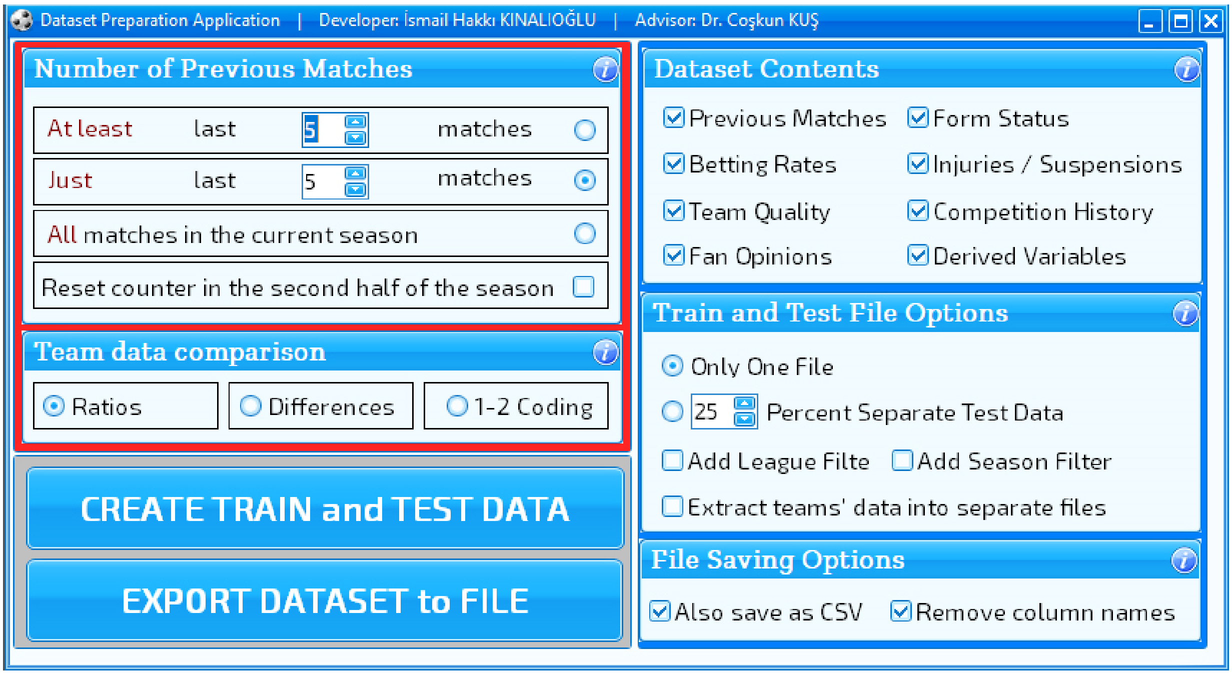

Figure 2: Dataset preparation software

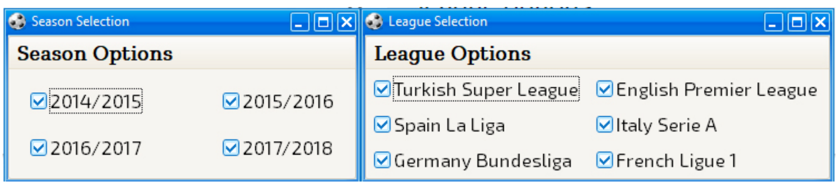

Figure 3: Season and league filter

2.2.3 Calculating Performance Values of Teams

This step is performed for all matches in the fixture table. Performance values are the average of the statistics the teams obtained in a certain number of matches played in the previous weeks. In order to calculate these values, it is necessary to decide which matches played by the teams in the previous weeks to consider. This preference is selected in the "Interval of Previous Matches" section of the dataset preparation software. This part of the software is shown in Fig. 4.

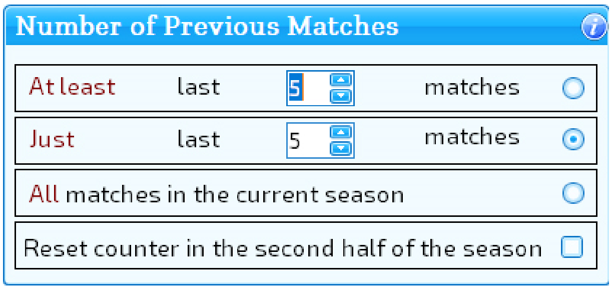

Figure 4: Interval of previous matches

These attributes, which provide information about the performance statistics obtained in previous matches, can be calculated using three different methods. These are "At least the last M match", "Only the last M match" and "All Matches in the Current Season". The variables used in these methods are as follows.

A,B	:	teams names,	M	:	limit of number matches,
i	:	variable index (1,2,...,65),	j	:	week index,
X_A^{ij}	:	Team A's i variable in week j,	X_B^{ij}	:	Team B's i variable in week j,
M_A	:	number of matches of Team A,	M_B	:	number of matches of Team B,

If "at least the last M match" option is selected;

Step 1: The match-stats table is filtered separately for both A and B teams. In this way, the number and information of the matches played by both teams until that date are accessed.

Step 2: If the number of matches of both teams is equal or more than M limit, continue with Step 3. Otherwise, no calculation is made and return to Step 1 with new match information from the fixture table.

Step 3: The average of the performance statistics of A and B teams in previous matches is calculated. As a result of the averaging process, it creates a value vector in the dimension of for both teams. The vector of values is given by

$$\mathbf{X}_A = \frac{1}{M}\left(\sum_{j=(M_A-M+1)}^{M_A} X_A^{1j}, \sum_{j=(M_A-M+1)}^{M_A} X_A^{2j}, \ldots, \sum_{j=(M_A-M+1)}^{M_A} X_A^{65j}\right)$$

and

$$\mathbf{X}_B = \frac{1}{M}\left(\sum_{j=(M_B-M+1)}^{M_B} X_B^{1j}, \sum_{j=(M_B-M+1)}^{M_B} X_B^{2j}, \ldots, \sum_{j=(M_B-M+1)}^{M_B} X_B^{65j}\right).$$

If "all matches in the current season" option is selected;

Step 1: The match-stats table is filtered separately for both A and B teams. In this way, the number and information of the matches played by both teams until that date are accessed.

Step 2: The average of the performance statistics of A and B teams in previous matches is calculated. As a result of the averaging process, it creates one each value vector in the dimension of for both teams. These value vectors are given by

$$\mathbf{X}_A = \frac{1}{M_A}\left(\sum_{j=1}^{M_A} X_A^{1j}, \sum_{j=1}^{M_A} X_A^{2j}, \ldots, \sum_{j=1}^{M_A} X_A^{65j}\right)$$

and

$$\mathbf{X}_B = \frac{1}{M_B}\left(\sum_{j=1}^{M_B} X_B^{1j}, \sum_{j=1}^{M_B} X_B^{2j}, \ldots, \sum_{j=1}^{M_B} X_B^{65j}\right).$$

In this section, there is an option called "reset counter in the second half of the season" that can be used with other three options. This option makes the second half of the season independent from the first half. The purpose of this option is to consider effect of the change in the teams' strength in the middle of the season. As a result of these calculations, dimensional value vectors are formed, which show the averages of the performance values obtained for both teams in the previous weeks. These vectors are then compared according to the preference in the "Tool Data Comparison" section is presented in Fig. 5. The purpose of this comparison is to convert two value vectors into one vector with the same size. In this section, there are the "differences", "rates" and "1-2 coding" options detailed below.

Figure 5: Team data comparison

The value vectors showing the average performances of A and B teams in previous matches are given by

$$\mathbf{X}_A = \left[\overline{X}_A^1, \overline{X}_A^2, \ldots, \overline{X}_A^{65}\right]_{1\times 65} \text{ and } \mathbf{X}_B = \left[\overline{X}_B^1, \overline{X}_B^2, \ldots, \overline{X}_B^{65}\right]_{1\times 65}.$$

If the *"differences"* option is selected, the combined vector $\mathbf{X}_{AB}$ is defined by subtracting the data of the Team B from the data of the Team A, that is.

$$\mathbf{X}_{AB} = (\mathbf{X}_A - \mathbf{X}_B) = \left[\left(\overline{X}_A^1 - \overline{X}_B^1\right), \left(\overline{X}_A^2 - \overline{X}_B^2\right), \ldots, \left(\overline{X}_A^{65} - \overline{X}_B^{65}\right)\right]_{1\times 65}$$

If the *"ratios"* option is selected, the combined vector $\mathbf{X}_{AB}$ is defined by dividing the data of Team A by the data of Team B. Hence, $\mathbf{X}_{AB}$ is given by

$$\mathbf{X}_{AB} = \frac{\mathbf{X}_A}{\mathbf{X}_B} = \left[\frac{\overline{X}_A^1}{\overline{X}_B^1}, \frac{\overline{X}_A^2}{\overline{X}_B^2}, \ldots, \frac{\overline{X}_A^{65}}{\overline{X}_B^{65}}\right]_{1\times 65}.$$

When the ratios option is selected, 65 attributes of the teams are mutually proportioned. During this process, it is checked whether the values of the compared attributes of the A and B teams differ from 0. If any of these values is equal to 0, the comparison process in the related attribute is performed with the "differences" option instead of the "ratios" option.

If *"1-2 coding option"* is selected, each value in the value vectors of teams A and B is compared. It is coded as 1 if the value of Team A is more, 2 if the value of Team B is more, and 0 if they are equal. Then the combined $\mathbf{X}_{AB}$ can be stated by

$$\mathbf{X}_{AB} = [AB_1, \ldots AB_{65}] ,$$

where

$$AB_i = \begin{cases} 1 & , \quad \overline{X}^i_A > \overline{X}^i_B \\ 0 & , \quad \overline{X}^i_A = \overline{X}^i_B \\ 2 & , \quad \overline{X}^i_A < \overline{X}^i_B \end{cases} ,$$

for i=1,2,...,65.

2.2.4 Dataset Preparation Steps

While preparing the study dataset, the following steps are performed for each of the 6396 matches in the fixture table, respectively.

Assume that the first match from the fixture table is played between teams A and B on the date T_1 Let the vector $\mathbf{V}_{AB}$ be included into the dataset as a result of processing the information of this match.

Step 1. If A and B teams played less than the M limit before the T_1 date in the current season, return to Step 1 and a new match information will come. If they have made at least M matches, continue with Step 2.

Step 2. According to the match from the *fixture* table, the *match-day-info* table is filtered, and *"match day data"* is obtained for the match between the A and B teams. This data consists of 58 attributes, as explained in the "matchday data" heading. With the addition of the match result as the target variable, the first version of the $\mathbf{V}_{AB}$ vector is given by

$$\mathbf{V}_{AB}{}^{(1)} = \mathbf{V}_{AB_{matchday}},$$

where

$$\mathbf{V}_{AB_{matchday}} = [x_{AB_1}, x_{AB_2}, \ldots, x_{AB_{58}}]_{1\times 58}.$$

Step 3. In this step, 65 new attributes are added to the dataset. These attributes are obtained by comparing the average performance values of A and B teams in the previous games. Detailed information about this process is given under the heading *"Calculation of the performance values of the teams"*. Thus, the $\mathbf{V}_{AB}$ vector has taken its new form:

$$\mathbf{V}_{AB}{}^{(2)} = \left[\mathbf{V}_{AB_{match\_day}}, \mathbf{V}_{AB_{previous\_matches}}\right] = [x_{AB_1}, x_{AB_2}, \ldots, x_{AB_{123}}]_{1\times 123},$$

where

$$\mathbf{V}_{AB_{previous\_matches}} = [x_{AB_{59}}, x_{AB_{60}}, \ldots, x_{AB_{123}}]_{1\times 65}.$$

Step 4. In this step, 68 new attributes are added, which are called produced attributes. Detailed information on obtaining these attributes is given in the heading *"produced attributes"*. With the addition of these attributes, the $\mathbf{V}_{AB}$ vector reaches its final form with:

$$\mathbf{V}_{AB}{}^{(3)} = \left[\mathbf{V}_{AB}{}^{(2)}, \mathbf{V}_{AB_{produced}}, V_{AB_{result}}\right] = [x_{AB_1}, x_{AB_2}, \ldots, x_{AB_{191}}, y]_{1\times 192},$$

where

$$\mathbf{V}_{AB_{produced}} = [x_{AB_{124}}, x_{AB_{125}}, \ldots, x_{AB_{191}}]_{1\times 68}$$

and

$$V_{AB_{result}} = y.$$

Assume that N of the 6396 matches in the *fixture* table reached Step 2. In this case, an sized matrix is created which can be used in study. This matrix is recorded in a database table called *"train"* for use in preparing alternative datasets to be used in the study.

$$train = \begin{bmatrix} x_{1_1} & x_{1_2} & \ldots & x_{1_{191}} & y_1 \\ x_{2_1} & x_{2_2} & \ldots & x_{2_{191}} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N_1} & x_{N_2} & \ldots & x_{N_{191}} & y_N \end{bmatrix}_{N\times 192}$$

2.3 Feature Selection and Feature Extraction

In this study, information gain ratio and correlation-based methods are used for feature selection. These methods are frequently used in the literature [21, 32]. Principal components analysis are used for feature extraction. Principal components analysis is used in different forms in the literature [47, 45, 48, 10].

Tests are carried out with datasets obtained as a result of feature selection and feature extraction. In these tests, better performance values are achieved with the datasets prepared by feature selection compared to the datasets prepared by feature extraction. For this reason, feature selection are used to reduce the dataset while performing the applications of the study. The attributes in the lower dimensional dataset obtained by feature selection are given in Table 6.

Table 6: Reduced dataset

x_1	home's 11 market value	x_{26}	blocked shots
x_2	away's 11 market value	x_{27}	6-yard box shots
x_3	home win betting odds	x_{28}	open play shots
x_4	away win betting odds	x_{29}	pass success %
x_5	half time home team win betting odds	x_{30}	key passes
x_6	half time away team win betting odds	x_{31}	through ball
x_7	half time / full time: 1/1 betting odds	x_{32}	home team win percentage
x_8	half time / full time: 1/2 betting odds	x_{33}	home team draw percentage
x_9	half time / full time: x/1 betting odds	x_{34}	away team win percentage
x_{10}	half time / full time: 2/2 betting odds	x_{35}	home team win percentage at home
x_{11}	home win voting rate	x_{36}	total Points Difference
x_{12}	number of home win tweets	x_{37}	is it the target for the home team?
x_{13}	number of away win tweets	x_{38}	distance of away team to last loss
x_{14}	home's injuries, suspensions players market value	x_{39}	market value ratio of teams 11
x_{15}	away's injuries, suspensions players market value	x_{40}	home team wins of last M matches
x_{16}	home's injuries, suspensions players performance rating	x_{41}	home team wins at home
x_{17}	away's injuries, suspensions players performance rating	x_{42}	away handicap wins
x_{18}	home's 11 contribution to the score	x_{43}	distance to draw for home team
x_{19}	rating	x_{44}	away team number of handicap wins
x_{20}	possession %	x_{45}	distance of home team to last loss
x_{21}	successful passes	x_{46}	away team total draws
x_{22}	aerials win %	x_{47}	away team longest draw streak
x_{23}	corners	x_{48}	home team number of one different loses
x_{24}	saves	x_{49}	distance of home team to last win
x_{25}	shots on target	x_{50}	home team longest loss streak

2.4 Classification, Clustering Methods and Structure of Proposed Hybrid Methods

The prediction of football match result is tackled as a classification problem. Classification problems are one of the most frequently studied topics in data mining and machine learning and is studied by researchers from many different disciplines over the past few decades [3]. In addition to classification, cluster analysis is also used in the proposed hybrid methods. In the literature, there are many classification and clustering methods developed by making use of various learning algorithms [1, 42]. In this study, Artificial neural network (ANN), K nearest neighbor (KNN), multinomial logistic regression (MLR), naive bayes (NB), random forest (RF), support vector machine (SVM) are used for classification while K means clustering (KM) and fuzzy C-means clustering (FCM) are used for clustering.

Today, there are many platforms used to implement these methods. Rich libraries of programming languages such as Python and R are good alternatives to implement these methods. There are many IDE developed to use these libraries. In this study, Rstudio is used for R libraries, and Pycharm are used for python libraries. R libraries are used for coding machine learning methods and python libraries are used for data collection. On the other hand, there are many software that can use machine learning methods. The most popular among these are software like Matlab, SPSS, Weka. WEKA (Waikato Environment for Knowledge Analysis) software is mostly used in the applications of this study due to its ease of use, speed and open-source structure. WEKA is an open-source machine learning software developed by the University of Waikato [20].

Within the scope of the study, hybrid methods are developed in order to increase the prediction success achieved by classification methods. In these methods, "K-Means Clustering (KM)" and "Fuzzy C Means Clustering (FCM)" clustering methods and classification methods are used together.

While creating hybrid methods with "K-Means Clustering" method, cluster analysis is performed on the dataset for different K values. After the cluster analysis, a new column is added to the dataset. This column shows the cluster labels that include each observation. Various datasets are prepared for different K values, and classification is made on these datasets. The K value with the highest prediction success is taken into consideration. The structure of hybrid methods prepared with K-Means Clustering is presented in Fig. 6.

Hybrid classification methods created by K-Means clustering;

- K-Means – Artificial Neural Network Hybrid Classifier (KM-ANN),
- K-Means – K-Nearest Neighbor Hybrid Classifier (KM-KNN),
- K-Means – Multinomial Logistic Regression Hybrid Classifier (KM-MLR),
- K-Means – Naive Bayes Hybrid Classifier (KM-NB),
- K-Means – Random Forest Hybrid Classifier (KM-RF),
- K-Means – Support Vector Machine Hybrid Classifier (KM-SVM),

it is named as.

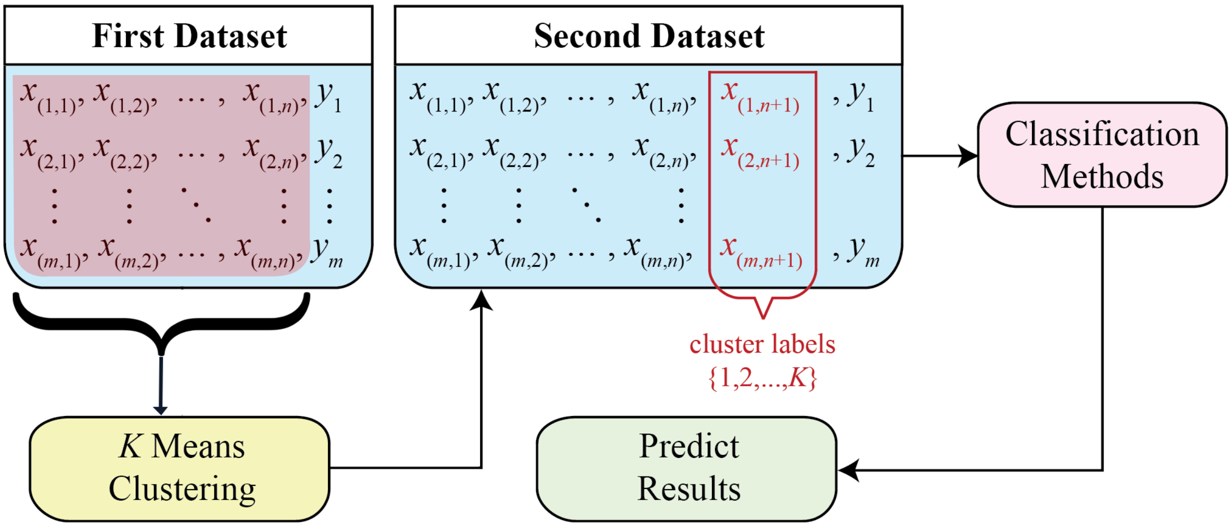

Figure 6: Structure of hybrid methods created by K-means clustering

In hybrid methods created with the Fuzzy C Means Clustering method, in addition to cluster labels, membership degrees are added to the dataset. Tries are made for different C values. After each trial, $(C + 1)$ new columns are added to the dataset. C columns show membership degrees, and one column shows the cluster labels of the samples. The structure of hybrid methods created using Fuzzy C Means Clustering is shown in Fig. 7.

Hybrid classification methods created by fuzzy C-Means clustering;

- Fuzzy C-Means – Artificial Neural Network Hybrid Classifier (FCM-ANN),
- Fuzzy C-Means – K-Nearest Neighbor Hybrid Classifier (FCM-KNN),
- Fuzzy C-Means – Multinomial Logistic Regression Hybrid Classifier (FCM-MLR),
- Fuzzy C-Means – Naive Bayes Hybrid Classifier (FCM-NB),
- Fuzzy C-Means – Random Forest Hybrid Classifier (FCM-RF),
- Fuzzy C-Means – Support Vector Machine Hybrid Classifier (FCM-SVM),

it is named as.

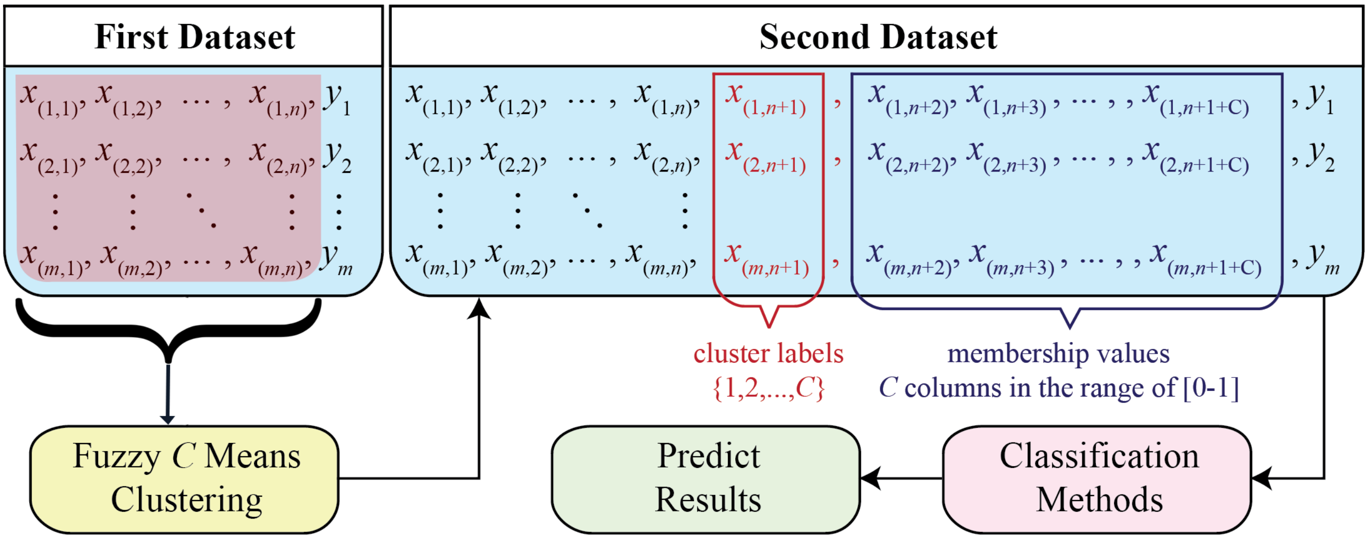

Figure 7: Structure of hybrid methods created by fuzzy C-means clustering

2.5 Performance Measures

The successes of the models used in machine learning studies are assessed according to various performance measures [13, 41]. These performance measures are calculated using the confusion matrix. Confusion matrices summarize the model's performance based on a number of test values [44]. These test values are defined according to the number of true and false classifications as "true positive (TP), false positive (FP), true negative (TN), and false negative (FN)". In many studies, the dependent variable has two class labels, positive and negative. The structure of the confusion matrix for these studies can be given as in Table 7.

Table 7: Confusion matrix structure

		Classification	
		1	0
Actual	1	TP (True Positive)	FN (False Negative) (Type 2 Error)
	0	FP (False Positive) (Type 1 Error)	TN (True Negative)

In this matrix;

TP: Number of observations classified as 1 when actually 1

TN: Number of observations classified as 0 when actually 0

FP: Number of observations classified as 1 even though it is actually 0

FN: Number of observations classified as 0 although actually 1

The performance measures calculated using the values in the confusion matrix are briefly described below.

Observed Accuracy (ACC) refers to the ratio of correctly classified observations. Expected accuracy *(EACC)* refers to the ratio of observations expected to be classified correctly. They are given, respectively, by

$$ACC = \frac{TP + TN}{TN + FN + FP + TP}$$

$$EACC = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TN + FN + FP + TP)^2}$$

Sensitivity (TPR, TP Rate, Recall) is the correct classification rate of positive observations. Specificity (TNR, TN Rate) is the rate of correct classification of negative observations. FP Rate (FPR) is the rate of false classified positive observations. FN Rate (FNR) is the ratio of misclassified negative observations.

$$TPR = \frac{TP}{TP+FN}\ ,\ TNR = \frac{TN}{TN+FP}\ ,\ FPR = \frac{FP}{FP+TN},\ FNR = \frac{FN}{TP+FN}$$

Precision (PPV) is the ratio of correctly classified positive observations to all positively classified observations. The negative prediction rate (NPV) is the ratio of correctly classified negative observations to all negative classified negative observations. F-Measure (FM) is the harmonic average of sensitivity and precision values. Kappa Statistics (KS) describes the relationship between expected and observed classification success to check whether the correctly classified samples are random. KS takes values between 0 and 1. The classifier's success increases as the KS gets closer to 1. As KS gets closer to 0, the predictions do not differ from random predictions.

$$PPV = \frac{TP}{TP+FP},\ NPV = \frac{TN}{TN+FN},$$

$$F-M = \frac{2}{\frac{1}{TPR}+\frac{1}{PPV}} = 2 \times \frac{PPV \times TPR}{PPV+TPR},\ KS = \frac{ACC-EACC}{1-EACC}.$$

Receiver Operating Characteristic (ROC) curves are drawn according to the updated TPR (sensitivity) and (1-specificity) FPR values after each observation classified in the test phase. As the area under the Roc curve gets closer to 1, the classifier's success increases. This value is referred to as AUC (area under the curve) in the literature.

In this study, the dependent variable has three class labels in the case of MR-1 and two in the case of MR-2 and MR-3. Generalized confusion matrices for MR-1, MR-2 and MR-3 are presented in Table 8-11.

Table 8: Structure of the general confusion matrix of the study

			Classification		
			Draw 0	Home 1	Away 2
Actual	Draw	0	K_{11}	K_{12}	K_{13}
	Home	1	K_{21}	K_{22}	K_{23}
	Away	2	K_{31}	K_{32}	K_{33}

Table 9: Generalized confusion matrix structure for draw

		Classification	
		Draw (0)	Home (1) or Away (2)
Actual	Draw (0)	TP=K_{11}	FN=K_{12}+K_{13}
	Home (1) or Away (2)	FP=K_{21}+K_{31}	TN=K_{22}+K_{23}+K_{32}+K_{33}

Table 10: Generalized confusion matrix structure for home

		Classification	
		Home (1)	Draw (0) or Away (2)
Actual	Home (1)	TP=K_{22}	FN=K_{21}+K_{23}
	Draw (0) or Away (2)	FP=K_{12}+K_{32}	TN=K_{11}+K_{13}+K_{31}+K_{33}

3 Results

K-Fold Cross Validation method is used for testing the prepared dataset. In this method, the dataset is divided into K equal parts. Respectively, one of these parts is used as test data and the rest as training data, and K times tests are performed. The average of these tests gives the general performance of the model.

Table 11: Generalized confusion matrix structure for away

		Classification	
		Away (2)	Draw (0) or Home (1)
Actual	Away (2)	TP=K_{33}	FN=K_{31}+K_{32}
	Draw (0) or Home (1)	FP=K_{13}+K_{23}	TN=K_{11}+K_{12}+K_{21}+K_{22}

The first title includes the findings obtained by classification algorithms. In the second title, there are the findings of the hybrid methods obtained by using the classification algorithms and k-means clustering method. The third title includes the findings of the hybrid methods obtained by using the classification algorithms together with the fuzzy c-means clustering method. In the last part, there are the findings showing the change in prediction success according to the content of the dataset. These four titles are divided into subtitles for three types of the target variable.

3.1 Findings of Classification Methods

As a result of the training and testing processes, the parameter values at which the classifiers used reached the best performance measures are determined. These values are given in Table 12. Findings reached by classification methods are shared under separate headings for MR-1, MR-2 and MR-3 cases.

Table 12: MR-1 model parameters

Classification Methods	**Parameters**	**First Dataset**	**Dimension Reduced Dataset**
ANN	Hidden Layers	3	3
	Neurons of Hidden Layers	7/3/11	5/3/09
	Momentum	0.2	0.2
	Learning Rate	0.3	0.5
KNN	K	91	34
	Neighbor Search Algorithm	linear search	linear search
	Distance Function	manhatten distance	manhatten distance
	Distance Weighting Method	1/distance	1/distance
MLR	Conjugate gradient descent	using	using
	Iteration Number	150	150
NB	Kernel Estimator	using	using
RF	Iteration Number	500	250
	Roots	1	1
SVM	c	3	3
	Kernel function	polly kernel	polly kernel
	Calibrator	Random forest	Random forest

The parameters that most affect the model performance in ANN are the hidden layer and number of neurons in hidden layers. In the tries conducted according to these parameters, it is observed that the difference between the prediction successes reached 10%. K value mostly affects model performance in KNN. The distance function is the second. In the tries conducted according to K value and distance function in KNN, it reaches 10% between the prediction successes. The parameter that most affects the model performance in MLR is the number of iterations. The highest prediction success in the tries is reached when the number of iterations is 150. Model performance increases when a kernel estimator is used in NB. In order to achieve the maximum model performance in RF, tries are made in the number of iterations and roots. Experiments are carried out on c value, kernel function, and calibrator parameters used in SVM. The kernel function used is the most significant parameter in prediction success. The difference between the prediction successes in the tries reached 20%.

3.1.1 Findings in the Case of MR-1 (H, D, A)

As a result of the applications performed for MR-1; confusion matrices are shared in Table 13, performance measures in Table 14, AUC values in Table 15, ROC curves in Fig. 8-10.

Table 13: Confusion matrices reached by classification methods for MR-1

		ANN			KNN			MLR			NB			RF			SVM		
		0	**1**	2	**0**	**1**	**2**	**0**	**1**	**2**	**0**	1	**2**	**0**	**1**	**2**	**0**	**1**	**2**
First Dataset	**0**	64	321	197	89	309	184	190	234	158	**232**	183	167	131	270	181	167	258	157
	1	43	790	86	41	779	99	131	694	92	235	593	91	41	**795**	83	99	756	64
	2	33	215	468	40	198	478	122	158	436	155	101	460	55	167	**494**	107	165	444
Dimension Reduced Dataset	**0**	99	275	208	110	273	199	106	271	195	**221**	197	164	145	262	175	65	348	169
	1	72	751	96	56	468	95	76	753	90	207	641	71	58	780	81	21	**823**	75
	2	53	175	**488**	72	157	487	78	156	482	145	111	460	75	165	476	63	206	447

Table 14: Performance measures for MR-1

		First Dataset						Dimension Reduced Dataset					
Class. Method	*Class*	*TPR*	*TNR*	*PPV*	*F-M*	*KS*	*ACC*	*TPR*	*TNR*	*PPV*	*F-M*	*KS*	*ACC*
ANN	0	0.11	0.954	0.457	0.177	0.36	59.63%	0.17	0.924	0.442	0.246	0.37	60.35%
	1	0.68	0.587	0.596	0.704			0.817	0.653	0.625	0.708		
	2	0.654	0.813	0.623	0.638			0.682	0.797	0.616	0.647		
KNN	0	0.153	0.95	0.524	0.237	0.38	60.71%	0.189	0.922	0.462	0.268	0.39	61.57%
	1	0.848	0.609	0.606	0.707			0.836	0.669	0.641	0.726		
	2	0.668	0.813	0.628	0.647			0.68	0.804	0.624	0.651		
MLR	0	0.326	0.845	0.429	0.371	0.37	59.63%	0.182	0.906	0.408	0.252	0.38	60.49%
	1	0.757	0.698	0.64	0.694			0.819	0.663	0.633	0.714		
	2	0.609	0.833	0.636	0.622			0.673	0.81	0.628	0.65		
NB	0	0.399	0.761	0.373	0.385	0.36	57.96%	0.38	0.788	0.386	0.383	0.38	59.63%
	1	0.645	0.781	0.676	0.66			0.697	0.763	0.675	0.686		
	2	0.642	0.828	0.641	0.642			0.642	0.843	0.662	0.652		
RF	0	**0.225**	**0.941**	**0.577**	**0.324**	**0.43**	**64.05%**	**0.249**	**0.919**	**0.522**	**0.337**	**0.42**	**63.19%**
	1	**0.865**	**0.663**	**0.645**	**0.739**			**0.849**	**0.671**	**0.646**	**0.734**		
	2	**0.69**	**0.824**	**0.652**	**0.67**			**0.665**	**0.829**	**0.65**	**0.657**		
SVM	0	0.287	0.874	0.448	0.35	0.4	61.66%	0.112	0.949	0.436	0.178	0.36	60.22%
	1	0.823	0.674	0.641	0.721			0.896	0.573	0.598	0.717		
	2	0.62	0.853	0.668	0.643			0.624	0.837	0.647	0.635		

According to the performance measures of the applications performed with the first dataset, the RF method achieved the highest prediction success (ACC: 64.05%). RF is followed by SVM (ACC: 61.66%), KNN (ACC: 60.71%), MLR (ACC: 59.63%), ANN (ACC: 59.63%) and NB (ACC: 57.96%) methods, respectively. RF has achieved the highest prediction success in the applications where dimension reduced dataset is used (ACC: 63.19%). RF is followed by KNN (ACC: 61.57%), MLR (ACC: 60.49%), ANN (ACC: 60.35%), SVM (ACC: 60.22%) and NB (ACC: 59.63) methods, respectively. KS range from 0.36 to 0.43; this indicates that the classification is approximately 40% better than a random classification. When the ROC curves and the areas under these curves are examined, it is observed that the prediction success is high in matches won by home or away team, and low in the draws.

3.1.2 Findings in the Case of MR-2 (HD, A)

Reached in applications for MR-2, confusion matrices, in Table 16, performance measures in Table 17, information on AUC values in Table 18, ROC curves are shared in Fig. 11.

According to the performance measures of the applications performed with the first dataset, the RF method achieved the highest prediction success (ACC: 79.25%). RF is followed by SVM (ACC: 77.94%), ANN (ACC: 77.45%), KNN (ACC: 77.08%), MLR (ACC: 76.59) and NB (ACC: 75.45%) methods, respectively. RF has achieved the highest prediction success in the applications where dimension reduced dataset is used (ACC: 79.03%). RF is followed by KNN (ACC: 78.21%), ANN (ACC: 77.58%), SVM (ACC: 76.91%), NB (ACC: 76.91%) and MLR (ACC: 76.68) methods, respectively. In the applications performed in the MR-2 case, prediction success of approximately 80% is achieved. KS range from 0.42 to 0.50; this indicates that a 50% better classification success is achieved than a random classification. Areas under ROC curves are higher than MS-1.

3.1.3 Findings in the Case of MR-3 (H, DA)

Reached in applications for MR-2, confusion matrices, in Table 16, performance measures in Table 17, information on AUC values in Table 18, ROC curves are shared in Fig. 11.

According to the performance measures of the applications performed with the first dataset, the RF method achieved the highest prediction success (ACC: 75.50%). RF is followed by SVM (ACC: 73.61%), ANN (ACC: 73.39%), NB (ACC: 73.16%), MLR (ACC: 72.67) and KNN (ACC: 72.12%) methods, respectively. RF has achieved the highest prediction success in applications where dimension reduced dataset is used (ACC: 74.79%). RF is followed by KNN (ACC: 73.93%), NB (ACC: 73.88%), MLR (ACC: 72.76%), ANN (ACC: 72.08%) and SVM (ACC: 71.04) methods, respectively. In the case of MS-3, the prediction successes ranging from 71.04% to 75.50% are achieved. Kappa statistics range from 0.37 to 0.51; this indicates that a 51% better classification success is achieved than a random classification. The areas under ROC curves are observed with values varying around 0.7.

Table 15: AUC values and confidence intervals for MR-1

	Class. Method	for home team win AUC	Std. Error	95% CI	for away team win AUC	Std. Error	95% CI	for draw AUC	Std. Error	95% CI
First Dataset	ANN	0.726	0.011	0.704 - 0.748	0.687	0.0096	0.668 - 0.706	0.521	0.0126	0.496 - 0.546
	KNN	0.731	0.011	0.709 - 0.753	0.668	0.01	0.648 - 0.688	0.551	0.0128	0.526 - 0.576
	MLR	0.692	0.0125	0.668 - 0.717	0.609	0.0111	0.587 - 0.631	0.58	0.0135	0.554 - 0.606
	NB	0.701	0.0126	0.676 - 0.726	0.624	0.0113	0.602 - 0.646	0.572	0.0135	0.546 - 0.598
	RF	0.743	0.0113	0.721 - 0.765	0.66	0.0104	0.640 - 0.680	0.574	0.0133	0.548 - 0.600
	SVM	0.706	0.0125	0.682 - 0.731	0.627	0.0108	0.606 - 0.648	0.573	0.0134	0.547 - 0.599
Dimension Reduced Dataset	ANN	0.728	0.0113	0.706 - 0.750	0.682	0.0101	0.662 - 0.702	0.53	0.0132	0.504 - 0.556
	KNN	0.719	0.0118	0.696 - 0.742	0.664	0.0104	0.644 - 0.684	0.542	0.0132	0.516 - 0.568
	MLR	0.717	0.0119	0.694 - 0.740	0.668	0.0104	0.648 - 0.688	0.535	0.0131	0.509 - 0.561
	NB	0.709	0.0126	0.684 - 0.734	0.63	0.0112	0.608 - 0.652	0.572	0.0136	0.545 - 0.599
	RF	0.726	0.0118	0.703 - 0.749	0.645	0.0106	0.624 - 0.666	0.574	0.0134	0.548 - 0.600
	SVM	0.705	0.012	0.681 - 0.729	0.657	0.0098	0.638 - 0.676	0.534	0.0123	0.510 - 0.558

(a) the first dataset, for home win **(b)** dimension red. dataset, for home win

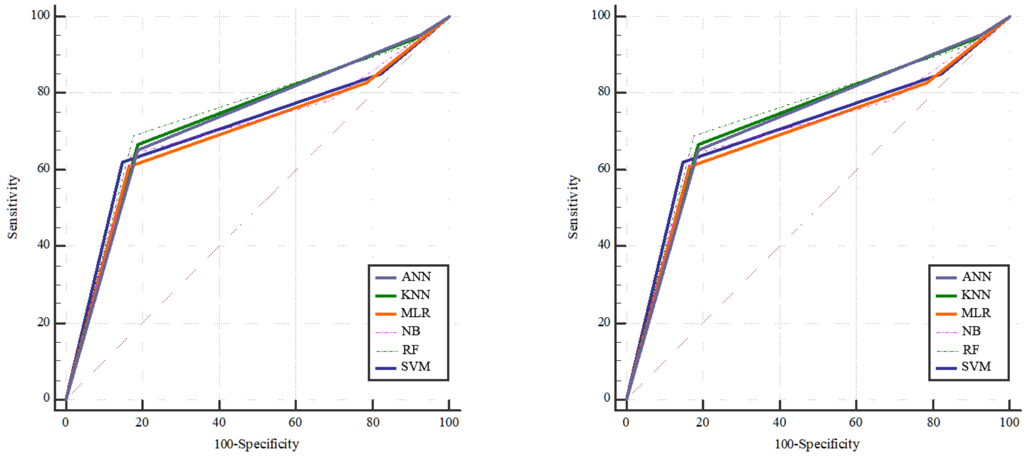

Figure 8: ROC curves according to “home team win” for MR-1

3.2 Findings of Hybrid Methods Created by K-Means Clustering

K-means cluster analysis is performed each time by changing the value of K from 2 to 15. As a result of these analyzes, 14 new attributes are created. These attributes show the class labels to which the observations belong. Afterward, these attributes are added to the data one by one, and 14 alternative datasets are created. The optimal K value is determined for each method by comparing the findings obtained with these datasets. Findings obtained with optimal K values are presented in Table 22-24 for MR-1, Table 25-27 for MR-2, and Table 28-30 for MR-3.

According to the findings of hybrid methods developed by using KM clustering for MR-1, the highest prediction success is achieved with KM-RF method (ACC: 65.46%). When the effect of hybrid methods on prediction success in MR-1 is examined, it is seen that the most significant change is in KM-NB. It is seen that the prediction success achieved with KM-NB is 3.80% higher than the prediction success achieved with NB.

In the tests performed for the case of MR-2 using hybrid methods developed with KM clustering, the highest prediction success is achieved with KM-SVM method (ACC: 81.49%). When the effect of hybrid methods on prediction success in MR-2 is examined, it is seen that the most significant change is in KM-SVM. It is seen that the prediction success achieved with KM-SVM is 3.55% higher than the prediction success achieved with SVM.

According to the findings of hybrid methods developed by using KM clustering for MR-3, the highest prediction successes is achieved with KM-KNN and KM-RF methods (ACC: 76.52%). When the effect of hybrid methods on prediction success in MR-3 is examined, it is seen that the most significant change is in KM-KNN. It is seen that the prediction success achieved with KM-KNN is 2.59% higher than the prediction success achieved with KNN.

3.3 Findings of Hybrid Methods Created by Fuzzy C-Means Clustering

The C value is changed from 2 to 15, and fuzzy C-Means cluster analysis is performed with each change. In each of these analyzes, C+1 new attribute is obtained and added to the dataset. Thus, 14 different alternative datasets are obtained. The optimal C value is determined for each method by comparing the findings obtained with these datasets.

(a) the first dataset, for home win **(b)** dimension red. dataset, for home win

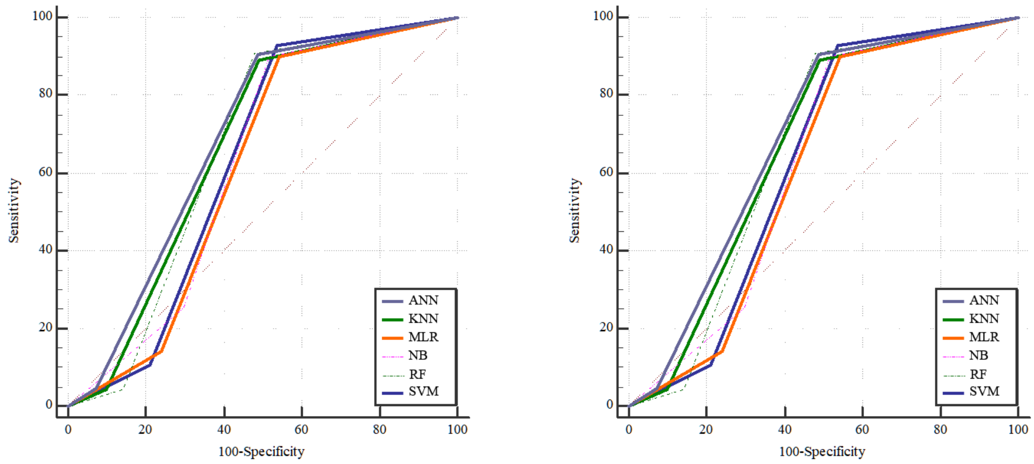

Figure 9: ROC curves according to "away team win" for MR-1

(a) the first dataset, for home win **(b)** dimension red. dataset, for home win

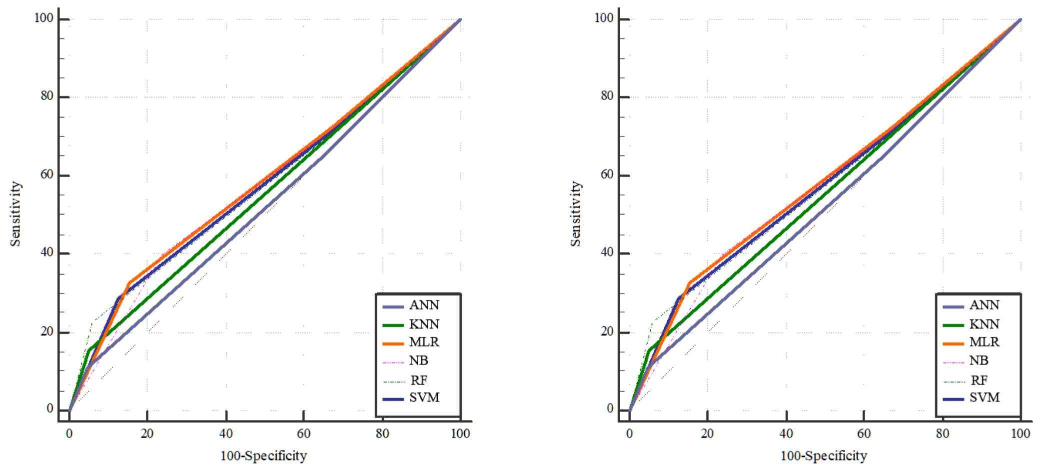

Figure 10: ROC curves according to "draw" for MR-1

Findings obtained with optimal C values are shared in Table 31-33 for MR-1, Table 34-36 for MR-2 and Table 37-39 for MR-3.

According to the findings of hybrid methods developed by using FCM clustering for the case of MR-1, the highest prediction success is achieved with FCM-RF method (ACC: 65.46%). When the effect of hybrid methods on prediction success in MR-1 is examined, it is seen that the most significant change is in FCM-MLR. It is seen that the prediction success achieved with FCM-MLR is 2.27% higher than the prediction success achieved with MLR.

In the tests performed for MR-2 status using hybrid methods developed with FCM clustering, the highest prediction success is achieved with FCM-SVM method (ACC: 81.49%). When the effect of hybrid methods on prediction success in MR-2 is examined, it is seen that the most significant change is in FCM-SVM. It is seen that the prediction success achieved with FCM-SVM is 3.82% higher than the prediction success achieved with SVM.

According to the findings of hybrid methods developed by using FCM clustering for MR-3, the highest prediction successes is achieved with FCM-RF method (ACC: 77.88%). When the effect of hybrid methods on prediction success in MR-3 is examined, it is seen that the most significant change is in FCM-RF. It is seen that the prediction success achieved with FCM-RF is 2.38% higher than the prediction success achieved with RF.

Table 16: Confusion matrices reached by classification methods for MR-2

		ANN 0	ANN 1	KNN 0	KNN 1	MLR 0	MLR 1	NB 0	NB 1	RF 0	RF 1	SVM 0	SVM 1
First Dataset	**0**	392	324	356	360	393	323	**493**	223	428	288	439	277
	1	176	1325	148	**1353**	196	1305	331	1170	172	1329	212	1289
Dimension Red. Dataset	**0**	426	290	406	310	362	354	**507**	209	426	290	372	344
	1	207	1294	173	1328	159	**1342**	303	1198	175	1326	168	1333

Table 17: Performance measures for MR-2

Class. Method	*Class*	First Dataset *TPR*	*TNR*	*PPV*	*F-M*	*KS*	*ACC*	Dimension Reduced Dataset *TPR*	*TNR*	*PPV*	*F-M*	*KS*	*ACC*
ANN	0	0.548	0.883	0.69	0.611	0.45	77.45%	0.595	0.862	0.673	0.632	0.47	77.58%
	1	0.883	0.547	0.804	0.841			0.862	0.595	0.817	0.839		
KNN	0	0.497	0.901	0.706	0.584	0.43	77.08%	0.567	0.885	0.701	0.627	0.48	78.21%
	1	0.901	0.497	0.79	0.842			0.885	0.567	0.811	0.846		
MLR	0	0.549	0.869	0.667	0.602	0.44	76.59%	0.506	0.894	0.695	0.585	0.43	76.86%
	1	0.869	0.549	0.802	0.834			0.894	0.506	0.791	0.84		
NB	0	0.689	0.779	0.598	0.64	0.45	75.01%	0.708	0.798	0.626	0.664	0.49	76.91%
	1	0.779	0.689	0.84	0.809			0.798	0.708	0.851	0.824		
RF	**0**	**0.598**	**0.885**	**0.713**	**0.65**	**0.5**	**79.25%**	**0.595**	**0.883**	**0.709**	**0.647**	**0.5**	**79.03%**
	1	**0.885**	**0.598**	**0.822**	**0.852**			**0.883**	**0.595**	**0.821**	**0.851**		
SVM	0	0.613	0.859	0.674	0.642	0.48	77.94%	0.52	0.888	0.689	0.592	0.44	76.91%
	1	0.859	0.613	0.823	0.841			0.888	0.52	0.795	0.839		

Table 18: AUC values and confidence intervals for MR-2

Class. Methods	First Dataset AUC	Std. Error	95% CI	Dimension Reduced Dataset AUC	Std. Error	95% CI
ANN	0.715	0.0102	0.695 - 0.735	0.729	0.0102	0.709 - 0.749
KNN	0.699	0.0101	0.679 - 0.719	0.726	0.0101	0.706 - 0.746
MLR	0.709	0.0103	0.689 - 0.729	0.700	0.0102	0.680 - 0.720
NB	0.734	0.0102	0.714 - 0.754	**0.753**	**0.0099**	**0.734 - 0.772**
RF	**0.742**	**0.0100**	**0.722 - 0.762**	0.739	0.0101	0.719 - 0.759
SVM	0.715	0.0102	0.695 - 0.735	0.729	0.0102	0.709 - 0.749

(a) the first dataset **(b)** dimension reduced dataset

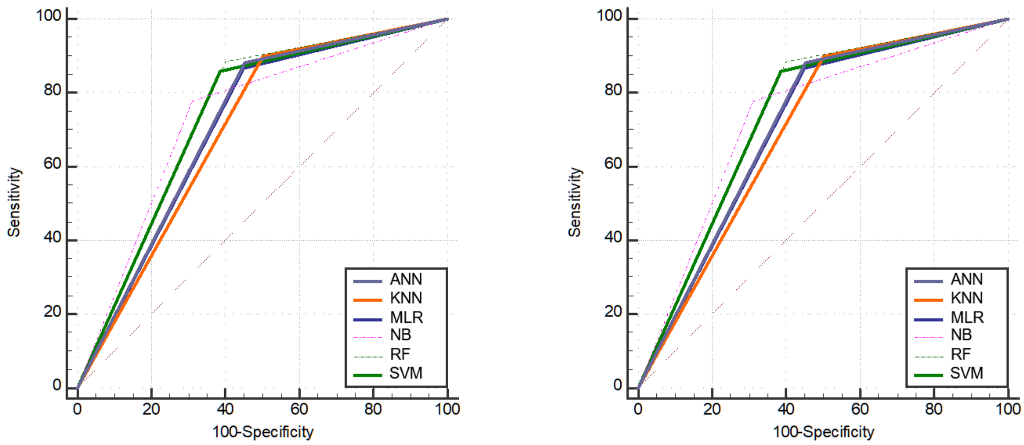

Figure 11: ROC curves for MR-2

Table 19: Confusion matrices reached by classification methods for MR-3

		ANN 0	ANN 1	KNN 0	KNN 1	MLR 0	MLR 1	NB 0	NB 1	RF 0	RF 1	SVM 0	SVM 1
First Dataset	**0**	676	322	**1019**	279	995	303	981	317	961	337	1003	295
	1	268	651	339	580	303	616	278	641	206	**713**	290	629
Dimension Red. Dataset	**0**	997	301	1015	283	1025	273	953	345	952	346	**1127**	171
	1	318	601	295	624	331	588	234	685	213	**706**	471	448

Table 20: Performance measures for MR-3

Class. Method	Class	First Dataset TPR	TNR	PPV	F-M	KS	ACC	Dimension Reduced Dataset TPR	TNR	PPV	F-M	KS	ACC
ANN	0	0.752	0.708	0.785	0.768	0.46	73.39%	0.768	0.654	0.758	0.763	0.42	72.08%
	1	0.708	0.752	0.669	0.688			0.654	0.768	0.666	0.654		
KNN	0	0.785	0.631	0.75	0.767	0.42	72.12%	0.782	0.679	0.775	0.778	0.46	73.93%
	1	0.631	0.785	0.675	0.652			0.679	0.782	0.688	0.683		
MLR	0	0.767	0.67	0.767	0.767	0.44	72.67%	0.79	0.64	0.756	0.772	0.43	72.76%
	1	0.67	0.767	0.67	0.67			0.64	0.79	0.683	0.661		
NB	0	0.756	0.697	0.779	0.756	0.45	73.16%	0.734	0.745	0.803	0.734	0.47	73.88%
	1	0.697	0.756	0.669	0.697			0.745	0.734	0.665	0.703		
RF	**0**	**0.74**	**0.776**	**0.823**	**0.78**	**0.51**	**75.50%**	**0.733**	**0.768**	**0.817**	**0.773**	**0.49**	**74.79%**
	1	**0.776**	**0.74**	**0.679**	**0.724**			**0.768**	**0.733**	**0.671**	**0.716**		
SVM	0	0.773	0.684	0.776	0.774	0.46	73.61%	0.868	0.487	0.705	0.778	0.37	71.04%
	1	0.684	0.773	0.681	0.683			0.487	0.868	0.724	0.487		

Table 21: AUC values and confidence intervals for MR-3

Class. Methods	First Dataset AUC	Std. Error	95% CI	Dimension Reduced Dataset AUC	Std. Error	95% CI
ANN	0.73	0.0096	0.711 - 0.749	0.711	0.0098	0.692 - 0.730
KNN	0.708	0.00979	0.689 - 0.727	0.73	0.0096	0.711 - 0.749
MLR	0.718	0.00973	0.699 - 0.737	0.715	0.00974	0.696 - 0.734
NB	0.727	0.00965	0.708 - 0.746	0.74	0.00945	0.721 - 0.759
RF	**0.758**	**0.00919**	**0.740 - 0.776**	**0.751**	**0.00928**	**0.733 - 0.769**
SVM	0.729	0.00963	0.710 - 0.748	0.678	0.00949	0.659 - 0.697

(a) the first dataset **(b)** dimension reduced dataset

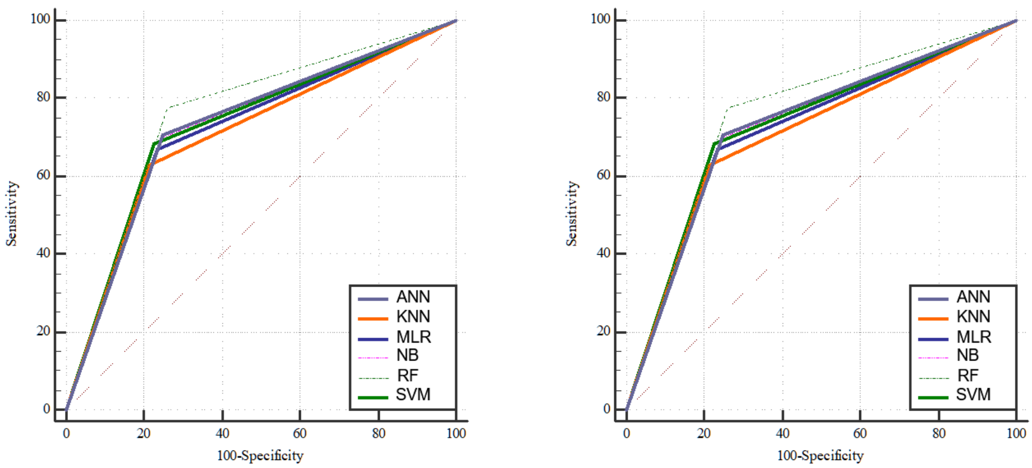

Figure 12: ROC curves for MR-3

Table 22: Confusion matrices obtained from hybrid methods created with KM clustering for MR-1

	KM-ANN 0	1	2	KM-KNN 0	1	2	KM-MLR 0	1	2	KM-NB 0	1	2	KM-RF 0	1	2	KM-SVM 0	1	2
0	259	204	125	154	309	125	239	269	80	**274**	219	95	140	328	120	214	284	90
1	124	725	70	45	795	79	84	775	60	159	715	45	34	**830**	55	70	805	45
2	138	128	450	66	169	481	174	143	399	184	113	419	66	164	**486**	133	184	399

Table 23: Performance measures obtained from hybrid methods created with KM clustering for MR-1

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
KM-ANN	12	0	0.441	0.84	0.5	0.468	0.45	64.56%
		1	0.789	0.744	0.689	0.736		
		2	0.629	0.871	0.693	0.659		
KM-KNN	5	0	0.263	0.932	0.585	0.363	0.44	64.33%
		1	0.865	0.632	0.627	0.727		
		2	0.671	0.865	0.696	0.684		
KM-MLR	5	0	0.407	0.843	0.485	0.442	0.43	63.66%
		1	0.843	0.682	0.655	0.738		
		2	0.557	0.908	0.736	0.634		
KM-NB	12	0	0.466	0.791	0.447	0.456	0.44	63.43%
		1	0.778	0.744	0.686	0.729		
		2	0.586	0.908	0.745	0.656		
KM-RF	12	0	0.237	0.938	0.583	0.337	0.45	**65.46%**
		1	0.903	0.62	0.63	0.742		
		2	0.679	0.884	0.731	0.704		
KM-SVM	5	0	0.364	0.877	0.518	0.428	0.43	63.89%
		1	0.876	0.64	0.635	0.736		
		2	0.557	0.911	0.743	0.637		

Table 24: Effect of hybrid methods created with KM clustering on prediction successes in MR-1

Methods	Previous ACC	Hybrid ACC	Change
KM-ANN	63.35%	64.56%	1.21%
KM-KNN	61.57%	64.33%	2.76%
KM-MLR	60.49%	63.66%	3.19%
KM-NB	59.63%	63.43%	**3.80%**
KM-RF	**64.05%**	**65.46%**	1.41%
KM-SVM	61.66%	63.89%	2.29%

Table 25: Confusion matrices obtained from hybrid methods created with KM clustering for MR-2

	KM-ANN		KM-KNN		KM-MLR		KM-NB		KM-RF		KM-SVM	
	0	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
0	389	327	327	389	389	327	**491**	225	384	332	430	286
1	139	1362	84	**1417**	114	1387	218	1283	94	1407	129	1372

Table 26: Performance measures obtained from hybrid methods created with KM clustering for MR-2

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
KM-ANN	5	0	0.543	0.908	0.731	0.623	0.48	79.23%
		1	0.908	0.543	0.811	0.857		
KM-KNN	9	0	0.457	0.944	0.79	0.579	0.45	79.00%
		1	0.944	0.457	0.79	0.86		
KM-MLR	3	0	0.543	0.924	0.768	0.636	0.51	80.36%
		1	0.924	0.543	0.814	0.866		
KM-NB	12	0	0.686	0.855	0.686	0.686	0.54	80.14%
		1	0.855	0.686	0.855	0.855		
KM-RF	3	0	0.536	0.937	0.798	0.641	0.52	81.04%
		1	0.937	0.536	0.814	0.871		
KM-SVM	3	0	0.600	0.914	0.764	0.672	0.55	**81.49%**
		1	0.914	0.600	0.832	0.871		

Table 27: Effect of hybrid methods created with KM clustering on prediction successes in MR-2

Methods	Previous ACC	Hybrid ACC	Change
KM-ANN	77.58%	79.23%	1.65%
KM-KNN	78.21%	79.00%	0.79%
KM-MLR	76.86%	80.36%	3.50%
KM-NB	76.91%	80.14%	3.23%
KM-RF	**79.25%**	81.04%	1.79%
KM-SVM	77.94%	**81.49%**	**3.55%**

Table 28: Confusion matrices obtained from hybrid methods created with KM clustering for MR-3

	KM-ANN		KM-KNN		KM-MLR		KM-NB		KM-RF		KM-SVM	
	0	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
0	976	322	971	327	**986**	312	971	327	906	392	946	352
1	238	681	194	725	253	666	199	720	129	**790**	189	730

Table 29: Performance measures obtained from hybrid methods created with KM clustering for MR-3

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
KM-ANN	12	0	0.752	0.741	0.802	0.776	0.49	74.72%
		1	0.741	0.752	0.682	0.71		
KM-KNN	3	0	0.748	0.789	0.832	0.788	0.53	**76.52%**
		1	0.789	0.748	0.692	0.737		
KM-MLR	9	0	0.76	0.724	0.794	0.776	0.48	74.50%
		1	0.724	0.76	0.684	0.703		
KM-NB	12	0	0.748	0.784	0.828	0.786	0.52	76.30%
		1	0.784	0.748	0.69	0.734		
KM-RF	5	0	0.698	0.859	0.874	0.776	0.54	**76.52%**
		1	0.859	0.698	0.671	0.859		
KM-SVM	9	0	0.729	0.795	0.832	0.777	0.51	75.62%
		1	0.795	0.729	0.677	0.795		

Table 30: Effect of hybrid methods created with KM clustering on prediction successes in MR-3

Methods	Previous ACC	Hybrid ACC	Change
KM-ANN	73.39%	74.72%	1.33%
KM-KNN	73.93%	**76.52%**	**2.59%**
KM-MLR	72.76%	74.50%	1.74%
KM-NB	73.88%	76.30%	2.42%
KM-RF	**75.50%**	**76.52%**	1.02%
KM-SVM	73.61%	75.62%	2.01%

Table 31: Confusion matrices obtained from hybrid methods created with FCM clustering for MR-1

	FCM-ANN			FCM-KNN			FCM-MLR			FCM-NB			FCM-RF			FCM-SVM		
	0	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**
0	130	299	159	60	394	135	219	274	95	**279**	214	95	159	329	100	194	309	85
1	40	785	94	20	**830**	70	94	760	65	174	695	50	30	820	70	50	820	50
2	61	148	**506**	56	189	471	148	153	414	205	113	399	77	164	476	128	199	389

Table 32: Performance measures obtained from hybrid methods created with FCM clustering for MR-1

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
FCM-ANN	12	0	0.22	0.938	0.565	0.317	0.43	63.88%
		1	0.854	0.655	0.64	0.731		
		2	0.707	0.832	0.66	0.683		
FCM-KNN	5	0	0.144	0.957	0.548	0.228	0.4	62.76%
		1	0.903	0.574	0.603	0.903		
		2	0.671	0.865	0.696	0.684		
FCM-MLR	3	0	0.373	0.852	0.478	0.419	0.42	62.76%
		1	0.827	0.671	0.643	0.723		
		2	0.579	0.894	0.717	0.64		
FCM-NB	12	0	0.475	0.769	0.427	0.45	0.42	61.86%
		1	0.757	0.748	0.683	0.718		
		2	0.557	0.904	0.729	0.632		
FCM-RF	12	0	0.271	0.935	0.604	0.374	0.45	**65.46%**
		1	0.892	0.620	0.627	0.737		
		2	0.664	0.888	0.732	0.697		
FCM-SVM	5	0	0.331	0.892	0.527	0.406	0.42	63.21%
		1	0.892	0.609	0.62	0.392		
		2	0.543	0.911	0.738	0.626		

Table 33: Effect of hybrid methods created with FCM clustering on prediction successes in MR-1

Methods	Previous ACC	Hybrid ACC	Change
FCM-ANN	63.35%	63.88%	0.53%
FCM-KNN	61.57%	62.76%	1.19%
FCM-MLR	60.49%	62.76%	**2.27%**
FCM-NB	59.63%	61.86%	2.23%
FCM-RF	**64.05%**	**65.46%**	1.41%
FCM-SVM	61.66%	63.21%	1.55%

Table 34: Confusion matrices obtained from hybrid methods created with FCM clustering for MR-2

	FCM-ANN		FCM-KNN		FCM-MLR		FCM-NB		FCM-RF		FCM-SVM	
	0	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**
0	465	251	317	399	368	348	**471**	245	399	317	424	292
1	193	1308	79	**1422**	104	1397	208	1293	104	1397	119	1382

Table 35: Performance measures obtained from hybrid methods created with FCM clustering for MR-2

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
FCM-ANN	3	0	0.65	0.871	0.7	0.674	0.53	80.14%
		1	0.871	0.65	0.843	0.857		
FCM-KNN	12	0	0.443	0.947	0.795	0.569	0.44	78.78%
		1	0.947	0.443	0.786	0.859		
FCM-MLR	3	0	0.514	0.931	0.774	0.618	0.49	79.91%
		1	0.931	0.514	0.806	0.684		
FCM-NB	3	0	0.657	0.861	0.687	0.682	0.52	79.69%
		1	0.861	0.657	0.845	0.853		
FCM-RF	3	0	0.557	0.931	0.788	0.653	0.53	81.26%
		1	0.931	0.557	0.82	0.872		
FCM-SVM	3	0	0.593	0.921	0.776	0.672	0.55	**81.76%**
		1	0.921	0.593	0.83	0.873		

Table 36: Effect of hybrid methods created with FCM clustering on prediction successes in MR-2

Methods	Previous ACC	Hybrid ACC	Change
FCM-ANN	77.58%	80.14%	2.56%
FCM-KNN	78.21%	78.78%	0.57%
FCM-MLR	76.86%	79.91%	3.05%
FCM-NB	76.91%	79.69%	2.78%
FCM-RF	**79.25%**	81.26%	2.01%
FCM-SVM	77.94%	**81.76%**	**3.82%**

Table 37: Confusion matrices obtained from hybrid methods created with FCM clustering for MR-3

	FCM-ANN		FCM-KNN		FCM-MLR		FCM-NB		FCM-RF		FCM-SVM	
	0	1	0	1	0	1	0	1	0	1	0	1
0	1001	297	**1051**	247	986	312	971	327	931	367	946	352
1	278	641	298	621	258	661	204	715	124	**795**	199	720

Table 38: Performance measures obtained from hybrid methods created with FCM clustering for MR-3

Hybrid Method	K	Class	TPR	TNR	PPV	F-M	KS	ACC
FCM-ANN	9	0	0.771	0.697	0.78	0.776	0.47	74.04%
		1	0.697	0.771	0.686	0.692		
FCM-KNN	3	0	0.81	0.676	0.777	0.793	0.49	75.40%
		1	0.676	0.81	0.718	0.693		
FCM-MLR	3	0	0.76	0.719	0.79	0.775	0.48	74.27%
		1	0.719	0.76	0.682	0.7		
FCM-NB	9	0	0.748	0.778	0.825	0.785	0.52	76.07%
		1	0.778	0.748	0.689	0.778		
FCM-RF	3	0	0.717	0.865	0.881	0.791	0.56	**77.88%**
		1	0.865	0.717	0.687	0.766		
FCM-SVM	3	0	0.729	0.784	0.825	0.774	0.5	75.17%
		1	0.784	0.729	0.674	0.725		

Table 39: Effect of hybrid methods created with FCM clustering on prediction successes in MR-3

Methods	Previous ACC	Hybrid ACC	Change
FCM-ANN	73.39%	74.04%	0.65%
FCM-KNN	73.93%	75.40%	1.47%
FCM-MLR	72.76%	74.27%	1.51%
FCM-NB	73.88%	76.07%	2.19%
FCM-RF	**75.50%**	**77.88%**	**2.38%**
FCM-SVM	73.61%	75.17%	1.56%

3.4 Findings Regarding the Content of the Dataset

The attributes in the dataset are divided into 5 parts. These parts are betting odds, team line-up, fans opinion, stats of previous matches and produced attributes. The effect of each part on prediction success is investigated. In the first stage, it is investigated what change would be in prediction success if any of these parts are not included in the dataset. In the second stage, it is investigated what change in prediction success would be if the dataset consists of only one of these parts. Findings obtained as a result of these tries; It is shared in Fig. 13-14 for MR-1, in Fig. 15-16 for MR-2 and in Fig. 17-18 for MR-3.

Figure 13: The effect of excluded dataset parts on the prediction success for MR-1

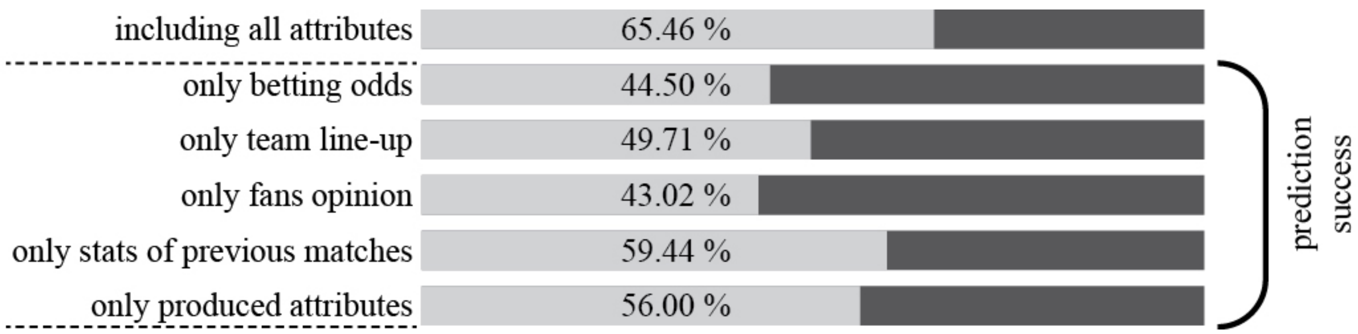

Figure 14: Prediction successes obtained with each part of the dataset in MR-1

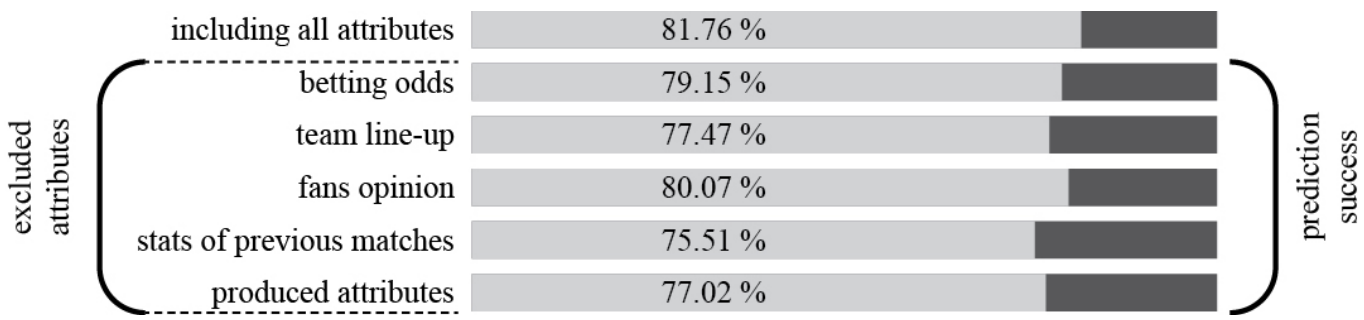

Figure 15: The effect of excluded dataset parts on the prediction success for MR-2

As shown in Figure 13-18, the part of the dataset that contributed the most to prediction success is the performance statistic from the previous matches. Produced attributes are part of the dataset that provides the second-highest contribution. Teams line-up is in third place. Betting odds and fan opinions are parts of the dataset that give the least contribution to prediction success.

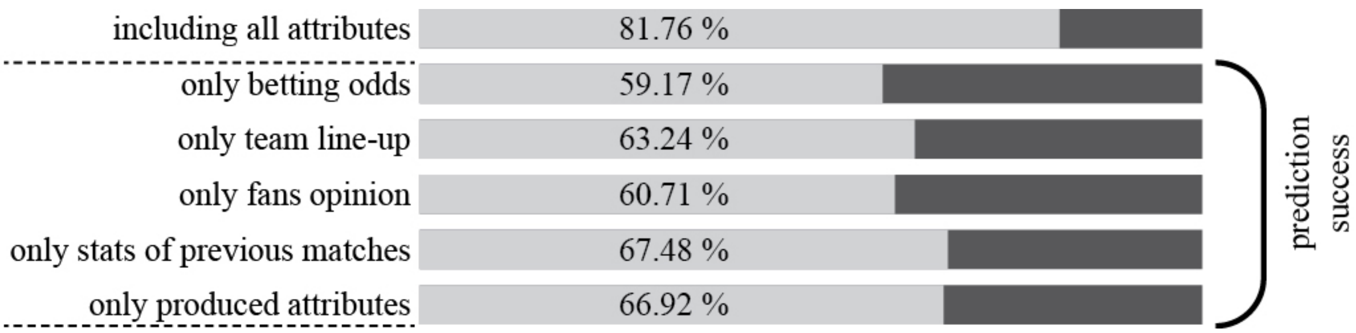

Figure 16: Prediction successes obtained with each part of the dataset in MR-2

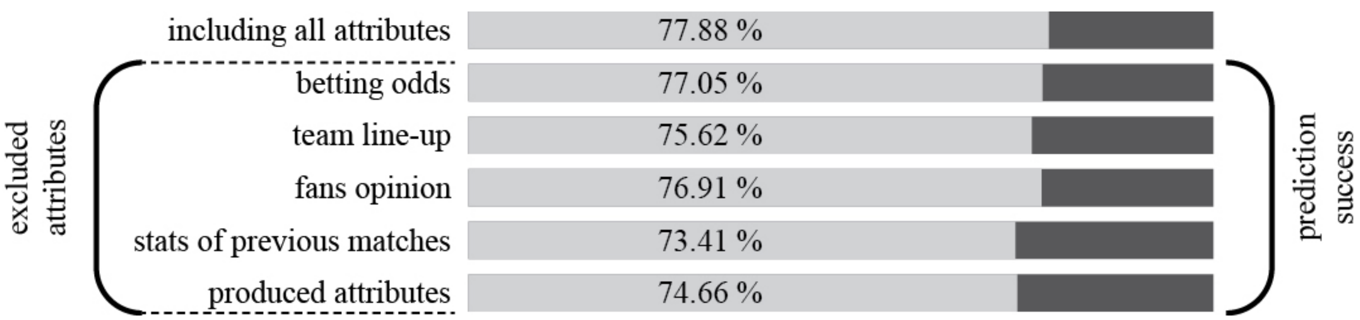

Figure 17: The effect of excluded dataset parts on the prediction success for MR-3

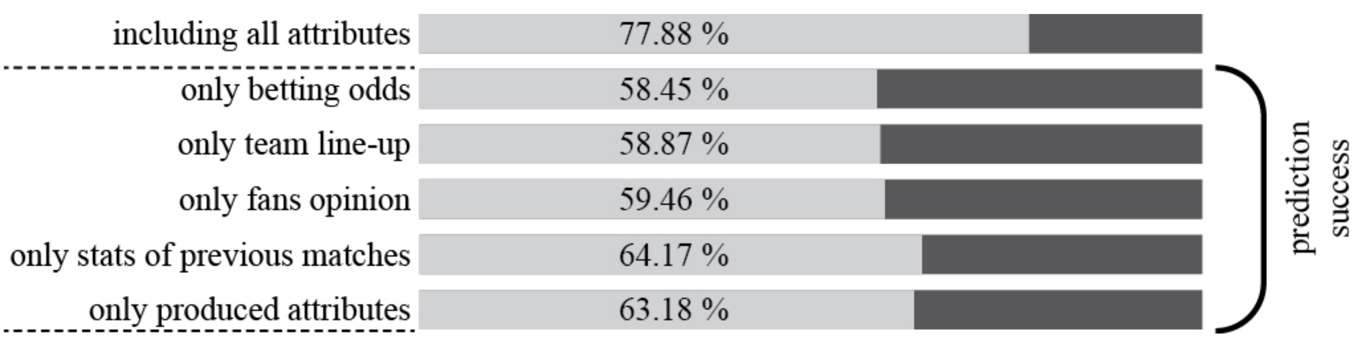

Figure 18: Prediction successes obtained with each part of the dataset in MR-3

4 Discussion

This study proposes hybrid classification methods to predict the results of football matches. Classification algorithms and clustering algorithms are used together in these hybrid methods. Artificial neural network, K nearest neighbor, multinomial logistic regression, naive bayes, random forest, support vector machine are used as the classification methods, and K-means clustering and fuzzy C-means clustering are used as the clustering methods.

As a first step, the prediction success of the classification algorithms on the existing dataset is examined. The findings are summarized as follows.

- In the case of MR-1, the most successful method on average is RF (64.05%). The most successful method in home win is SVM (89.60%),the most successful method in away win is RF (69.00%), and the most successful method in draw is NB (39.90%).
- In the case of MR-2, the most successful method on average is RF (79.25%). The most successful method in the group where home win and draw classes are combined is KNN (90.13%). The most successful method in the away win is NB (70.81%).
- In the case of MR-3, the most successful method, on average, is RF (75.05%). RF is the most successful method in home win (77.58%). SVM (86.80%) is the most successful method in the group where draw and away win classes are combined.

In some of the studies in the literature, the match result variable consists of 3 classes (Home, Draw, Away), while most studies consider the match result variable as 2 classes and ignore the matches that ended in a draw. Because it is very difficult to predict the results of football matches that ended in a draw. The data of these matches are very similar to the data of the matches won by the home or away team. For this reason, most classification methods cannot distinguish between matches that ended in a draw.

In this study, in the case of MR-1, the match result variable is considered as 3 classes and the values reached when the model is tested left behind similar studies [27, 29, 43] in the literature. The success rates obtained in MR-2 and MR-3 cases, where the match result variable consists of 2 classes, show how successful machine learning methods are in predicting the results of football matches when the draw is ignored.

In the second stage, hybrid models are proposed by using clustering algorithms to improve the success rates obtained with classification algorithms. The proposed hybrid methods significantly improve prediction success in each of the MR-1, MR-2 and MR-3 cases.

- In the case of MR-1, the most successful method is KM-RF (65.46%). The best significant improvement in performance is observed with the KM-NB method (+3.80%).
- In the case of MR-2, KM-SVM is the method in which both the most successful and the best significant performance improvement is observed (81.49%, +3.55%).
- In the case of MR-3, the most successful methods are KM-KNN and KM-RF (76.52%). The best significant improvement in performance is observed with the KM-KNN method (+2.59%).
- In the case of MR-1, the most successful method is FCM-RF (65.46%). The best significant improvement in performance is observed with the FCM-MLR method (+2.27%).
- In the case of MR-2, FCM-SVM is the method in which both the most successful and the best significant performance improvement is observed (81.76%, +3.82%).
- In the case of MR-3, FCM-RF is the method in which both the most successful and the best significant performance improvement is observed (77.88%, +2.38%).

The features in the dataset are divided into five sections: betting odds, team line-up, fan opinions, statistics of previous matches, and produced attributes. The contribution of these parts to the prediction success is investigated.

According to the findings, the part that includes the variables that give the performance values of the previous matches made the most contribution to the prediction success. A certain number of matches are taken into account in order to determine the previous performances of the teams. How many matches to take into account here is an important problem. Various attempts are made to determine this number. The number that maximized the prediction success is determined as six. Thus, the average performance power of the teams is calculated according to the values they reached in the last six games.

The attributes produced with expert's opinions are the second most contributing parts to the prediction success. In this part of the dataset, there are some variables that are not directly included in the performance values, and these variables generally play an important role in increasing the estimation success.

In the literature, there are various studies in which both social media shares and betting odds are used alone to predict match results [16, 17, 31, 40]. In this study, when the social media data and betting odds are used alone to predict match results, a lower success rate is achieved than the overall success of the study. With the entire dataset, prediction successes of 65.45% for MR-1, 81.76% for MR-2, and 77.88% for MR-3 are achieved. When social media data is used alone, these rates decrease to 43.02% for MR-1, 60.71% for MR-2, and 59.06% for MR-3. When the bet rates are used alone, it decreases to 44.50% for MR-1, 59.17% for MR-2, and 58.45% for MR-3.

5 Conclusion

The existence of predictive models that can predict the results of sports competitions is very important. Today, the development of these models and increasing their estimation success are the subject of many studies. One of the most frequently studied sports branches in this field is football. Football is a team game and there are many measurable and unmeasurable variables that affect its outcome. Therefore, it is very difficult to predict the results of football matches.

With the hybrid methods developed in this study, it is tried to overcome this difficulty and very high estimation successes are achieved. The prediction successes achieved reveal that artificial intelligence-based methods are quite successful in predicting the results of football matches. The developed model can be used by football teams, football-based betting platforms and companies working on sports statistics. In addition to predicting the match results, it can also inform the users about the effects of the variables in the dataset on the match result.

In order to increase the success of the model, the study can be developed in two directions. First of all, model success can be increased by expandig the dataset content. For this purpose;

- Some organizations archive high-dimensional football data but do not share these data on the internet. By connecting with these organizations, new data can be obtained to enrich the dataset.
- More data can be collected that provides insight into the individual performances and strengths of the players. In order to determine the variables to be included in these data, various studies investigating the effect of player performances on match results can be used. Cortez, A. et al. [12] have done a very comprehensive study for this purpose.
- With deep learning methods, information can be obtained from the teams' data from previous matches.
- A few features affect the outcome of the match but cannot be measured statistically. Motivation level of the player or technical staff, coach change, fan pressure etc. features are examples. There are various studies on the effects of such psychological factors on athlete success [2, 36]. The dataset can be enriched by making use of such studies.
- Sentiment analysis can be made on the articles shared by experts before the match. Results obtained from these analyzes can be added to the dataset.
- City, stadium, date, time, the number of fans, geographical conditions, etc. information can be added to the dataset.

In addition to expanding the content of the dataset, the success of the prediction can be increased by working on the methods used. For this purpose;

- Other classification and clustering methods can be used.
- New classification methods and hybrid solutions can be developed to increase prediction success.
- In order to better predict matches that ended in a draw, NB-based studies can be performed because NB is the most successful method for predicting draws.

Acknowledgment

This study is a part of PhD thesis of İsmail Hakkı KINALIOĞLU titled "Prediction of Football Match Results by Using Artificial Intelligence Based Methods and Proposals on Hybrid Model", which was completed at the Selcuk University, Institute of Science, Department of Statistics.

References

[1] A.A. Abbasi and M. Younis, *A survey on clustering algorithms for wireless sensor networks*, Comput. Commun. **30** (2007), no. 14-15, 2826–2841.

[2] M.R. Abdullah, R.M. Musa, A.B.H.M.B. Maliki, N.A. Kosni, and P.K. Suppiah, *Role of psychological factors on the performance of elite soccer players*, J. Phys. Educ. Sport **16** (2016), no. 1, 170.

[3] C.C. Aggarwal, *Data classification: algorithms and applications*, CRC Press, 2014.

[4] F. Amadin and J.C. Obi, *English premier league (epl) soccer matches prediction using an adaptive neuro-fuzzy inference system (anfis)*, Trans. Machine Learn. Artif. Intel. **3** (2015), no. 2, 34.

[5] W. Andreff and N. Scelles, *Walter c. neale 50 years after:beyond competitive balance, the league standing effect tested with french football data*, J. Sports Econ. **16** (2015), no. 8, 819–834.

[6] S.M. Arabzad, M.E. Tayebi Araghi, S. Sadi-Nezhad, and N. Ghofrani, *Football match results prediction using artificial neural networks; the case of iran pro league*, J. Appl. Res. Ind. Engin. **1** (2014), no. 3, 159–179.

[7] R. Baboota and H. Kaur, *Predictive analysis and modelling football results using machine learning approach for english premier league*, Int. J. Forecast. **35** (2019), no. 2, 741–755.

[8] G. Baio and M. Blangiardo, *Bayesian hierarchical model for the prediction of football results*, J. Appl. Statist. **37** (2010), no. 2, 253–264.

[9] G. Boshnakov, T. Kharrat, and I.G. McHale, *A bivariate weibull count model for forecasting association football scores*, Int. J. Forecast. **33** (2017), no. 2, 458–466.

[10] E.J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, J. ACM (JACM) **58** (2011), no. 3, 11.

[11] A. Caraffa, G. Cerulli, M. Projetti, G. Aisa, and A. Rizzo, *Prevention of anterior cruciate ligament injuries in soccer*, Knee Surgery Sports Traumatol. Arthros. **4** (1996), no. 1, 19–21.

[12] A. Cortez, A. Trigo, and N. Loureiro, *Predicting physiological variables of players that make a winning football team: A machine learning approach*, Int. Conf. Comput. Sci. Appl., Springer, 2021, pp. 3–15.

[13] E. Costa, A. Lorena, A.C.P.L.F. Carvalho, and title = A review of performance evaluation measures for hierarchical classifiers booktitle = Evaluation Methods for Machine Learning II: papers from the AAAI-2007 Workshop pages = 1-6 type = Conference Proceedings Freitas, A.

[14] A. Decrop and C. Derbaix, *Pride in contemporary sport consumption: A marketing perspective*, J. Acad. Market. Sci. **38** (2010), no. 5, 586–603.

[15] M.J. Dixon and S.G. Coles, *Modelling association football scores and inefficiencies in the football betting market*, J. Royal Statist. Soc.: Ser. C (Applied Statistics) **46** (1997), no. 2, 265–280.

[16] M.J. Dixon and P.F. Pope, *The value of statistical forecasts in the uk association football betting market*, Int. J. Forecast. **20** (2004), no. 4, 697–711.

[17] Engin Esme and M.S. Kiran, *Prediction of football match outcomes based on bookmaker odds by using k-nearest neighbor algorithm*, Int. J. Machine Learn. Comput. **8** (2018), no. 1, 26–32.

[18] C.W. Fuller, J. Ekstrand, A. Junge, T.E. Andersen, R. Bahr, J. Dvorak, M. Hägglund, P. McCrory, and W.H. Meeuwisse, *Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries*, Scand. J. Med. Sci. Sports **16** (2006), no. 2, 83–92.

[19] J. Goddard, *Regression models for forecasting goals and match results in association football*, Int. J. Forecast. **21** (2005), no. 2, 331–340.

[20] M. Hall, E. Frank, G. Holmes, P. Pfahringer, B.and Reutemann, and I.H. Witten, *The weka data mining software: An update*, ACM SIGKDD Explor. Newslett. **11** (2009), no. 1, 10–18.

[21] M.A. Hall, *Correlation-based feature selection for machine learning*, (1999).

[22] R.S. Heidt, L.M. Sweeterman, R.L. Carlonas, J.A. Traub, and F.X. Tekulve, *Avoidance of soccer injuries with preseason conditioning*, Amer. J. Sports Med. **28** (2000), no. 5, 659–662.

[23] R.G. Houston and D.P. Wilson, *Income, leisure and proficiency: an economic study of football performance*, Appl. Econ. Lett. **9** (2002), no. 14, 939–943.

[24] O. Hubáček, G. Šourek, and F. Železný, *Learning to predict soccer results from relational data with gradient boosted trees*, Machine Learn. **108** (2019), no. 1, 29–47.

[25] J. Hucaljuk and A. Rakipović, *Predicting football scores using machine learning techniques*, 2011 Proceedings of the 34th International Convention MIPRO, pp. 1623–1627.

[26] L.M. Hvattum and H. Arntzen, *Using elo ratings for match result prediction in association football*, Int. J. Forecast. **26** (2010), no. 3, 460–470.

[27] C.P. Igiri, *Support vector machine—based prediction system for a football match result*, IOSR J. Comput. Engin. (IOSR-JCE) **17** (2015), no. 3, 21–26.

[28] D. Jones and M. Green, *Deloitte annual review of football finance 2019*, Deloitte, 2019.

[29] A. Joseph, N.E. Fenton, and M. Neil, *Predicting football results using bayesian nets and other machine learning techniques*, Knowledge-Based Syst. **19** (2006), no. 7, 544–553.

[30] A. Junge and J. Dvorak, *Soccer injuries*, Sports Med. **34** (2004), no. 13, 929–938.

[31] S. Kampakis and A. Adamides, *Using twitter to predict football outcomes*, arXiv preprint arXiv:1411.1243 (2014).

[32] A.G. Karegowda, A.S. Manjunath, and M.A. Jayaram, *Comparative study of attribute selection using gain ratio and correlation based feature selection*, Int. J. Inf. Technol. Knowledge Manag. **2** (2010), no. 2, 271–277.

[33] C.K. Leung and K.W. Joseph, *Sports data mining: Predicting results for the college football games*, Procedia Comput. Sci. **35** (2014), 710–719.

[34] R.G. Martins, A.S. Martins, L.A. Neves, L.V. Lima, E.L. Flores, and M.Z. do Nascimento, *Exploring polynomial classifier to predict match results in football championships*, Expert Syst. Appl. **83** (2017), 79–93.

[35] B. Min, J. Kim, C. Choe, H. Eom, and R.B. McKay, *A compound framework for sports results prediction: A football case study*, Knowledge-Based Syst. **21** (2008), no. 7, 551–562.

[36] T. Morris, *Psychological characteristics and talent identification in soccer*, J. Sports Sci. **18** (2000), no. 9, 715–726.

[37] Md.A. Rahman, *A deep learning framework for football match prediction*, SN Appl. Sci. **2** (2020), no. 2, 165.

[38] Zion Market Research, *Research: Sports betting market size 2019-2024*, 2019.

[39] S. Robertson, N. Back, and J.D. Bartlett, *Explaining match outcome in elite australian rules football using team performance indicators*, J. Sports Sci. **34** (2016), no. 7, 637–644.

[40] R.P. Schumaker, A.T. Jarmoszko, and C.S. Labedz Jr, *Predicting wins and spread in the premier league using a sentiment analysis of twitter*, Decision Support Syst. **88** (2016), 76–84.

[41] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, *Rocr: visualizing classifier performance in r*, Bioinf. **21** (2005), no. 20, 3940–3941.

[42] J. Tang, S. Alelyani, and H. Liu, *Feature selection for classification: A review*, Data classif.: Algorithms Appl. (2014), 37.

[43] A.S. Timmaraju, A. Palnitkar, and V. Khanna, *Game on! predicting english premier league match outcomes*, 2013.

[44] K.M. Ting, *Confusion matrix*, pp. 260–260, Springer US, Boston, MA, 2017.

[45] M.E. Tipping and C.M. Bishop, *Probabilistic principal component analysis*, J. Royal Statist. Soc.: Ser. B (Statis-

tical Methodology) **61** (1999), no. 3, 611–622.

[46] N. Vlastakis, G. Dotsis, and R.N. Markellos, *Nonlinear modelling of european football scores using support vector machines*, Appl. Econ. **40** (2008), no. 1, 111–118.

[47] S. Wold, K. Esbensen, and P. Geladi, *Principal component analysis*, Chemomet. Intell. Lab. Syst. **2** (1987), no. 1-3, 37–52.

[48] H. Zou, T. Hastie, and R. Tibshirani, *Sparse principal component analysis*, J. Comput. Graph. Statist. **15** (2006), no. 2, 265–286.