Cumhuriyet Science Journal
e-ISSN: 2587-246X
ISSN: 2587-2680
Cumhuriyet Sci. J., 41(4) (2020) 951-967
http://dx.doi.org/10.17776/csj.780391

Prediction of UEFA champions league elimination rounds winners using machine learning algorithms

İsmail Hakkı KINALIOĞLU[1,*], *Coşkun KUŞ*[2]

[1]*Uşak University, Department of Computer Programming, USAK/ TURKEY*

[2]*Selçuk University, Department of Statistics, KONYA/ TURKEY*

Abstract

In this study, the teams that qualified for the next round as a result of two-legged matchups are predicted using the data collected from the UEFA (Union of European Football Associations) Champions League group stage matches. The study contributes to the literature in terms of variety of methods used and content of the dataset compared to other studies conducted on football data. It is also a pioneering study to predict the outcome of a two-legged matchup. The data are collected from the matches played in the Champions League organizations held between 2010-2018. Classification methods as Artificial Neural Network, K-Nearest Neighbors, Logistic Regression Analysis, Naive Bayes Classifier, Random Forest and Support Vector Machine are used for the prediction. Two applications are carried out to test the successes of the classification models. In the first application, the most successful method is naive bayes classifier (86.66%) and in the second application, the most successful method is random forest (74.81%).

Article info

History:
Received: 14.08.2020
Accepted: 18.11.2020

Keywords:
Data classification,
Football data mining,
Match result prediction,
Machine learning

1. Introduction

Football or soccer is the most popular sport branch of our time thanks to its massive fan base worldwide. [1-5]. A great number of national and international football tournaments take place today. One of the most well-known organizations is the UEFA Champions League. The Champions League is held between the teams of the countries affiliated to UEFA that qualify for the tournament and consists of the qualifying rounds, group matches and final matches, respectively. Some of the teams gain direct entry to the group stage while others are qualified after taking part in different number of qualifying round ties. The group stage begins with 32 teams and these teams are split into eight groups of four teams. The top two teams from each group are advanced to the Round of 16 after the group stage. Thus, 15 matchups take place in the Champions League: Round of 16, quarter-final, semi-final and final, respectively. These matchups are shown in Fig 1.

*Corresponding author. *e-mail address: ismailhakkikinali@gmail.com*

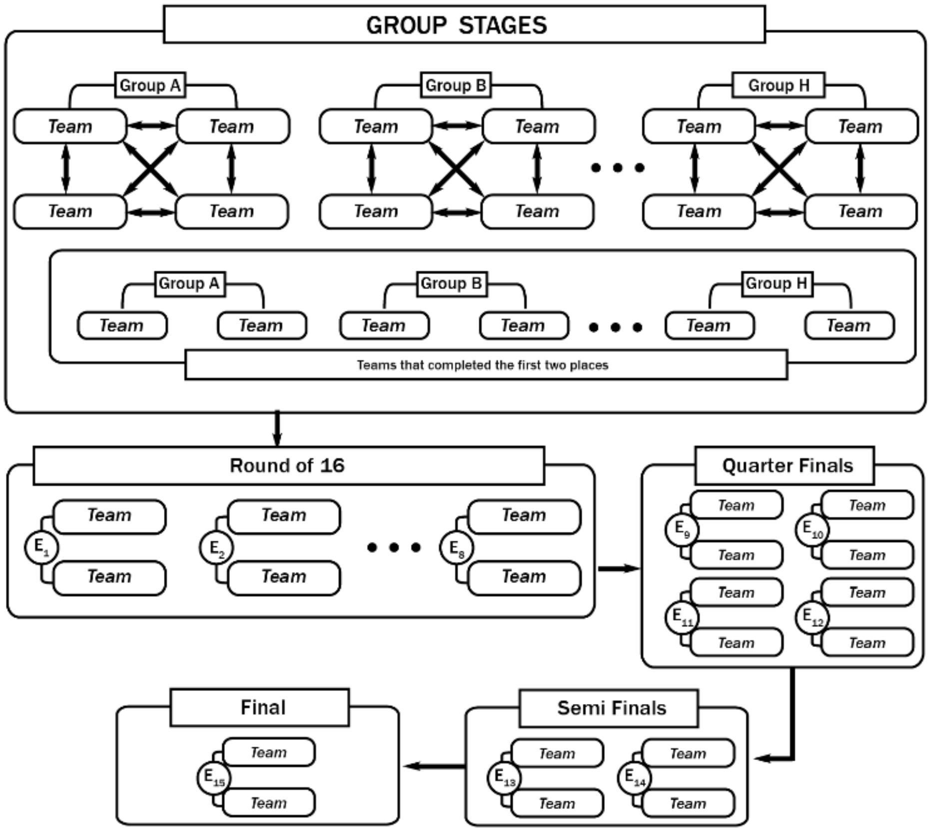

Figure 1. Organization of the champions league matches

Predicting the result of a future football match is important in several ways. A successful prediction model is important for betting organizations as well as contributing to sports analytics studies. It contributes to sports analytics studies on topics such as evaluating performance data, explaining the relationship between these data and the result of the match, determining team and player strength. On the other hand, an effective prediction model is crucial for both bettors and organizers to increase profitability. While the prediction models are used to calculate the bet odds by organizers, it is also used by players to gain more from the bet.

Results of the football competitions are based on many independent variables (predictors). Variables such as the number of passes, the number of shots, and the percentage of playing the ball are related to the result of the match which is considered as dependent variable. In addition to these variables, the methods used to predict the match result are important. When the literature is examined, it is seen that machine learning algorithms are frequently used in predicting match results. Methods such as Artificial Neural Network (ANN), K-Nearest Neighborhood (KNN), Logistic Regression Analysis (LRA), Naive Bayes Classifier (NBC), Random Forest (RF) and Support Vector Machine (SVM), are the most commonly used. Some of these studies that differ in terms of dataset contents and methods used are given below.

Goddard [6] used 25 years of data obtained from the English Premier League in order to predict the goals and results of matches. He conducted a bivariate Poisson regression analysis to predict the scores and “ordered probit regression” for prediction. Huang and Chen [7] predicted the next rounds’ results using artificial neural networks with the data obtained from 2006 FIFA World Cup group results. They considered the number of goals and shots, shot on target, corner kick, free kick, indirect free kick and fouls as independent variables and achieved an accuracy rate of 85.7% in the second round, 66.7% in the quarter-final and 50% in the semi-final while they correctly predicted the final match. The general accuracy rate is calculated as 76.9%. Timmaraju et al. [8] considered the independent variables as goal, corner and shot on target and predicted the results of the English Premier League matches using the Multinomial Logistic Regression and Support Vector Machine. The highest accurate prediction rate, 66.7%, is obtained through the Support Vector Machine. Ulmer and Fernandez [9], intended to predict the results using artificial intelligence and machine learning with the data obtained from English Premier League clubs. Historical data, match day data and certain independent variables of current team performance are used for the classification. Additionally, the methods as Linear Regression with Stochastic Gradient Descent, Naive Bayes, Hidden Markov Model, Support Vector Machine and Random Forest are implemented. Linear Regression with Stochastic Gradient Descent (0.48), Random Forest (0.50) and Support Vector Machine (0.50) provided better error rates. Late and Gupta [10] predicted the results using artificial intelligence and machine learning. Historical data and current data of the clubs are used as independent variables. The highest accurate prediction rate is found to be 42.2% in their study. Igiri [11] used support vector machines to predict the results of football matches. There are 38 features related to the performances of the teams in the study dataset. The prediction success achieved in the tests is 53.3%. Baboota and Kaur [12] predicted the match results using gaussian naive bayes, random forest, support vector machine, and gradient boosting methods. In the dataset, there are 33 attributes that show the past performance of the teams. As a result of the tests, they achieved the highest prediction success with the gradient boosting method (57%). Rahman [13] predicted the results of the 2018 World Cup group stage matches with deep learning. He proposed a neural network model using LSTM. The dataset is prepared by evaluating the teams' performances in international matches between 1872-2018. As a result of the tests, 63.3% prediction success is achieved. For more details on prediction of the match results please see Rotshtein et al. [14], Joseph et al. [15], Hvattum and Arntzen [16], Owen [17], Constantinou et al. [18], Igiri and Nwachukwu [19], Koopman and Lit [20], Amadin and Obi [21], Robertson et al. [22], Gevaria et al. [23], Prasetio and Dra. Harlili [24], Martins et al. [25], Bunker and Thabtah [26].

In this study, the winners of the two-legged machups are predicted with the information obtained from the group stage matches of the UEFA champions league tournaments between 2010 and 2018. ANN, KNN, LRA, NBC, RF and SVM methods are used for prediction. Prediction methods, evaluation measures and information about data preparation process is provided in Section 2. Application results and accurate prediction rates of the methods are presented in Section 3. In Section 4, the results are discussed and information on future studies is provided.

2. Materials and Methods

2.1. Classification methods and evaluation measures

Classification is defined as assigning the observations in the dataset to certain labels according to their characteristics [27]. In this study, predicting the results of football matches is considered as a classification problem. The concept of classification is one of the most popular topics studied in data mining and machine learning and is also used in many different disciplines recently [28]. There are many methods used for classification in the literature. In this study, ANN, KNN, LRA, NBC, RF and SVM methods are used and brief information about these methods is given below.

2.1.1. Artificial neural network (ANN)

The idea of artificial neural network began with a paper written by neuroscientist, Warren McCulloch, and a mathematician, Walter Pitts [29]. Artificial neural network is a learning system that focuses on simulating biological human brain entirely different from traditional computers [30]. It is used to solve various problems like classification, prediction, identification, modeling, etc. An artificial neural network should have learning ability in order to solve the problem. It requires a wealth of data to learn, just as the human brain that needs neuronal data transmitted by the sense organs, also called knowledge. The network creates an output value as a result of a series of operations performed by inserting this data into the network as input. This value corresponds to the solution to be reached.

2.1.2. K-nearest neighbors (KNN)

KNN algorithm may be used as an alternative for classification procedures when there is no information about data distribution or limited [31]. In 1951, Fix and Hodges [32] introduced the method, which is a nonparametric model called the nearest neighbor decision rule, and used for pattern classification. Certain characteristics of the KNN algorithm are later specified by Cover and Hart [33] in 1967. In time, new approaches are presented for the algorithm, which is outlined with these studies. This algorithm calculates the distance between the classified data and others within the dataset. There are various distance measures used to calculate this distance. Euclidean and Manhattan distance measurements are the most common. The calculated distance values are listed ascending and the closest neighbors are determined and the data are classified. This simple and easy-to-use method can provide successful results even when it is compared to the most complex machine learning methods [34].

2.1.3. Logistic regression analysis (LRA)

The origins of the logistic regression go back to the early 19th century [35]. It differs from the linear regression model due to the categorical and multi-class dependent variable [36]. Independent variables can be categorical or continuous, while the dependent variable is only categorical [37]. When the dependent variable has two categories, the binary logistic regression is applied. If the dependent variable has more than two categories, multinomial logistic regression can be performed. Logistic regression models give odds ratios and related confidence intervals. It provides a solution for various problems as classification and pattern recognition.

2.1.4. Naive bayes classifier (NBC)

The Naive Bayes classification algorithm is based on the Bayes Theorem. Naive Bayes is a simple form of the Bayesian network where all attributes are independent for the class variable [38]. Although the Naive Bayes classifier is a structurally simple classification method, it is a very effective and baseline classifier [39]. It is widely used in many applications as text classification, medical diagnosis and performance management [40].

2.1.5. Random forest (RF)

The model, first called Random Decision Forest by Ho [41], is identified with its widely used name, Random Forest, by Breiman [42] in 1999. Random forests are a combination of tree predictors depend on the values of a vector which are independent and identical distributed [43]. The structure of the random forests is built through multiple decision trees, instead of individuals, trained with different datasets derived from the original using bootstrapping. The prediction of the forest is achieved as a result of the predictions obtained from each tree within this structure that may have a different number of variables.

2.1.6. Support vector machine (SVM)

The Support Vector Machines are based on the Statistical Learning Theory, known as the Vapnik-Chervonenkis (VC), and they are introduced by Vladimir Vapnik in 1960 [44]. They are mainly a type of machine learning systems that use the binary classification method. This method aims to separate two classes from each other according to the hyperplane that will be formed by transforming the data into a higher dimension [45]. Support Vector Machines use a linear separation function and their purpose is to predict the most appropriate function to split data. The estimation of this function is the solution for the optimization problem which ensures the maximum distance between the two boundaries. Support vector machines designed to solve two-class problems are also used in multi-class classification problems with various modifications [46-51].

2.1.7. Evaluation measures

The Confusion Matrix given in Table 1 is used to compare the classification methods. Some evaluation measures based on the Confusion Matrix are given below.

Table 1. Demonstration of the confusion matrix

		Actual Result of Elimination Tour	
		Team 1 (Negative)	Team 2 (Positive)
Prediction	Team 1 (Negative)	TN (True Negative)	FP (False Positive)
	Team 2 (Positive)	FN (False Negative)	TP (True Positive)

The structure of the confusion matrix given in Table 1 is adapted for this classification problem as follows;

Negative when A (Team 1) is qualified for the next round,

Positive when B (Team 2) is qualified for the next round,

TN: The number of those classified as "Team 1 qualified" when it is qualified for the next round,

FP: The number of those classified as "Team 1 qualified" when it is not qualified for the next round,

FN: The number of those classified as "Team 2 qualified" when it is not qualified for the next round,

TP: The number of those classified as "Team 2 qualified" when it is qualified for the next round,

Based upon these variables, the evaluation statistics are defined as follows:

$$\boldsymbol{PO} = \frac{TP + TN}{TN + FN + FP + TP} \qquad \boldsymbol{PE} = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TN + FN + FP + TP)^2}$$

$$\boldsymbol{TP\ Rate} = \frac{TP}{TP + FN} \qquad \boldsymbol{FP\ Rate} = \frac{FP}{FP + TN} \qquad \boldsymbol{TN\ Rate} = \frac{TN}{TN + FP} \qquad \boldsymbol{FN\ Rate} = \frac{FN}{TP + FN}$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \qquad \boldsymbol{Kappa\ Statistic} = \frac{PO - PE}{1 - PE} \qquad \boldsymbol{F\text{ - }Measure} = \frac{2TP}{(2TP + FP + FN)}$$

$$\boldsymbol{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

TP rate is also known as sensitivity or Recall and TN rate is known as specificity. High values of Accuracy, TP Rates, TN Rates, Precision and F-Measure indicate the success of classification methods. MCC is a correlation coefficient between the predicted and observed binary classifications. On the other hand, the kappa statistic is interested in the coherency between results of the predicted and observed classifications to check sample points correctly classified by chance, and it ranges from 0 to 1. Kappa value implies a successful classifier for large kappa values (near 1). If the kappa values almost zero, it indicates that the classifier is equivalent to chance.

2.2. Data description and analysis

It is essential to have adequate information on the teams before the elimination stage that meant to be predicted.

It is essential to have adequate information on the teams before the elimination stage that meant to be predicted. In order to have information about the teams, their performances in previous matches should be evaluated. For this purpose, statistical data of champions league matches played between 2010-2018 are collected. These data are obtained from the website named "WhoScored.Com".

Each team plays six matches in group stage. 40 independent variables are observed for each team from these six matches. These variables are presented in Table 2. 1x40 sized feature vectors characterizing each team are derived from the mean of the independent variables. These feature vectors are used to compare the matching teams in qualifying rounds. The feature vectors difference of the two teams in the qualifying round consists of independent variables of the training and test dataset. The dependent variable is qualified team's label (1 or 2). These dependent and independent variables are identified as follows:

Let us $X_{kij}^{(\ell)}$ denotes the jth variable of first team of kth elimination match in ith group match at the ℓ organization and let $Y_{kij}^{(\ell)}$ denotes the jth variable of second team of the kth elimination match in the ith group match at the ℓ organization, where $\ell = 1, 2, ..., 9,$ $k = 1, 2, ..., 15,$ $i = 1, 2, ..., 6,$ and $j = 1, 2, ..., 40$. Let us define

$$\overline{X}_{k \cdot j}^{(\ell)} = \frac{1}{6}\sum_{i=1}^{6} X_{kij}^{(\ell)} \quad \text{and} \quad \overline{Y}_{k \cdot j}^{(\ell)} = \frac{1}{6}\sum_{i=1}^{6} Y_{kij}^{(\ell)}, \tag{1}$$

and

$$\overline{D}_{k \cdot j}^{(\ell)} = \overline{X}_{k \cdot j}^{(\ell)} - \overline{Y}_{k \cdot j}^{(\ell)}. \tag{2}$$

Then independent variables are compiled by

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{(1)} \\ \mathbf{T}^{(2)} \\ \vdots \\ \mathbf{T}^{(9)} \end{pmatrix}_{135 \times 40}, \tag{3}$$

where

$$\mathbf{T}^{(\ell)} = \begin{pmatrix} \overline{D}_{1 \cdot j}^{(\ell)} & \overline{D}_{1 \cdot 2}^{(\ell)} & \cdots & \overline{D}_{1 \cdot 40}^{(\ell)} \\ \overline{D}_{2 \cdot j}^{(\ell)} & \overline{D}_{2 \cdot 2}^{(\ell)} & \cdots & \overline{D}_{2 \cdot 40}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{D}_{15 \cdot j}^{(\ell)} & \overline{D}_{15 \cdot 2}^{(\ell)} & \cdots & \overline{D}_{15 \cdot 40}^{(\ell)} \end{pmatrix}_{15 \times 40}. \tag{4}$$

Let us also define random variables for $r = 1, 2, ..., 15$

$$Z_r = \begin{cases} 1 & , \quad \textit{if the first team is qualified} \\ 2 & , \quad \textit{if the second team is qualified} \end{cases}$$

Main point of the study is prediction of the Z_r (predict qualified team) based on matrix $\mathbf{T}$.

Table 2. The attributes without feature selection and feature extraction

Id	Attribute Name	Id	Attribute Name	Id	Attribute Name
$\overline{\boldsymbol{D}}_{k\cdot 1}^{(\ell)}$	Possession %	$\overline{D}_{k\cdot 15}^{(\ell)}$	Goals Six Yard Box	$\overline{D}_{k\cdot 29}^{(\ell)}$	Total Aerial
$\overline{\boldsymbol{D}}_{k\cdot 2}^{(\ell)}$	Total Attempted Tackles	$\overline{D}_{k\cdot 16}^{(\ell)}$	Goals Penalty Area	$\overline{D}_{k\cdot 30}^{(\ell)}$	Successful Aerial %
$\overline{\boldsymbol{D}}_{k\cdot 3}^{(\ell)}$	Successful Tackles %	$\overline{D}_{k\cdot 17}^{(\ell)}$	Goals Out of Box	$\overline{D}_{k\cdot 31}^{(\ell)}$	Total Passes
$\overline{\boldsymbol{D}}_{k\cdot 4}^{(\ell)}$	Interception	$\overline{D}_{k\cdot 18}^{(\ell)}$	Goals Open Play	$\overline{D}_{k\cdot 32}^{(\ell)}$	Successful Passes %
$\overline{\boldsymbol{D}}_{k\cdot 5}^{(\ell)}$	Fouled	$\overline{D}_{k\cdot 19}^{(\ell)}$	Goals Counter	$\overline{D}_{k\cdot 33}^{(\ell)}$	Total Key Passes
$\overline{\boldsymbol{D}}_{k\cdot 6}^{(\ell)}$	Fouls	$\overline{D}_{k\cdot 20}^{(\ell)}$	Goals Set Piece	$\overline{D}_{k\cdot 34}^{(\ell)}$	Rating
$\overline{\boldsymbol{D}}_{k\cdot 7}^{(\ell)}$	Shots Out of Box	$\overline{D}_{k\cdot 21}^{(\ell)}$	Goals Penalty	$\overline{D}_{k\cdot 35}^{(\ell)}$	Attack Sides Left
$\overline{\boldsymbol{D}}_{k\cdot 8}^{(\ell)}$	Shots Six Yard Box	$\overline{D}_{k\cdot 22}^{(\ell)}$	Goals Normal	$\overline{D}_{k\cdot 36}^{(\ell)}$	Attack Sides Middle
$\overline{\boldsymbol{D}}_{k\cdot 9}^{(\ell)}$	Shots Penalty Area	$\overline{D}_{k\cdot 23}^{(\ell)}$	Goals Foot	$\overline{D}_{k\cdot 37}^{(\ell)}$	Attack Sides Right
$\overline{\boldsymbol{D}}_{k\cdot 10}^{(\ell)}$	Shots Off Target	$\overline{D}_{k\cdot 24}^{(\ell)}$	Goals Head	$\overline{D}_{k\cdot 38}^{(\ell)}$	Own Third Action Zone
$\overline{\boldsymbol{D}}_{k\cdot 11}^{(\ell)}$	Shots on Target	$\overline{D}_{k\cdot 25}^{(\ell)}$	Total Dribbles	$\overline{D}_{k\cdot 39}^{(\ell)}$	Middle Third Action Zone
$\overline{\boldsymbol{D}}_{k\cdot 12}^{(\ell)}$	Shots Blocked	$\overline{D}_{k\cdot 26}^{(\ell)}$	Successful dribbles %	$\overline{D}_{k\cdot 40}^{(\ell)}$	Opposition Third Action Zone
$\overline{\boldsymbol{D}}_{k\cdot 13}^{(\ell)}$	Shots Foot	$\overline{D}_{k\cdot 27}^{(\ell)}$	Unsuccessful Touches		
$\overline{\boldsymbol{D}}_{k\cdot 14}^{(\ell)}$	Shots Head	$\overline{D}_{k\cdot 28}^{(\ell)}$	Dispossessed		

In one season of the Champions League, 15 matchups are played after the group stages. There are a total of 135 matchups in 9 seasons between 2010 and 2018. Each of these matchups is considered as an observation in the study.

2.2.1. Data normalization

Since the mean and variances of the variables in the dataset are significantly different from each other, it may not be appropriate to classify the raw data. [52]. In order to make it more suitable for classification, firstly, the normalization process is performed on the dataset. Thanks to normalization methods such as Min-Max, Z-Score and Sigmoid in the literature, different alternatives of predictors can be used in the classification process. These normalization methods are introduced below.

Score-Z normalized independent variables are defined by

$$\mathbf{ZT} = \begin{pmatrix} \mathbf{ZT}^{(1)} \\ \mathbf{ZT}^{(2)} \\ \vdots \\ \mathbf{ZT}^{(9)} \end{pmatrix}_{135\times 40}, \tag{5}$$

where

$$\mathbf{ZT}^{(\ell)} = \begin{pmatrix} \overline{ZD}_{1\cdot j}^{(\ell)} & \overline{ZD}_{1\cdot 2}^{(\ell)} & \cdots & \overline{ZD}_{1\cdot 40}^{(\ell)} \\ \overline{ZD}_{2\cdot j}^{(\ell)} & \overline{ZD}_{2\cdot 2}^{(\ell)} & \cdots & \overline{ZD}_{2\cdot 40}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{ZD}_{15\cdot j}^{(\ell)} & \overline{ZD}_{15\cdot 2}^{(\ell)} & \cdots & \overline{ZD}_{15\cdot 40}^{(\ell)} \end{pmatrix}_{15\times 40}, \tag{6}$$

$$\overline{ZD}_{k\cdot j}^{(\ell)} = \frac{\overline{D}_{k\cdot j}^{(\ell)} - \overline{\overline{D}}_{k\cdot j}^{(\ell)}}{S_{\overline{D}_{k\cdot j}^{(\ell)}}}, 1 \le \ell \le 9, 1 \le k \le 15, \tag{7}$$

$$\overline{\overline{D}}_{k\cdot j}^{(\ell)} = \frac{1}{135}\sum_{\ell=1}^{9}\sum_{k=1}^{15}\overline{D}_{k\cdot j}^{(\ell)}, \tag{8}$$

and

$$S_{\overline{D}_{k\cdot j}^{(\ell)}} = \frac{1}{134}\sum_{\ell=1}^{9}\sum_{k=1}^{15}\left(\overline{D}_{k\cdot j}^{(\ell)} - \overline{\overline{D}}_{k\cdot j}^{(\ell)}\right)^2, \tag{9}$$

Min-Max normalized independent variables are defined by

$$\mathbf{MT} = \begin{pmatrix} \mathbf{MT}^{(1)} \\ \mathbf{MT}^{(2)} \\ \vdots \\ \mathbf{MT}^{(9)} \end{pmatrix}_{135\times 40}, \tag{10}$$

where

$$\mathbf{MT}^{(\ell)} = \begin{pmatrix} \overline{MD}_{1\cdot j}^{(\ell)} & \overline{MD}_{1\cdot 2}^{(\ell)} & \cdots & \overline{MD}_{1\cdot 40}^{(\ell)} \\ \overline{MD}_{2\cdot j}^{(\ell)} & \overline{MD}_{2\cdot 2}^{(\ell)} & \cdots & \overline{MD}_{2\cdot 40}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{MD}_{15\cdot j}^{(\ell)} & \overline{MD}_{15\cdot 2}^{(\ell)} & \cdots & \overline{MD}_{15\cdot 40}^{(\ell)} \end{pmatrix}_{15\times 40} \tag{11}$$

and

$$\overline{MD}_{k\cdot j}^{(\ell)} = \frac{\overline{D}_{k\cdot j}^{(\ell)} - \min\limits_{1\le\ell\le 9, 1\le k\le 15} \overline{D}_{k\cdot j}^{(\ell)}}{\max\limits_{1\le\ell\le 9, 1\le k\le 15} \overline{D}_{k\cdot j}^{(\ell)} - \min\limits_{1\le\ell\le 9, 1\le k\le 15} \overline{D}_{k\cdot j}^{(\ell)}}, \tag{12}$$

$$1 \le \ell \le 9, \quad 1 \le k \le 15$$

Sigmoid normalized independent variables are defined as,

$$\mathbf{ST} = \begin{pmatrix} \mathbf{ST}^{(1)} \\ \mathbf{ST}^{(2)} \\ \vdots \\ \mathbf{ST}^{(9)} \end{pmatrix}_{135\times 40}, \tag{13}$$

where

$$\mathbf{ST}^{(\ell)} = \begin{pmatrix} \overline{SD}_{1\cdot j}^{(\ell)} & \overline{SD}_{1\cdot 2}^{(\ell)} & \cdots & \overline{SD}_{1\cdot 40}^{(\ell)} \\ \overline{SD}_{2\cdot j}^{(\ell)} & \overline{SD}_{2\cdot 2}^{(\ell)} & \cdots & \overline{SD}_{2\cdot 40}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{SD}_{15\cdot j}^{(\ell)} & \overline{SD}_{15\cdot 2}^{(\ell)} & \cdots & \overline{SD}_{15\cdot 40}^{(\ell)} \end{pmatrix}_{15\times 40} \tag{14}$$

and

$$\overline{SD}_{k\cdot j}^{(\ell)} = \frac{\exp\left(\overline{D}_{k\cdot j}^{(\ell)}\right) - \exp\left(-\overline{D}_{k\cdot j}^{(\ell)}\right)}{\exp\left(\overline{D}_{k\cdot j}^{(\ell)}\right) + \exp\left(-\overline{D}_{k\cdot j}^{(\ell)}\right)}, \tag{15}$$

$$1 \le \ell \le 9, \; 1 \le k \le 15,$$

The Min-Max Normalization method provides the highest success rates and it is used according to the analyses discussed in Section 3.

2.2.2. Feature selection and extraction

Feature selection and feature extraction are used for datasets in order to increase accuracy rate of the classification process after performing normalization. It is aimed to create a lower-dimensional dataset that will be replaced with the current dataset. It is a two-way process: A small subset of the dimensions are chosen among the originals to represent datasets within the feature selection while extraction transforms the originals into new dimensional datasets. Since there is no difference between feature selection and extraction in our study, the correlation-based feature subset selection [53] is applied to 40 independent variables, and selected ones are given in Table 3.

Table 3. Predictors after applying future selection

j	Variable Name	j	Variable Name
$\boldsymbol{\overline{D}_{k\cdot 1}^{(\ell)}}$	Possession %	$\overline{D}_{k\cdot 22}^{(\ell)}$	Goals Normal
$\boldsymbol{\overline{D}_{k\cdot 7}^{(\ell)}}$	Shots Out Of Box	$\overline{D}_{k\cdot 30}^{(\ell)}$	Successful Aerial %
$\boldsymbol{\overline{D}_{k\cdot 11}^{(\ell)}}$	Shots On Target	$\overline{D}_{k\cdot 31}^{(\ell)}$	Total Passes

3. Applications and Results

In this section, classification applications are performed using ANN, KNN, LRA, NBC, RF and SVM on the data obtained from the Champions League tournaments between 2010 and 2018. The applications are divided into two separate phases. In the first application, data of eight seasons between 2010-2017 are used as training data, and 2018 season data are used as test data. In the second application, methods are tested with K-Fold Cross validity by using all data between 2010-2018. In two applications, tests are carried out both in the original form of the dataset and in its reduced dimension.

During the testing process, it benefited from the WEKA software tools and the libraries of the R programming language. Explorer and Knowledge Flow tools from WEKA software are used. R programming language libraries are used via RStudio, a popular IDE.

3.1. Results of the application 1

120 of the elimination stage matches between 2010 and 2017 are used as training data, and 15 of them in 2018 are used as test data in Application 1. The results are indicated separately based on the feature selection.

Table 4. Confusion matrix for Application 1

	ANN		KNN		LRA		NBC		RF		SVM	
	1	2	1	2	1	2	1	2	1	2	1	2
Without Feature Selection	7	3	4	3	4	3	5	2	5	2	5	2
	0	5	1	4	4	4	0	8	1	7	2	6
With Feature Selection	5	1	6	1	5	2	5	2	5	2	4	3
	2	7	2	6	3	5	0	8	2	6	0	8

The confusion matrices obtained in Application 1 are shown in Table 4. There was no change in TP and TN values after feature selection in ANN and NBC methods. TP and TN values increased after feature selection in KNN, LRA and SVM methods. In the RF method, there is a decrease. When TN and TP values are examined separately, there is a decrease in the TN value after the property selection in ANN, while an increase is observed in the TP value, which compensated this decrease. Since the increase in TP value in SVM is more than the decrease in TN value, there is an increase in total. In the KNN and LRA methods, the amount of increase in TN and TP values are equal.

Table 5. The model parameters for Application 1

Model	Parameter	Without Feature Selection	With Feature Selection
ANN	hidden layers	7,10,3	8.12.14
	learning rate	0.3	0.1
	momentum	0.2	0.2
KNN	*k*	1	4
RF	iteration	100	100
	seed	1	2
SVM	kernel	PUK Kernel	PUK Kernel
	sigma,omega, *c*	1	1

The parameters of the related models are changed through the trial-and-error method during the application process, and the outcome is observed in terms of prediction accuracy. The most reliable parameters are given in Table 5.

Table 6. Performance measures of the classifiers for Application 1

	Method	Class	TP Rate	FP Rate	Precision	F-Measure	MCC	Kappa St.	Accuracy
Without Feature Selection	ANN	1	1.000	0.375	0.700	0.824	0.661	0.6087	%80.00
		2	0.625	0.000	1.000	0.769	0.661		
	KNN	1	0.571	0.125	0.800	0.667	0.472	0.4545	%73.33
		2	0.875	0.429	0.700	0.778	0.472		
	LRA	1	0.571	0.500	0.500	0.533	0.071	0.0708	%53.33
		2	0.500	0.429	0.571	0.533	0.071		
	NBC	1	0.714	0.000	1.000	0.833	0.756	0.7273	%86.66
		2	1.000	0.286	0.800	0.889	0.756		
	RF	1	0.714	0.125	0.833	0.769	0.600	0.5946	%80.00
		2	0.875	0.286	0.778	0.824	0.600		
	SVM	1	0.714	0.250	0.714	0.714	0.464	0.4643	%73.33
		2	0.750	0.286	0.750	0.714	0.464		
With Feature Selection	ANN	1	1.000	0.375	0.700	0.824	0.661	0.6087	%80.00
		2	0.625	0.000	1.000	0.769	0.661		
	KNN	1	0.857	0.250	0.750	0.800	0.607	0.6018	%80.00
		2	0.750	0.143	0.857	0.800	0.607		
	LRA	1	0.714	0.375	0.625	0.667	0.339	0.3363	%66.66
		2	0.625	0.286	0.714	0.667	0.339		
	NBC	1	0.714	0.000	1.000	0.833	0.756	0.7273	%86.66
		2	1.000	0.286	0.800	0.889	0.756		
	RF	1	0.714	0.250	0.714	0.714	0.464	0.4643	%73.33
		2	0.750	0.286	0.750	0.750	0.464		
	SVM	1	0.571	0.000	1.000	0.727	0.645	0.5872	%80.00
		2	1.000	0.429	0.727	0.842	0.645		

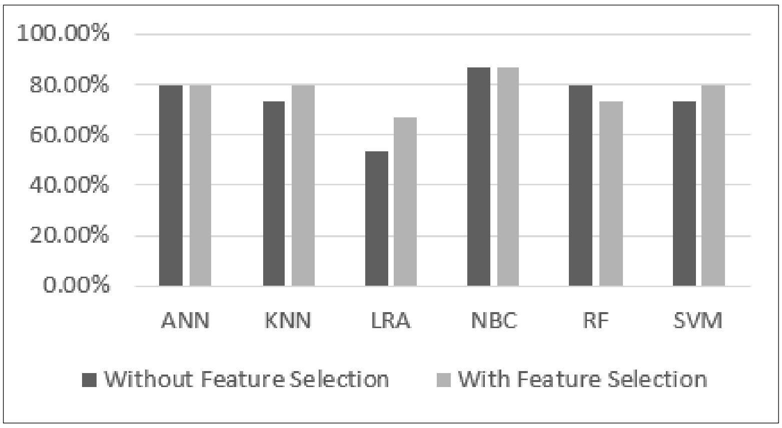

Figure 2. Accuracy change after feature selection for Application 1

The performance measures obtained in Application-1 are shared in Table 6 and Fig 2. Before and after feature selection, the most successful prediction rates are reached by the NBC method with 86.66%. NBC is followed by ANN and RF with 80% before feature selection, followed by ANN, KNN and SVM with 80% after feature selection. While the prediction success of KNN and SVM increased after feature selection, the successes of NBC and ANN decreased. Although LRA has increased its success from 53.33% to 66.66% after feature selection, it is the most unsuccessful prediction method of Application-1.

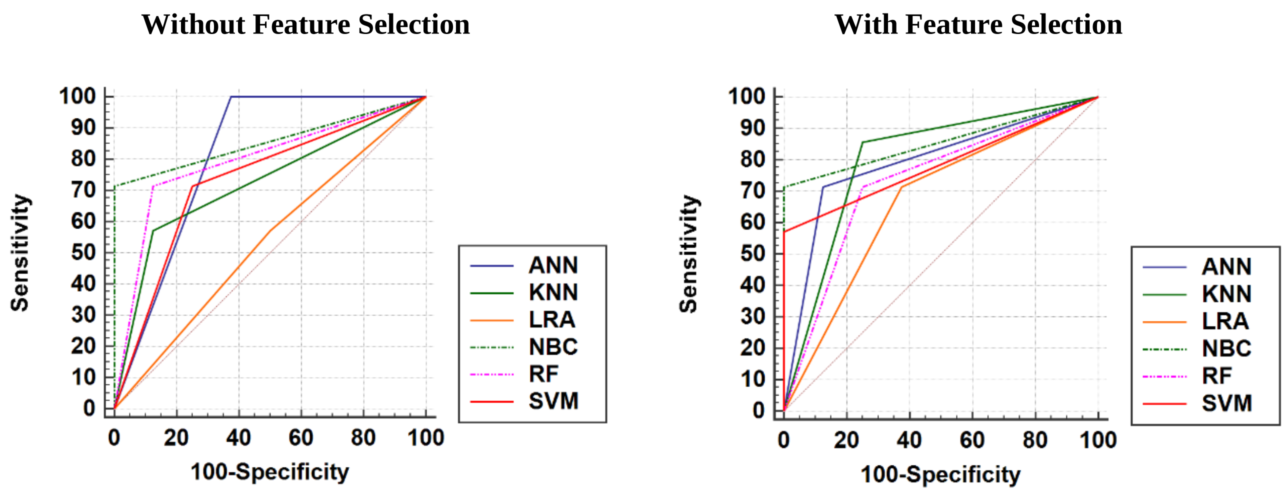

Figure 3. The roc curve graphs for Application 1

Table 7. Area under the curve (AUC) and confidence intervals for Application 1

Method	Without Feature Selection			With Feature Selection		
	AUC	SE [a]	95% CI [b]	AUC	SE [a]	95% CI [b]
ANN	0.813	0.0915	(0.533, 0.962)	0.795	0.111	(0.513, 0.954)
KNN	0.723	0.119	(0.439, 0.916)	0.804	0.109	(0.523, 0.958)
LRA	0.536	0.138	(0.268, 0.789)	0.670	0.130	(0.387, 0.884)
NBC	0.857	0.0922	(0.584, 0.980)	0.857	0.0922	(0.584, 0.980)
RF	0.795	0.111	(0.513, 0.954)	0.732	0.123	(0.448, 0.921)
SVM	0.732	0.123	(0.448, 0.921)	0.786	0.101	(0.504, 0.950)

[a][54] [b] Binomial exact

The ROC curves resulting in Application - 1 are shown in Fig 3. The values for the areas under the ROC curves are given in Table 7. The most successful method before and after feature selection is NBC (AUC = 0.857). The second most successful method before feature selection is ANN (AUC = 0.813). After feature selection, the second most successful method is KNN (AUC = 0.804).

Table 8. Predictions for Application 1

		Actual Result														
		Bayern Munich Besiktas	Roma Shakhtar Donetsk	Manchester Unt. Sevilla	Juventus Tottenham	FC Basel 1893 Manchester City	Paris St-Germain Real Madrid	FC Porto Liverpool	Barcelona Chelsea	Juventus Real Madrid	Bayern Munich Sevilla	Liverpool Manchester City	Barcelona Roma	Bayern Munich Real Madrid	Liverpool Roma	Liverpool Real Madrid
		1	**1**	**2**	**1**	**2**	**2**	**2**	**1**	**2**	**1**	**1**	**2**	**2**	**1**	**2**
Without Feature Selection	ANN	1	1	1	1	2	2	2	1	2	1	1	1	2	1	1
	KNN	1	2	2	2	2	2	2	1	2	2	1	1	2	1	2
	LRA	2	1	1	2	2	1	1	1	2	2	1	2	2	1	1
	NBC	1	1	2	2	2	2	2	2	2	1	1	2	2	1	2
	RF	1	1	2	2	2	2	2	2	2	1	1	2	2	1	1
	SVM	1	1	2	2	2	1	2	2	2	1	1	2	2	1	1
With Feature Selection	ANN	1	1	2	2	2	2	2	2	2	1	1	1	2	1	2
	KNN	1	1	2	2	2	1	2	1	2	1	1	2	2	1	1
	LRA	1	1	1	2	2	1	2	2	2	1	1	2	2	1	1
	NBC	1	1	2	2	2	2	2	2	2	1	1	2	2	1	2
	RF	1	1	1	2	2	2	2	2	2	1	1	2	2	1	1
	SVM	1	1	2	2	2	2	2	2	2	1	2	2	2	1	2

Correct Prediction   Incorrect Prediction

Table 8 contains the estimates of the methods used for 15 test matches and the comparison of actual results. The prediction results obtained before and after feature selection are presented separately. Table 8 contains remarkable details. While all methods before and after feature selection failed in the match between Juventus and Tottenham, ANN successfully predicted this matching result. While all methods successfully predicted the match between Bayern Munich and Besiktaş before and after feature selection, LRA failed before feature selection. In the match between Roma and Shakhtar Donetsk, all methods before and after feature selection made a successful prediction but failed before KNN feature selection.

3.2. Results of the application 2

In Application 2, the tests are carried out with the K-fold cross-validation method. K value is taken as 9 to represent the number of seasons. 135 qualifying matches between 2010 and 2018 are divided into 9 parts, 120 of which are used as training and the remaining 15 as test data. By taking the average of the performance criteria obtained in 9 tests performed in this way, the values obtained in Application-2 are calculated.

Table 9. Confusion matrix for Application 2

	ANN		KNN		LRA		NBC		RF		SVM	
	1	1	1	2	1	2	1	2	1	2	1	2
Without Feature Selection	44	21	41	24	43	22	45	20	40	25	39	26
	27	43	18	52	31	39	25	45	21	49	16	54
With Feature Selection	40	25	48	17	41	24	50	15	46	19	42	23
	19	51	22	48	25	45	24	46	15	55	18	52

The confusion matrix obtained for Application 2 is given in Table 9. From Table 9, it can be concluded that NBC reaches the highest TN value before applying feature selection while SVM reaches the highest TP. Similarly, as in this case, NBC achieves the highest TN value while it is SVM for TN after applying the selection.

Table 10. The model parameters for Application 2

Model	Parameter	Without Feature Selection	With Feature Selection
ANN	hidden layers	7,2,4	8,2,7
	learning rate	0.3	0.3
	momentum	0.2	0.2
KNN	k	39	45
RF	iteration	500	1000
	seed	100	100
SVM	kernel	PUK Kernel	PUK Kernel
	sigma,omega, c	1	1

Model parameters used for Application 2 are given in Table 10.

Table 11. Performance measures of the classifiers for Application 2

	Method	Class	TP Rate	FP Rate	Precision	F-Measure	MCC	Kappa St.	Accuracy
Without Feature Selection	ANN	1	0.677	0.386	0.620	0.647	0.291	0.2903	%64.44
		2	0.614	0.323	0.672	0.642	0.291		
	KNN	1	0.631	0.257	0.695	0.661	0.376	0.3749	%68.89
		2	0.743	0.369	0.684	0.712	0.376		
	LRA	1	0.662	0.443	0.581	0.619	0.220	0.2176	%60.74
		2	0.557	0.338	0.639	0.595	0.220		
	NBC	1	0.692	0.357	0.643	0.667	0.335	0.3342	%66.66
		2	0.643	0.308	0.692	0.667	0.335		
	RF	1	0.615	0.300	0.656	0.635	0.317	0.3161	%65.93
		2	0.700	0.385	0.662	0.681	0.317		
	SVM	1	0.600	0.229	0.709	0.650	0.378	0.3735	%68.89
		2	0.771	0.400	0.675	0.720	0.378		
With Feature Selection	ANN	1	0.615	0.271	0.678	0.645	0.346	0.3451	%67.40
		2	0.729	0.385	0.671	0.699	0.346		
	KNN	1	0.738	0.314	0.686	0.711	0.424	0.4230	%71.11
		2	0.686	0.262	0.738	0.711	0.424		
	LRA	1	0.631	0.357	0.621	0.626	0.274	0.2735	%63.70
		2	0.643	0.369	0.652	0.647	0.274		
	NBC	1	0.769	0.343	0.676	0.719	0.428	0.4243	%71.11
		2	0.657	0.231	0.754	0.702	0.428		
	RF	1	0.708	0.214	0.754	0.730	0.495	0.4945	%74.81
		2	0.786	0.292	0.743	0.764	0.495		
	SVM	1	0.646	0.257	0.700	0.672	0.391	0.3901	%69.63
		2	0.743	0.354	0.693	0.717	0.391		

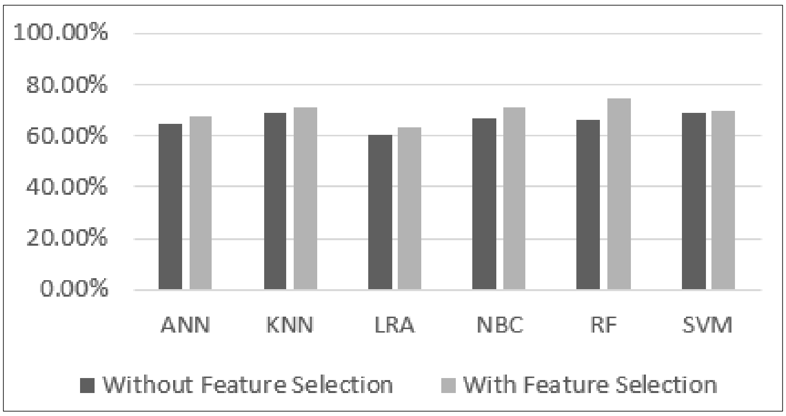

Figure 4. Accuracy change after feature selection for Application 2

The highest accuracy rate is 68.89% with KNN and SVM before applying feature selection, according to Table 11 and Fig 4, demonstrating the performance measures for Application 2. RF is the most effective method, with 74.81%, while KNN and NBC reach 71.11%.

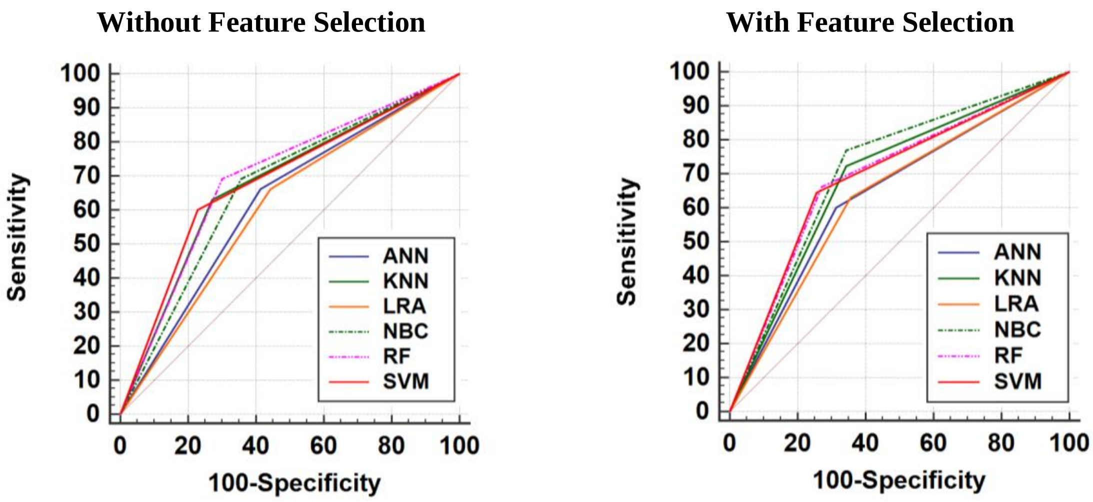

Figure 5. The roc curve graphs for Application 2

Table 12. Area under the curve (AUC) and confidence intervals for Application 2

Method	Without Feature Selection			With Feature Selection		
	AUC	SE [a]	95% CI [b]	AUC	SE [a]	95% CI [b]
ANN	0.624	0.0419	0.536, 0.705	0.643	0.0415	0.556, 0.723
KNN	0.680	0.0403	0.594, 0.757	0.690	0.0400	0.605, 0.767
LRA	0.609	0.0421	0.522, 0.692	0.637	0.0417	0.550, 0.718
NBC	0.668	0.0408	0.581, 0.746	0.713	0.0389	0.629, 0.788
RF	0.696	0.0399	0.611, 0.772	0.695	0.0399	0.610, 0.771
SVM	0.686	0.0397	0.600, 0.763	0.695	0.0398	0.609, 0.771

[a][54] [b] Binomial exact

The ROC curves obtained in Application 2 are given in Fig 5, and AUC values and accuracy ratio are given in Table 12. From Figure 5 and Table 12 are examined, it can be seen that RF reaches the highest AUC value before applying feature selection while it is NBC afterward.

4. Discussion and Future Studies

For many years, football is interpreted through numerical values that contain various information about teams and players. The use of statistical data so much causes an increase in the number of scientific researches based on data in football. One of the frequently studied topics in these scientific studies is the prediction of match results. Our study is an example of these studies. The match results are predicted using six different machine learning algorithms. These methods are Artificial Neural Network, K-Nearest Neighbors, Logistic Regression Analysis, Naive Bayes Classifier, Random Forest and Support Vector Machine. The study dataset's scope is the matches played in the UEFA Champions League between 2010-2018. With the information obtained from these matches, a dataset containing forty features is obtained. Two different applications are carried out to test the performance of the created models. In the first application, the last season matches are used as test data and the rest as training data. In the second application, each season's matches are separated, and tests are carried out with the K-Fold Cross-validation method. In both applications, the tests are repeated by performing dimension reduction on the dataset. The comparison of the results achieved is shown in Table 13.

Table 13. Success average of the methods

	Application 1			Application 2		
	FS-*	FS+*	Change	FS-*	FS+*	Change
ANN	80	80	0	64.44	67.4	2.96
KNN	73.33	80	6.67	68.89	71.11	2.22
LRA	53.33	66.66	13.33	60.74	63.7	2.96
NBC	86.66	86.66	0	66.66	71.11	4.45
RF	80	73.33	-6.67	65.93	74.81	8.88
SVM	73.33	80	6.67	68.89	69.63	0.74

FS-: Without Feature Selection, FS+: With Feature Selection

The most successful method for each form of the dataset is NBC (86.66%) in the first application. Then ANN and KNN are the most successful methods (80%). In the second application, KNN and SVM are the most successful methods for the first form of the dataset (68.89%). RF is the most successful method for dimension reduced dataset (74.81%). It is the method in which the lowest success rates are obtained in logistic regression tests. The dimension reduction process performed on the dataset is increased the prediction successes.

In order to increase the prediction successes obtained in the study, enrichment can be performed both in the study dataset and in the methods used. Although football is defined as a sport branch that can be explained by statistics, there are also variables that affect the match results but are not taken into account. Variables such as injuries, suspensions, the player and technical staff's motivation, climatic conditions, instant developments, and transfers that are ignored in most scientific studies can also directly affect the match results. There are also variables that do not directly affect the match results but can be useful in the prediction process. For example, fan opinions, social media posts, bet odds, football news, expert opinions. For these reasons, instead of using a dataset consisting of only the performances of the teams and players, a large dataset can be used, which also includes variables such as injuries, suspensions, fan opinions, bet odds. In addition to enriching the dataset content, the number of methods used can also be increased. Hybrid classifiers can also be used in addition to new classification methods.

Acknowledgment.

This study is a part of PhD thesis of Ismail Hakkı KINALIOGLU titled "Prediction of Football Match Results by Using Artificial Intelligence Based Methods and Proposals on Hybrid Model", which was completed at the Selcuk University, Graduate School of Natural and Applied Sciences, Department of Statistics, Konya, Turkey.

Conflicts of interest

There is no conflict of interest.

References

[1] Caraffa, A., Cerulli G., Projetti M., Aisa G. and Rizzo, A., Prevention of anterior cruciate ligament injuries in soccer, *Knee Surgery, Sports Traumatology, Arthroscopy,* 4 (1) (1996) 19-21.

[2] Heidt R. S., Sweeterman L. M., Carlonas R. L., Traub J. A. and Tekulve F. X., Avoidance of Soccer Injuries with Preseason Conditioning, *The American Journal of Sports Medicine,* 28 (5) (2000) 659-662.

[3] Houston R. G., Wilson D. P., Income, leisure and proficiency: an economic study of football performance, *Applied Economics Letters*, 9 (14) (2002) 939-943.

[4] Junge A., Dvorak J., Soccer Injuries, *Sports Medicine,* 34 (13) (2004) 929-938.

[5] Decrop A., Derbaix C., Pride in contemporary sport consumption: a marketing perspective, *Journal of the Academy of Marketing Science*, 38 (5) (2010) 586-603.

[6] Goddard J., Regression models for forecasting goals and match results in association football, *International Journal of forecasting,* 21 (2) (2005) 331-340.

[7] Huang K.-Y., Chen K.-J., Multilayer perceptron for prediction of 2006 world cup football game, *Advances in Artificial Neural Systems,* 2011 (2011) 11.

[8] Timmaraju A. S., Palnitkar A. and Khanna V., Game ON! Predicting English Premier League Match Outcomes, (2013).

[9] Ulmer B., Fernandez M. and Peterson M., *Predicting Soccer Match Results in the English Premier League*, Doctoral dissertation, Ph. D. dissertation, Stanford, (2013).

[10] Lata K., Gupta P., Forecasting English Premier League Match Results, *International Journal of Modern Trends in Engineering and Research,* 03 (02) (2016) 261-268.

[11] Igiri C. P., Support Vector Machine—Based Prediction System for a Football Match Result, *IOSR Journal of Computer Engineering,* 17 (3) (2015) 21-26.

[12] Baboota R., Kaur H., Predictive analysis and modelling football results using machine learning approach for English Premier League, *International Journal of Forecasting,* 35 (2) (2019) 741-755.

[13] Rahman M. A., A deep learning framework for football match prediction, *SN Applied Sciences,* 2 (2) (2020) 165.

[14] Rotshtein A. P., Posner M., Rakityanskaya A., Football predictions based on a fuzzy model with genetic and neural tuning, *Cybernetics and Systems Analysis,* 41 (4) (2005) 619-630.

[15] Joseph A., Fenton N. E. and Neil M., Predicting football results using Bayesian nets and other machine learning techniques, *Knowl-Based Syst*, 19 (7) (2006) 544-553.

[16] Hvattum L. M., Arntzen H., Using ELO ratings for match result prediction in association football, *International Journal of forecasting,* 26 (3) (2010) 460-470.

[17] Owen A., Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter, *IMA Journal of Management Mathematics*, 22 (2) (2011) 99-113.

[18] Constantinou A. C., Fenton N. E. and Neil M., pi-football: A Bayesian network model for forecasting Association Football match outcomes, *Knowl-Based Syst,* 36 (2012) 322-339.

[19] Igiri C. P., Nwachukwu E. O., An improved prediction system for football a match result, *IOSR Journal of Engineering,* 4 (2014) 12-20.

[20] Koopman S. J., Lit R. A., dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League, *Journal of the Royal Statistical Society: Series A,* 178 (1) (2015) 167-186.

[21] Amadin F., Obi, J. C.i English Premier League (EPL) Soccer Matches Prediction using An Adaptive Neuro-Fuzzy Inference System (ANFIS), *Transactions on Machine Learning and Artificial Intelligence*, 3 (2) (2015) 34.

[22] Robertson S., Back N. and Bartlett J. D., Explaining match outcome in elite Australian Rules football using team performance indicators, *Journal of Sports Sciences,* 34 (7) (2016) 637-644.

[23] Gevaria K., Sanghavi H., Vaidya S. and Deulkar K., Football Match Winner Prediction. (2015).

[24] Prasetio D., Dra. Harlili M. S., in Advanced Informatics: Concepts, Theory And Application, International Conference On., 2016, 1-5.

[25] Martins R. G. *et al.,* Exploring polynomial classifier to predict match results in football championships, *Expert Syst. Appl.*, 83 (2017) 79-93.

[26] Bunker R. P., Thabtah F., A machine learning framework for sport result prediction, *Applied Computing and Informatics,* 15 (1) (2017) 27-33.

[27] Mennis J., Guo D., Spatial data mining and geographic knowledge discovery—An introduction, *Computers, Environment and Urban Systems*, 33 (6) (2009) 403-408.

[28] Aggarwal C. C., Data classification, algorithms and applications, New York: CRC Press, 2014.

[29] Elmas Ç., Yapay zeka uygulamaları: (yapay sinir ağı, bulanık mantık, genetik algoritma), İstanbul: Seçkin Yayıncılık, 2010.

[30] Haykin S., Neural networks: a comprehensive foundation, Prentice Hall PTR, 1994.

[31] Peterson L. E., K-nearest neighbor. *Scholarpedia,* 4 (2) (2009) 1883.

[32] Fix E., Hodges Jr, J. L., Discriminatory analysis-nonparametric discrimination: consistency properties, California Univ Berkeley, 1951.

[33] Cover T. M., Hart P. E., Nearest neighbor pattern classification, *IEEE transactions on information theory*, 13 (1) (1967) 21-27.

[34] Song Y., Huang J., Zhou D., Zhan H. and Giles C. L., in European Conference on Principles of Data Mining and Knowledge Discovery, 248-264.

[35] Cramer J. S., in *T*inbergen Institute Discussion Paper, No. 02-119/4, Tinbergen Institute, Amsterdam and Rotterdam, 2002.

[36] Hosmer D. W., Lemeshow S., Applied Logistic Regression, New York: John Wiley & Sons, 2000.

[37] Özdamar K., Paket programlar ile istatistiksel veri analizi (çok değişkenli analizler), Kaan Kitabevi, 2004).

[38] Zhang H. The optimality of naive Bayes, *AA* 1 (2) (2004) 3.

[39] Kantarcıoglu M., Vaidya J. and Clifton C. in IEEE ICDM workshop on privacy preserving data mining, 3-9.

[40] Rish I., in IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. 41-46.

[41] Ho T. K, in Proceedings of 3rd international conference on document analysis and recognition. 278-282 (IEEE).

[42] Breiman L., Random Forests, UC Berkeley TR567 (1999).

[43] Breiman L., Random Forests, Mach Learn 45 (1) (2001) 5-32.

[44] Schölkopf B., Smola A. J. and Bach F., Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2002.

[45] Kavzoğlu T., Çölkesen İ., Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi, *Harita Dergisi* 144 (7) (2010) 73-82.

[46] Yang Z.-X., Shao Y.-H. and Zhang X.-S., Multiple birth support vector machine for multi-class classification, *Neural Computing and Applications*, 22 (1) (2013) 153-161.

[47] Crammer K. and Singer Y., On the algorithmic implementation of multiclass kernel-based vector machines, *Journal of machine learning research* ,2 (Dec) (2001) 265-292.

[48] Platt J. C., Cristianini N. and Shawe-Taylor J., in Advances in neural information processing systems. 547-553.

[49] Dietterich T. G., Bakiri G., Solving multiclass learning problems via error-correcting output codes, *Journal of artificial intelligence research*, 2 (1994) 263-286.

[50] Crammer K., Singer Y., On the learnability and design of output codes for multiclass problems, *Mach Learn.*, 47 (2-3) (2002) 201-233.

[51] Allwein E. L., Schapire R. E. and Singer Y., Reducing multiclass to binary: A unifying approach for margin classifiers, *Journal of machine learning research*, 1 (2000) 113-141.

[52] Özkan Y., Veri madenciliği yöntemleri, Papatya Yayıncılık Eğitim, 2016.

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

[54] DeLong, E. R., DeLong D. M. and Clarke-Pearson D. L., Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics*, 44 (3) (1988) 837-845