Decision Analytics Journal 14 (2025) 100537

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A predictive analytics framework for forecasting soccer match outcomes using machine learning models

Albert Wong [a,*], Eugene Li [a], Huan Le [a], Gurbir Bhangu [a], Suveer Bhatia [b]

[a] Department of Statistics and Data Analytics, Langara College, Vancouver, British Columbia, Canada, V5Y 2Z6
[b] School of Computing, University of the Fraser Valley, Abbotsford, British Columbia, Canada, V2S 7M8

ARTICLE INFO

Keywords:
Predictive analytics
Machine learning
Convolutional Neural Network
Ensemble technique
Sports analytics
Forecasting

ABSTRACT

Predicting the outcome of a sports game is a favourite pastime for sports fans and researchers. The interest has intensified in recent years due to data availability, the development and successful implementation of machine learning algorithms, and the proliferation of internet gaming. This research focuses on developing a predictive analytics framework using machine learning or artificial intelligence models, as well as publicly available game results and weather data, to accurately predict outcomes of games in the English Premier League. Development efforts include experimentation using weather data and constructs such as fatigue and momentum. Ensemble techniques such as stacking or voting are also explored to improve the accuracy of basic machine learning models. The results are compared with those derived from the odds given by the major bookmakers to gauge the usefulness and potential applications in sports betting.

1. Introduction

Predicting the outcome of a sports game is a favourite pastime for sports fans and researchers. With the advent of the Internet and the proliferation and availability of game data, research on how to use quantitative techniques and, more recently, machine learning or artificial intelligence algorithms to improve prediction has intensified. Interest is also fuelled by the increase in sports betting through various gaming websites.

It is well recognized that a profitable betting strategy is based on the optimal use of odds data from the bookmakers and accurate predictions of match outcomes. Following the footsteps of recent research identified in the following section, this research focuses on developing a machine learning (ML) or artificial intelligence (AI) model to accurately predict match outcomes. The research uses publicly available game results data for the English Premier League (EPL) to train various models to predict the outcome of a soccer match in the League.

The main contributions of the research are twofold. We use weather data alongside team-based box score data, as well as features such as fatigue and momentum, to enhance predictive models. On the other hand, the research deploys several machine learning algorithms to develop initial prediction models. These models are combined using an ensemble technique such as Stacking or Voting to produce a final model that delivers more accurate results. We believe that using this approach is promising in predicting game outcomes.

This paper is organized as follows: Section 2 reviews the research on the prediction of the outcomes of soccer games and other sports outcomes. Section 3 presents the methodology used in this research. Section 4 presents results using several machine learning models and ensemble techniques. Section 5 includes discussions of the results and suggestions for future work. We close with a brief Conclusion section.

2. Existing work

The use of ML algorithms to develop sports betting strategies has increased in the last three years. For example, these papers [1–9] were published in 2023. Several overviews can also be found in [10] or [8]. The survey paper by Tokic et al. [11] is also useful, though it was published in 2022.

2.1. Using machine learning models for Predicting Soccer Game Outcomes

Using ML algorithms to predict the outcomes of a soccer match dates back to the mid-2010s. For example, using a data set extracted from 110 games played in the 2014–2015 English Premier League season with features that include Home and Away goals, Home and Away shots, Home and Away corners, Home and Away Odds, Home and Away attack strength, Home and Away Players' performance index, Home and

* Correspondence to: Department of Statistics and Data Analytics, Langara College, 100 W 49th Ave, Vancouver, BC, Canada, V5Y 2Z6 .
E-mail address: alwong@langara.ca (A. Wong).

https://doi.org/10.1016/j.dajour.2024.100537
Received 16 October 2024; Received in revised form 8 December 2024; Accepted 16 December 2024
Available online 20 December 2024

Away Managers' performance index, Home and Away managers' win, as well as Home and Away streak, Igiri et al. [12], in 2014, implemented a logistic Regression (LR) and an Artificial Neural Network (ANN) model to predict the outcome of 20 matches in 2014–2015. Results of the research showed that an accuracy rate (not explicitly defined in the paper) of 85% with ANN and 93% with LR is possible. One interesting aspect of the approach by the authors was using a genetic algorithm to "calculate the weights of the features" of the data set before input to ML algorithms.

Constantinou in 2019 [13] described an ambitious project to build Dolores, "a model designed to predict the results of football matches in one country by observing matches in multiple other countries". This model used a mixture of dynamic ratings and the hybrid Bayesian Networks approach to train a single data set that incorporates match outcomes from 52 football leagues worldwide and was used to predict 206 future match outcomes from 26 different leagues from March 31 to April 9 in 2017. Dolores placed 2nd in the international special issue competition, Machine Learning for Soccer.

Carloni et al. in 2021 [14] presented the design and implementation of ML models to predict the outcome of soccer matches for sports betting. Data were extracted from the Web through a scraping process and used as input to a comprehensive suite of ML algorithms such as Logistic Regression (LR), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), and a Four-layer Artificial Neural Network (ANN). Based on the accuracy criteria (not clearly defined in the paper), the ANN model developed is the best for the 2.5 market under. Further experiments also showed "encouraging performance in terms of the Return-on-Investment (ROI) criteria".

Wheatcroft and Sienkiewicz [15] examined, also in 2021, the need to calibrate probability forecasts of soccer matches produced by various ML models. The authors found that it is essential to evaluate the performance of the models with and without calibration to ensure that the extra step is indeed required. Using the forecasts and two betting strategies, the authors demonstrated how the hyperparameters of ML models could be selected to optimize the betting or not-betting decisions.

Nivetha et al. [16]in 2022 explored using a deep learning framework to forecast soccer outcomes. The project team used a long-short-term memory (LSTM) algorithm, a modified version of a recurring neural network (RNN). Using data from the data sets of the "International Football results from 1872 to 2018", "FIFA World Cup 2018" and the "FIFA Soccer Rankings", Rahman [17] also used an LSTM deep learning network, but a different data set to predict the outcome of the matches in the group stage of the FIFA World Cup 2018 and achieve an accuracy rate of 63.3%. The author believes that "this accuracy can be increased with proper datasets and more accurate information of the teams. For more accurate prediction performance, prior and more information about each team, player, and match is desirable".

Peters and Pacheco [18] used the lineup statistics of the EPL teams from 2020 to 2022 to build heuristic and ML models to predict soccer scores. The research found that the Support Vector Regression (SVR) algorithm outperforms others in predicting the final scores. Moreover, the researchers managed to "predict correctly all relegated teams after forecast 100 consecutive matches" and using the forecasts was profitable (42% return). Interestingly, the researchers concluded that "lineups do not improve predictions".

The research by Yao et al. [19] focused on developing a "two-stage model for soccer in-game outcome prediction (IGSOP)". The idea was that when the duration of a soccer game is divided into sufficiently small time intervals, the goals scored by the participating team in each interval could be considered as a Bernoulli random variable. The IGSOP used a state-based machine learning method to predict the probability of an event of a goal in a future time interval. The IGSOP model then simulated "the remainder of the game to estimate the outcome of a game". Data from the Chinese Super League was used for model training and evaluations. The authors stated that the results of the study "demonstrate that IGSOP outperforms existing methods, especially in predicting draws and prediction during the final moments of games".

Rodrigues and Pinto [20] developed several ML models using historical endgame statistics and player attributes for the two teams involved in the games to predict the results of soccer matches. The experimental results of the study showed encouraging performance "in terms of the profit margin of the football bets".

Mattera [21] proposed a simple statistical framework in 2023 and obtained accurate forecasts of binary outcomes in a soccer match. As illustrations, the researcher conducted experiments in the English Premier League and the Italian Serie A League and predicted the occurrence of red cards, under/over, and Goal/Goal events. It seems that this framework could be extended to predict the win/ no-win outcome of a soccer match.

Atta Mills et el. [22] in 2024 introduced a novel framework for the prediction of the outcome of soccer games in the Dutch Eredivisie League, the Scottish Premiership, and the Belgian Jupiler Pro League using algorithms of ML and deep learning (DL). Models built with the Logistic Regression, XGBoost, Random Forest, SVM, Naive Bayes, Feed-forward Neural Network, and the Recurrent Neural Network algorithm were evaluated using performance measures such as accuracy, recall, precision, F1 score, and area under the ROC Curve. The distinctive feature of this modelling exercise was the use of real-time features such as half-time results and goals in addition to end-of-game features. They also found that a voting model that integrates the Random Forest and XGBoost algorithms provides higher accuracy.

A Convolutional Neural Network (CNN) is a special deep learning algorithm designed for classification and other pattern detection tasks such as image classification. Several research projects in the last three years have used this algorithm to predict the outcomes of sports events and soccer games. Wagenaar et al. [23] were the first research team to use the CNN algorithm in the sports prediction arena. Instead of predicting the outcome of a soccer match, the team uses a "deep CNN" to predict goal scoring opportunities in soccer using position data from "29 matches played by a German Bundesliga team". Although their ideas are exciting, they cannot be used directly to predict the outcome of a soccer match.

On the other hand, Chen [24] used three typical machine learning algorithms, CNN, RF, and SVM, to develop models to predict the result of a football match based on a "player ability index" derived from data found on the website of the International Federation of Association Football (FIFA). The accuracy of these three methods is between 54% and 58%. The researcher argued that these results "are acceptable, since they are all higher than the prediction accuracy of the famous BBC football analyst, Mark Lawrenson, which is only 52%. In addition, the accuracy of the convolutional neural network is higher than the prediction accuracy of the authoritative football gambling organization Pinnacle Sports, which is only 55%". In any case, the research concluded that the four-layer CNN algorithm produced the best results.

Randrianasolo [25] used CNN to develop models to predict the 2020 men's EURO and 2022 women's EURO, resulting "in a 69.8% accuracy for predicting the 2020 men's EURO and an 80% accuracy for predicting the 2022 women's EURO". The research claimed that these results "represent a 2% improvement in predicting the 2020 men's EURO and a 10% improvement in predicting the 2022 women's EURO"

Based on these results, the CNN algorithm appears to be a viable alternative to developing models to predict the outcome of a soccer match. The other side of the coin would be the structure of the network, as well as the selection and engineering of features that should be fed to the algorithm.

For women's soccer, Grollet al. [26] used several ML algorithms to predict the FIFA Women's World Cup 2019 outcome.

2.2. Feature selection

The choice of features used as input to ML models to predict the outcomes of sports games is very diverse. The features could differ depending on the game itself, the League within which it is played, and the ML algorithms employed.

As mentioned, Chen used a player ability index as input to the predictive models [24]. Peters et al. [18] analysed "the role of lineups in the final scores using machine learning prediction models" they developed. Historical data and machine learning models were used to create these models to predict the outcome of English Premier League (EPL) games between 2020 and 2022. Through the analysis of features used, they found that statistics on goalkeepers are more critical than those on attackers in predicted goals scored. Two other findings of interest are that data on lineups do not improve predictions and that Support Vector Regression outperforms other algorithms in predicting final scores.

Rose et al. [27] explored different machine learning techniques to predict the score and outcome of soccer matches using custom-generated features. The features are used to train the selected algorithms.

According to Zimmermann et al. [28], the features are more important than the algorithms used to predict the outcome of a basketball game of the National Collegiate Athletic Association (NCAA). This sentiment is echoed in the paper by Iskandaryan et al. [29], which stated that an accurate prediction of soccer game outcomes depends on the features included as much as the algorithm used. For example, the authors pointed to the work by Kampakis and Adamides [30] that showed that Twitter integration could significantly improve the final results. Iskandaryan et al. then analysed several research studies [31–37] that related weather conditions to different features of a sports game, such as successful passes, injuries, etc. From these studies, the authors used data from five seasons of Spanish La Liga and Segunda division games between 2013–2014 and 2017–2018 and weather data to answer the research question: "How do weather conditions affect the results of soccer matches?" [29]

Another research question the authors attempted to answer is: "which ML methods can predict the results of soccer matches with greater accuracy, taking the weather conditions into account as a principal feature?" [29]. The study concluded that including weather-related variables significantly increased the accuracy of the predictions in the implemented models.

2.3. Machine learning models for predicting game outcomes in other sports

ML has also been used to build models for predicting outcomes for other sports. Beal et al. [38] built and evaluated nine algorithms using data sets of 1280 games over five seasons to predict outcomes of National Football League (NFL) games. Their experiments showed that the Naive Bayes algorithm was the best classifier with an accuracy of 67.5% and an F1 Score of 0.67. The team believed that this algorithm performs the best when the features used are independent and the dependencies between the features are similar. Gifford and Bayrak [39] used data from the 2002 to 2017 seasons and constructed predictive models to predict the outcomes of NFL games in a season using the decision tree and logistic regression algorithms. On the other hand, Roumani's research [40] focused on the application of sports analytics in the National Football League and compared the performance of the C4.5, Neural Network, and Random Forest algorithms in predicting the Superbowl winner based on data from the regular season.

Hubacek et al. in 2019 [41] incorporated the bookmaker's predictions into their prediction models by reducing correlation as an optimization tool. The models were developed by implementing a CNN with a convolution layer that "enabled to leverage a vast number of player-related statistics as its input". Combining elements of modern portfolio theory, the authors yielded positive cumulative profits in experiments with data from the National Basketball Association (NBA). More recently, Liu [42] in 2020, Chen et al. [43] in 2021, Alsonso et a. [44] in 2022, and Liu et al. [45] used the ML technique to predict the outcome of a basketball game in the NBA.

Interestingly, professional baseball teams have used a quantitative approach to manage games for many years [46]. However, few studies have used ML to predict game outcomes in Major League Baseball [47, 48].

2.4. Model evaluation

Chicco et al. [49] show the advantage of using the Matthews correlation coefficient (MCC), first introduced by Matthews [50], over the F1 score and accuracy in the evaluation of binary classification. There are many references in this regard; see, for example, [51].

3. Methodology

Given the problem's symmetry, we will focus on predicting a home team's win or loss. Therefore, our problem could be viewed as classifying the outcome of a soccer match into one of these two states: Win (W) or Not Win (NW) of the home team by an ML model developed using the data from previous seasons.

3.1. Data collection and processing

Data on soccer, especially for those in the EPL, are widely available. For this research, we have chosen the data set made available to the public for free on the website [52]. Data on matches in the 2019–2022 seasons were downloaded and used in the modelling process.

Unlike Iskandaryan et al. [29] reviewed above, the research team speculated that the use of weather-related variables associated with the field of the game and without considering those associated with the home field of the opposition team could have an impact on the outcome of the matches. As weather forecasts are readily available, they could be used to predict match outcomes if their inclusion improves the accuracy of the prediction. Consequently, weather data associated with soccer matches were also downloaded from [53] (variables 24–29 listed below).

While the main objective of the research is the game's team endgame data to predict game outcomes, the research team has considered using player-specific metrics such as performance rating, injuries, substitution, etc., and team-based dynamics such as recent strategy changes in the modelling process. It was recognized that the collection or quantification of the required data will be demanding for the project's current phase, and the considerations will be revisited after the success of the prediction analytics process is established.

Note that the data set utilized for this study comprised actual statistical data from English Premier League (EPL) matches and the corresponding weather conditions, with no missing values present in the raw data sets. However, during the feature engineering process, additional features were derived by summarizing historical data, specifically from the last match and the preceding three matches. Consequently, the summarization features contained missing values for matches within the first three weeks of both the training and testing datasets. Records with missing values were excluded from the analysis to address this problem and ensure data integrity.

3.2. Exploratory data analysis

Before splitting the data set into training and testing data sets, we performed exploratory data analysis and used the results for feature engineering.

Fig. 1 Shows the distribution of home team wins during the 2019–2021 seasons. It confirms that the data set is not severely imbalanced and that we do not need to pay particular attention to this problem.

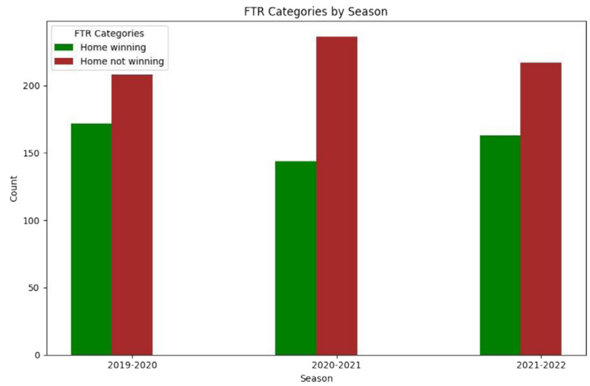

Fig. 1. Home Wins — 2019 to 2021.

3.3. Feature engineering

The extracted data set includes the following variables:

1. Match Date (dd/mm/yy) — identifier; not used in modelling
2. Home Team — identifier; not used in modelling
3. Opposition Team — identifier; not used in modelling
4. Time of match kick-off (afternoon, evening)
5. Game result (W, NW)
6. Home Team Goals (not used in the model)
7. Opposite Team Goals (not used in the model)
8. Home Team Previous Game (W, NW)
9. Opposition Team Previous Game (W, NW)
10. Home Team Previous Game Goals
11. Opposition Team Previous Game Goals
12. Home Team Previous Game Shots
13. Opposition Team Previous Game Shots
14. Home Team Previous Game Shots on Target
15. Opposition Team Previous Game Shots on Target
16. Home Team Previous Game Corners
17. Opposition Team Previous Game Corners
18. Home Team Previous Game Fouls Committed
19. Opposition Team Previous Game Fouls Committed
20. Home Team Previous Game Yellow Cards
21. Opposition Team Previous Game Yellow Cards
22. Home Team Previous Game Red Cards
23. Opposition Team Previous Game Red Cards
24. Temperature (degrees Celsius)
25. Relative Humidity (%)
26. Precipitation (mm)
27. Wind Direction (degrees)
28. Wind Speed (m/s)
29. Weather Conditions (e.g., Clouds, Rains, Clear, etc.)

Variable (5) is the target variable of the models.

3.3.1. Features related to weather, fatigue, and momentum

To include the potential impact of the game environment and the weather, we used two one-hot variables to represent the kickoff time and eight one-hot variables to represent the eight possible weather conditions. Together, we used 15 features of the environment and the weather area to build predictive models, including the temperature, relative humidity, precipitation, wind direction, and wind speed.

Intuitively, fatigue should be considered in any predictive model of game outcomes. A more rested team should have a higher chance of winning. However, there have been few studies on the measurement and impact of fatigue on the performance of professional sports teams. Draper et al. in 2024 [54] explored the association between travel and match outcomes in an elite North American Professional Soccer League through a retrospective observational study. Their results are not conclusive, but indicate that fatigue affects game outcomes. The question remains as to which proxy feature(s) should be used to represent the fatigue factor. For this study, we used the following features as the proxies for the fatigue factor:

- Time between the previous and current game — home team
- Time between the previous and current game — opposition team
- Average number of days of elapsed time between the last three games — home team
- Average number of days of elapsed time between the last three games — opposition team

Similarly, many fans and sports analysts believe in the impact of momentum on a team's winning, for example, Steeger et al. in hockey [55], Weimer et al. [56] in basketball and Zhong et al. [57] in tennis. Again, the challenge is to define and select the feature(s) to represent the momentum at the team level from game to game. Our first attempt in this regard was as follows.

- Average number of goals in the last three previous home team games
- Average number of goals in the last three previous opposition team games
- Average number of shots in the last three previous home team games
- Average number of shots in the last three previous opposition team games
- Average number of shots on target in the last three previous home team games
- Average number of shots on target in the last three previous opposition team games
- Average number of corners in the last three previous home team games
- Average number of corners in the last three previous opposition team games
- Average number of fouls in the last three previous home team games
- Average number of fouls in the last three previous opposition team games
- Average number of yellow cards in the last three previous home team games
- Average number of yellow cards in the last three previous opposition team games
- Average number of red cards in the last three previous home team games
- Average number of red cards in the last three previous opposition team games
- the number of wins in the last three previous home team games
- the number of wins in the last three previous opposition team games

The numerical features were standardized before the models were built. Together, 52 features were used in the modelling process.

To fully evaluate the potential of using historical data and a machine learning algorithm to predict match outcomes, data on bookmaker odds should not be included as features for the models [41]. Instead, we will contextualize our prediction results with the bookmakers' predictions through the odds.

3.4. Training and testing data

The data set's first two seasons (2019–2021) will be used to train the ML models using the algorithms discussed below. Data from the third season (2022) will be used as the testing data set.

3.5. Machine learning algorithms

In the research, four basic ML algorithms were considered: logistic regression (LR), random forest (RF), support vector machine (SVM), and XGboost (XGB). They were selected for their differences in structure and approach in the modelling process. An excellent overview of the use of some of these basic algorithms in sports betting can be found in [58].

The recently developed Light Gradient Boosting Machine (GBM) algorithm [59] has provided good predictive results (see, for example, Hu et al. [60] and Chang et al. [61]). We used the Bayesian Optimization option of the LightGBM algorithm in the modelling process.

As documented above, a Convolutional Neural Network (CNN) has been implemented to predict soccer and other sports events, for example, Randrianasolo [25], Chen et al. [24], or Wagenaar et al. [23]. This algorithm was also used to build a prediction model to gauge whether a more advanced neural network structure would improve predictive accuracy.

Together, these six algorithms represent the spectrum of basic statistical and machine learning techniques for binary classification. More importantly, we believe that employing an ensemble technique to combine these classifiers in some way will probably give better classification accuracy, for example, under the Condorcet paradigm [62].

Stacking and Voting are the two ensemble techniques used to combine selected classifiers once they are developed to improve performance. For example, the Stacking technique was used by Jaiyeoba et al. [63] to increase accuracy in classifying skin diseases, in detecting diabetes [64], and in education [65]. Similarly, the Voting technique was recently used in Atitallah et al. [66] to detect IoT malware. Other recent applications of this technique can be found in [65], kuman et al. [67], and Farooqi et al. [68]. Both techniques were implemented in the Scikit Learn package in Python ([69,70]) and are therefore readily available. We explored the use of these two techniques and examined the extent to which the two ensemble techniques can improve the accuracy of predicting the outcome of a soccer game. Fig. 2 shows the model development process:

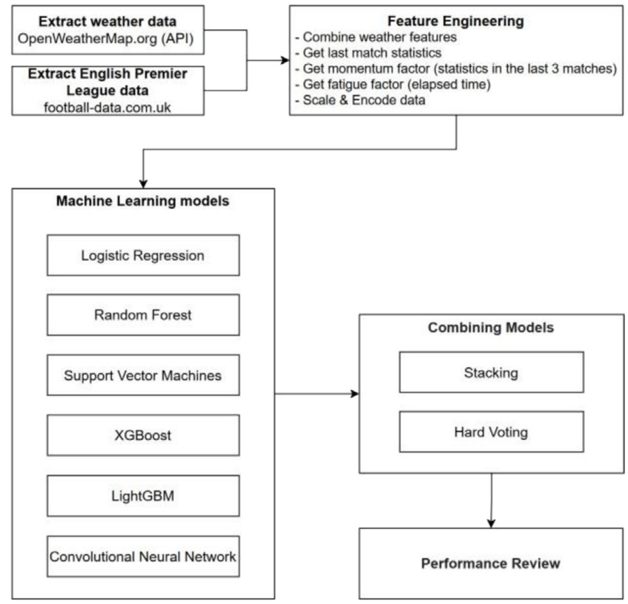

Fig. 2. Model Development Process.

3.6. Performance evaluation

The popular measurement of performance in classification is based on a confusion matrix. The matrix compares the number of predictions for each class that are correct and those that are incorrect. There are four numbers to pay attention to:

- True Positive (TP) — The number of positive observations (home wins) the model correctly predicted as positive (home wins).
- False Positive (FP) — The number of negative observations (non-home win) the model incorrectly predicted as positive (home wins).
- True Negative (TN) — The number of negative observations (non-home win) the model correctly predicted as negative (non-home win).

False Negative (FN) — The number of positive observations (home wins) the model incorrectly predicted as negative (non-home win).

Using these four numbers, we define the following performance measures for the classifier.

3.6.1. Accuracy

The overall accuracy of a model is simply the number of correct predictions divided by the total number of predictions. An accuracy score will give a value between 0 and 1, a value of 1 would indicate a perfect model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3.6.2. Precision

Precision measures how well the model is at correctly identifying the positive class. In other words, out of all predictions for the positive class, how many were correct? Using this metric alone to optimize a model would minimize false positives.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3.6.3. Recall

The percentage of recall measures how well a model predicts all the positive observations in the data set. However, it does not include information on false positives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The two percentages, precision and recall, are typically evaluated through a precision–recall curve.

3.6.4. F1 score

From the formula below, we can see that the F1 score is defined as the harmonic mean of precision and recall:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

In its construction, the F1 score takes a value between zero and one and could be expressed as a percentage. If precision and recall are 1, the F1 score is also 1. On the other hand, the F1 score is 0 when precision or recall is 0.

3.6.5. Normalized Matthews correlation coefficient

In addition to the traditional measures defined above, our research also uses the normalized Matthews correlation coefficient (UN MCC) as a measurement tool [50], giving its advantages as described in [49]. It is defined as follows.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{A}} \tag{5}$$

$$UNMCC = \frac{MCC + 1}{2} \tag{6}$$

Here A is calculated as:

$$A = (TP + FP)(TP + FN)(TN + FP)(TN + FN) \tag{7}$$

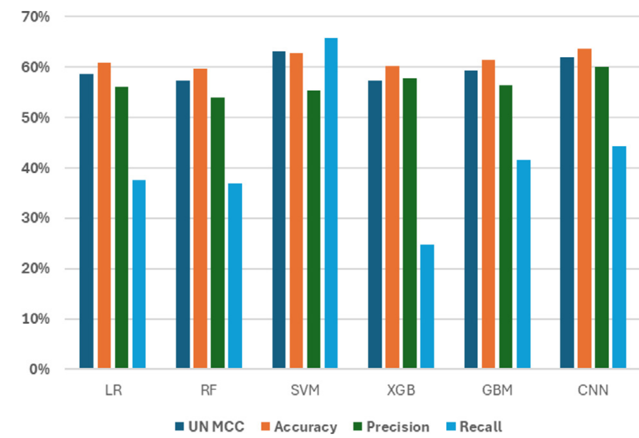

Fig. 3. Performance of the Basic Machine Learning Models Without the Weather Features.

Putting the classifier into perspective, we should also examine the performance of bookmakers. This examination is presented in the Discussion section.

4. Results

This section presents the results of the models built with the machine learning algorithms discussed in Section 3. The Logistic Regression, Random Forest, Support Vector Machine and XGBoost models were optimized through a grid-search cross-validation process. The LightGBM model was optimized with the Bayesian algorithm. Regularization methods were used where appropriate to reduce the risk of overfitting. Table 1 gives details of the models developed.

Models were first developed without them to gauge the impact of weather variables. Table 2 summarizes the results obtained by applying the models to the testing data set.

The CNN model had the highest accuracy rate, but a low recall rate. On the other hand, the SVM model had the second highest accuracy rate, the highest UN MCC score, and the best recall and F1 rate. It could be considered the best overall model (Fig. 3).

The stacking and voting (majority voting) algorithms with the six ML algorithms (stacking (all) and hard voting (all)) were next implemented. For comparison, the two ensemble algorithms were also implemented with only the LR, SVM and LightGBM (GBM) algorithms (Stacking and Voting). The results in Table 3 and Fig. 4 confirmed that the ensemble models that combined only the three algorithms were slightly better in all but one performance measure (recall). The Voting model had a slight edge over the Stacking model except for the recall rate. The Stacking model that combined all the ML models had a much better recall rate.

Another set of models was built with the same hyperparameter optimization process and the inclusion of the weather features to gauge their impact on the accuracy of the models (Table 5 and Fig. 5). The set of optimized hyperparameters used in the building of the models is presented in Table 4. The differences between these and those used to build models without weather data are highlighted in bold.

Based on Table 6, Fig. 6 and Fig. 7, except for the CNN model and the recall rate, the inclusion of weather features improved on all basic ML models. However, the results for the ensemble models were mixed. The ensemble models performed better in accuracy and precision with the weather features but the ensemble models that combined the three ML algorithms performed worse in UN MCC. Including weather features reduced the recall rate for all models.

These results demonstrated that it is worthwhile to consider weather data when building an ML model to predict the outcomes of sports games. In addition, the use of ensemble techniques to combine several predictive algorithms improves prediction accuracy [22].

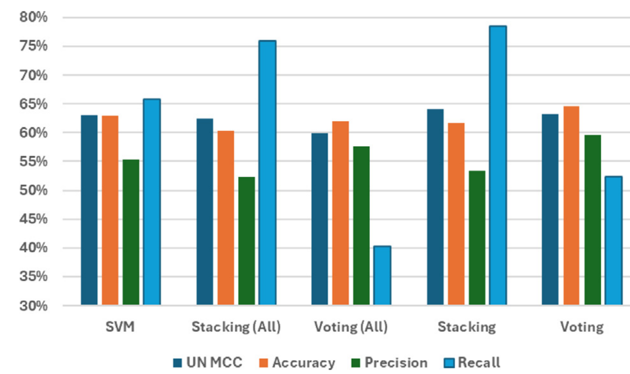

Fig. 4. Performance of the Ensemble Models Without Weather Features.

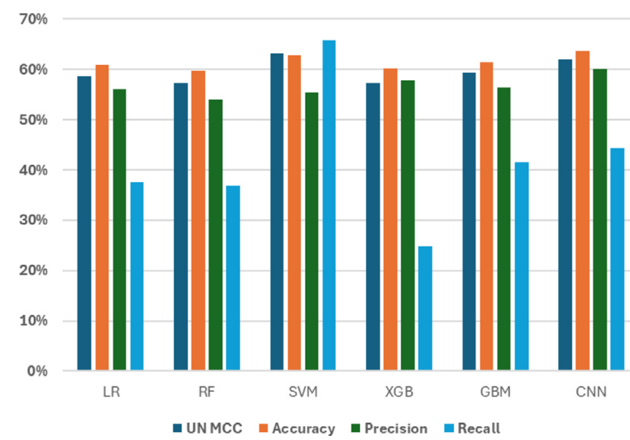

Fig. 5. Performance of the Basic Machine Learning Models with Weather Features.

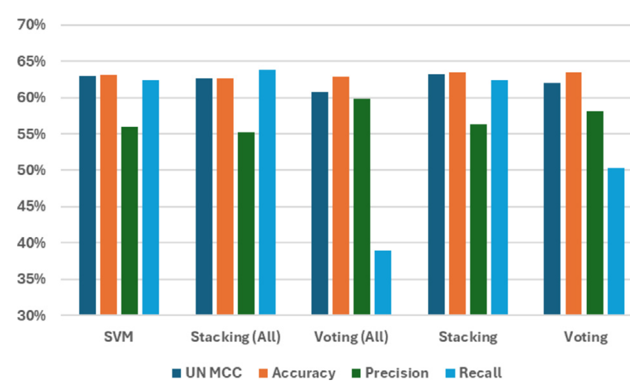

Fig. 6. Performance of the Ensemble Models With Weather Features.

5. Discussions and future work

As explained in [41], a highly accurate model is not as helpful if it coincides with those used by the bookmakers. However, an excellent predictive model should achieve a level of accuracy similar to or higher than that of the bookmakers. Therefore, it is important to compare any model developed with those articulated by the odds posted by the bookmakers. We used the matches in the testing data set and chose the odds produced by Bet365, William Hill (WH) and Pinnacle (PS) as a starting point (see Table 7).

To perform the analysis, we converted the odds produced by each bookmaker into a “prediction” for each match in the testing dataset: the bookmaker gives a prediction of a home win if the odd for the home win is lowest among the three odds (home team win, opposition

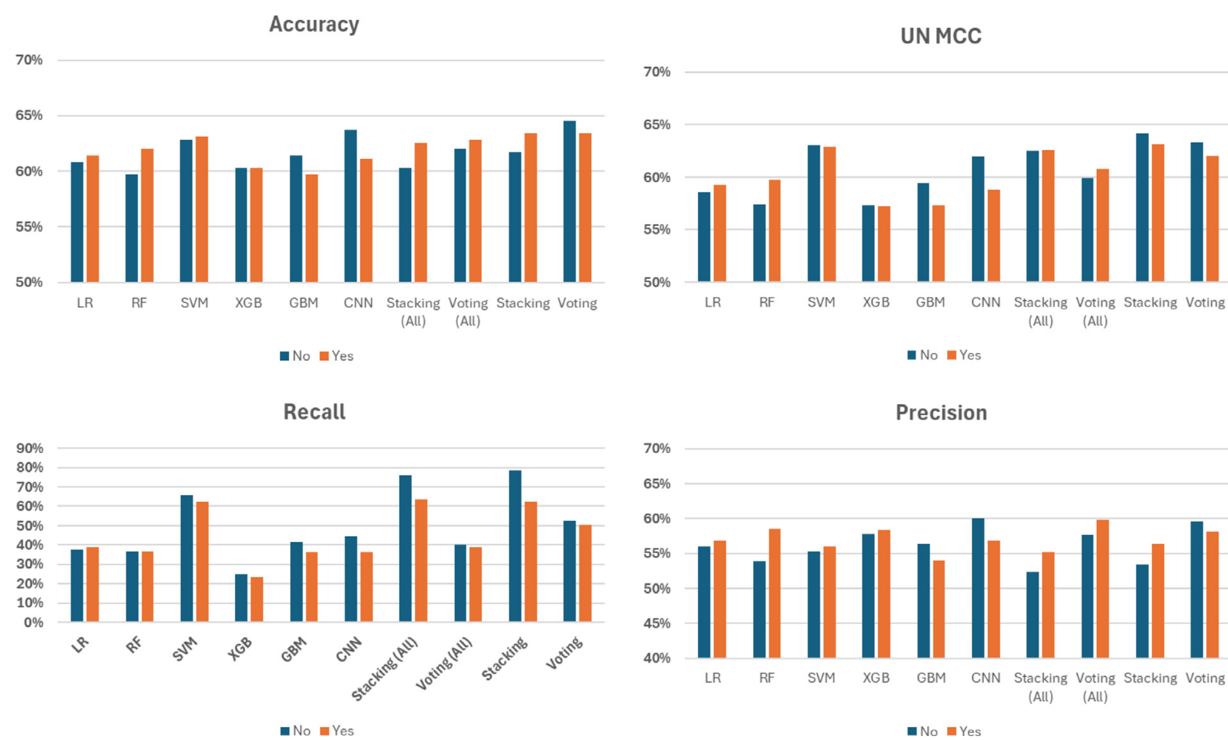

Fig. 7. Without (No) and With (Yes) Weather Features.

Table 1
Optimized hyper-parameters for models without weather data.

Model	Optimized Hyperparameters
Logistic Regression	Regularization Length=0.1, Penalty = “L2”, Solver = “lbfgs”, Tolerance=0.001
Random Forest	Max Depth=6, Min Samples Split=2, Min Samples Leaf=2, # Estimators=50 Max Features = 0.8, Max Samples = 0.8
Support Vector Machine	C=0.1, Class Weight = “balanced”, Kernel = “rbf”, Gamma = “scale”, Tolerance=0.0001
XGBoost	Max Depth=3, # Estimators=150, Learning Rate=0.01, subsample=0.8, Colsample bytree=1.0, Lambda=10
LightGBM	Max Depth=7, Num Leaves=41, Learning Rate=0.086, Min Child Samples=25, Subsample=0.87, colsample bytree=0.55, Lambda L1=4.32, Lambda L2=9.43
Convolutional Neural Network	Optimizer = “Adam”, Early Stopping, Filters=64, Kernel Size=3, Pool Size=2, Dropout=0.25, Dense Units=[64, 32] Activation=[relu, sigmoid], Regularization=l2 (0.001), Batch Size=32, Epochs=100

Table 2
Results of the machine learning models without the weather features.

Cell	LR	RF	SVM	XGB	GBM	CNN
TP	56	55	98	37	62	66
TN	157	154	122	174	153	157
FP	44	47	79	27	48	44
FN	93	94	51	112	87	83
UN MCC	58.6%	57.4%	63.1%	57.3%	59.4%	61.9%
Accuracy	60.9%	59.7%	62.9%	60.3%	61.4%	63.7%
Precision	56.0%	53.9%	55.4%	57.8%	56.4%	60.0%
Recall	37.6%	36.9%	65.8%	24.8%	41.6%	44.3%
F1	45.0%	43.8%	60.1%	34.7%	47.9%	51.0%

Table 3
Results of the ensemble models without weather features.

Cell	SVM	Stacking (All)	Voting (All)	Stacking	Voting
TP	98	113	60	117	78
TN	122	98	157	99	148
FP	79	103	44	102	53
FN	51	36	89	32	71
UN MCC	63.1%	62.5%	59.9%	64.2%	63.3%
Accuracy	62.9%	60.3%	62.0%	61.7%	64.6%
Precision	55.4%	52.3%	57.7%	53.4%	59.5%
Recall	65.8%	75.8%	40.3%	78.5%	52.3%
F1	60.1%	61.9%	47.4%	63.6%	55.7%

team win, and tie). With this definition, the interesting result from the table (Table 7 and Fig. 8) was that the bookmakers have a high level of agreement. More interesting is the fact that the accuracy rate of the bookmakers is only slightly higher than the 63.4% achieved by the

Table 4
Optimized hyper-parameters for models with weather data.

Model	Optimized Hyperparameters
Logistic Regression	Regularization Length=**1**, Penalty = "L2", Solver= "**liblinear**", Tolerance=**0.0001**
Random Forest	Max Depth=**5**, Min Samples Split=2, Min Samples Leaf=**5**, # Estimators=50 Max Features =0.8, Max Samples=**1.0**
Support Vector Machine	C=**1**, Class Weight = "balanced", Kernel = "rbf", Gamma = "**0.01**", Tolerance=**0.001**
XGBoost	Max Depth=**5**, # Estimators=150, Learning Rate=0.01, subsample=0.8, Colsample bytree=1.0, Lambda=10
LightGBM	Max Depth=**4**, Num Leaves=**42**, Learning Rate=**0.1387**, Min Child Samples=**57**, Subsample=**0.623**, colsample bytree=**0.634**, Lambda L1=**8.07**, Lambda L2=**2.89**
Convolutional Neural Network	Optimizer = "Adam", Early Stopping, Filters=64, Kernel Size=3, Pool Size=2, Dropout=0.25, Dense Units=[64, 32] Activation=[relu, sigmoid], Regularization=l2(0.001), Batch Size=32, Epochs=100

Table 5
Results of the machine learning models with weather features.

Cell	LR	RF	SVM	XGB	GBM	CNN
TP	58	55	93	35	54	54
TN	157	162	128	176	155	160
FP	44	39	73	25	46	41
FN	91	94	56	114	95	95
UN MCC	59.3%	59.8%	62.9%	57.2%	57.3%	58.8%
Accuracy	61.4%	62.0%	63.1%	60.3%	59.7%	61.1%
Precision	56.9%	58.5%	56.0%	58.3%	54.0%	56.8%
Recall	38.9%	36.9%	62.4%	23.5%	36.2%	36.2%
F1	46.2%	45.3%	59.0%	33.5%	43.4%	44.3%

Table 6
Results of the ensemble models with weather features.

Cell	SVM	Stacking (All)	Voting (All)	Stacking	Voting
TP	93	95	58	93	75
TN	128	124	162	129	147
FP	73	77	39	72	54
FN	56	54	91	56	74
UN MCC	62.9%	62.6%	60.8%	63.2%	62.0%
Accuracy	63.1%	62.6%	62.9%	63.4%	63.4%
Precision	56.0%	55.2%	59.8%	56.4%	58.1%
Recall	62.4%	63.8%	38.9%	62.4%	50.3%
F1	59.0%	59.2%	47.2%	59.2%	54.0%

Stacking and Voting algorithms. On the other hand, bookmakers are far superior in the recall rate and hence in F1.

Note that Bet 365, William Hill (WH), and Pinnacle (PS) are three of the largest bookmakers in England.

Given that the bookmakers were consistent in predicting the outcomes of the games in terms of home wins, we chose Best 365 as the representative bookmaker and compared it with the Voting classifier.

Interestingly, the Voting classifier was better at predicting home losses but less effective at predicting home wins. This was likely due to the slight imbalance of the training data set. In any case, we could conclude that there was considerable disagreement between the ML classifier and the predictions made by the bookmakers (Table 8). From a betting perspective, this should lead to profitable opportunities. Work for the future should, therefore, be as follows:

1. Conduct further feature engineering experiments on momentum and fatigue. Identify and evaluate other features that could be bought into the modelling process.

Table 7
Prediction by major bookmakers versus stacking/voting.

Cell	Stacking	Voting	Bet 365	WH	PS
TP	93	75	124	125	124
TN	129	147	109	109	109
FP	72	54	92	92	92
FN	56	74	25	24	25
UN MCC	63.2%	62.0%	69.0%	69.4%	69.0%
Accuracy	63.4%	63.4%	66.6%	66.9%	66.6%
Precision	56.4%	58.1%	57.4%	57.6%	57.4%
Recall	62.4%	50.3%	83.2%	83.9%	83.2%
F1	59.2%	54.0%	67.9%	68.3%	67.9%

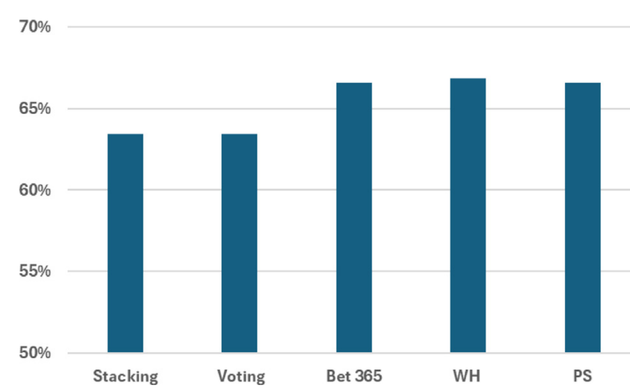

Fig. 8. Accuracy — Major Bookmakers Versus Stacking/Voting.

2. Make improvements to the classifiers focusing on re-structuring the CNN algorithm and further experimentations with stalking and voting.
3. Formulate a betting strategy taking into account the concurrence and discrepancies between the bookmaker (s) and the prediction.

In addition to these specific work related to the prediction of the outcome of a soccer game, research work could also be conducted on:

1. conditions under which the combined classifier is better than the individual classifiers

2. the best way to ensemble a set of classifiers to create a combined one with higher accuracy? What would be the extent of the improvement if this is possible?

Table 8
Actual VS bet 365 VS voting predictions.

Games	Count
Actual Home Loss	**201**
Bet 365 Predicted Home Loss	109
Hard Voting Predicted Home Loss	94
Hard Voting Predicted Home Win	15
Bet 365 Predicted Home Win	92
Hard Voting Predicted Home Lost	53
Hard Voting Predicted Home Win	39
Actual Home Win	**149**
Bet 365 Predicted Home Loss	25
Hard Voting Predicted Home Loss	20
Hard Voting Predicted Home Win	5
Bet 365 Predicted Home Win	124
Hard Voting Predicted Home Loss	54
Hard Voting Predicted Home Win	70
Total Number of Games	350

6. Conclusions

In this research, we develop several ML models to predict the outcome of soccer matches in the English Premier League. We introduce novel features such as momentum and fatigue as part of the model development process. Two ensemble approaches to combining classifiers were also used to arrive at the final predictive model. The prediction results are comparable to those of the bookmakers in terms of overall accuracy. It is recommended that further work be done to improve the models and develop a betting strategy based on them.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] O. Hubacek, G. Sır, Beating the market with a bad predictive model, Int. J. Forecast. 39 (2023) 691–719, http://dx.doi.org/10.1016/j.ijforecast.2022.02.001.
[2] J.H. Hewitt, O. Karakus, A machine learning approach for player and position adjusted expected goals in football (soccer), Franklin Open (2023) 100034, http://dx.doi.org/10.1016/j.fraope.2023.100034.
[3] C. Walsh, A. Joshi, Machine learning for sports betting: should forecasting models be optimised for accuracy or calibration? 2023, http://dx.doi.org/10.48550/arXiv.2303.06021, arXiv preprint arXiv:2303.06021.
[4] J.W. Kim, M. Magnusen, S. Jeong, March madness prediction: Different machine learning approaches with non-box score statistics, Manag. Decis. Econ. (2023) http://dx.doi.org/10.1002/mde.3814.
[5] V. Chang, K. Hall, L.M.T. Doan, Football results prediction and machine learning techniques, Int. J. Bus. Syst. Res. 17 (2023) 565–586, http://dx.doi.org/10.1504/IJBSR.2023.133178.
[6] R. Luo, V. Krishnamurthy, Who you play affects how you play: Predicting sports performance using graph attention networks with temporal convolution, 2023, http://dx.doi.org/10.48550/arXiv.2303.16741, arXiv preprint arXiv:2303.16741.
[7] V. Jain, P. Bhati, R. Singh, R. Maurya, R. Joshi, An exploratory study of ML techniques in football match's result prediction, J. Adv. Res. Embedded Syst. 9 (2023) 1–4, http://dx.doi.org/10.17148/IJARCCE.2022.11489.
[8] S. Muneer, M.B. Alvi, M. Al Sakhnani, H. Raza, T.M. Ghazal, M. Ahmad, Systematic review: Predictive models for the winning team of super leagues (SL), in: 2023 International Conference on Business Analytics for Technology and Security, ICBATS, IEEE, 2023, pp. 1–5, http://dx.doi.org/10.1109/ICBATS57792.2023.10111268.
[9] K. Mun, B. Cha, J. Lee, J. Kim, H. Jo, CompeteNet: Siamese networks for predicting win-loss outcomes in baseball games, in: 2023 IEEE International Conference on Big Data and Smart Computing, BigComp, IEEE, 2023, pp. 1–8, http://dx.doi.org/10.1109/BigComp57234.2023.00010.
[10] P.K. Czegledy, Crystal ball gazing: The future of sports betting, Gaming Law Rev. 27 (2023) 65–70, http://dx.doi.org/10.1089/glr2.2022.0046.
[11] S. Tokić, A. Panjkota, M. Matetić, Machine learning for data analysis in football: a survey of methods and problems, Ann. DAAAM Proc. 33 (2022) http://dx.doi.org/10.2507/33rd.daaam.proceedings.070.
[12] C.P. Igiri, E.O. Nwachukwu, An improved prediction system for football a match result, IOSR J. Eng. 4 (2014) 12–20, http://dx.doi.org/10.9790/3021-04124012020.
[13] A.C. Constantinou, Dolores: a model that predicts football match outcomes from all over the world, Mach. Learn. 108 (2019) 49–75.
[14] L. Carloni, A. De Angelis, G. Sansonetti, A. Micarelli, A machine learning approach to football match result prediction, in: HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July (2021) 24–29, Proceedings, Part II 23, Springer, 2021, pp. 473–480, http://dx.doi.org/10.1007/978-3-030-78642-7_63.
[15] E. Wheatcroft, E. Sienkiewicz, Calibration and hyperparameter tuning in football forecasting with machine learning, in: Mathsport International 8 2021 Proceedings, 2021, pp. 66–72, https://www.researchgate.net/publication/352846677.
[16] S.K. Nivetha, M. Geetha, R.C. Suganthe, R.M. Prabakaran, S. Madhuvanan, A.M. Sameer, A deep learning framework for football match prediction, in: 2022 International Conference on Computer Communication and Informatics, ICCCI, IEEE, 2022, pp. 1–7, http://dx.doi.org/10.1109/ICCCI54379.2022.9740760.
[17] M.A. Rahman, A deep learning framework for football match prediction, SN Appl. Sci. 2 (2020) 165, http://dx.doi.org/10.1007/s42452-019-1821-5.
[18] G. Peters, D. Pacheco, Betting the system: Using lineups to predict football scores, 2022, http://dx.doi.org/10.48550/arXiv.2210.06327, arXiv preprint arXiv:2210.06327.
[19] W. Yao, Y. Wang, M. Zhu, Y. Cao, D. Zeng, Goal or miss? A Bernoulli distribution for in-game outcome prediction in Soccer, Entropy 24 (2022) http://dx.doi.org/10.3390/e24070971.
[20] F. Rodrigues, A. Pinto, Prediction of football match results with machine learning, Procedia Comput. Sci. 204 (2022) 463–470.
[21] R. Mattera, Forecasting binary outcomes in Soccer, Ann. Oper. Res. 325 (2023) 115–134, http://dx.doi.org/10.1007/s10479-021-04224-8.
[22] E.F.E. Atta Mills, Z. Deng, Z. Zhong, J. Li, Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques, J. Big Data 11 (2024) 170, http://dx.doi.org/10.1186/s40537-024-01008-2, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-01008-210.1186/s40537-024-01008-2.
[23] M. Wagenaar, E. Okafor, W. Frencken, M. Wiering, Using deep convolutional neural networks to predict goal-scoring opportunities in Soccer, in: 6th International Conference on Pattern Recognition Applications and Methods, ICPRAM 2017, 2017, pp. 448–455.
[24] H. Chen, Neural network algorithm in predicting football match outcome based on player ability index, Adv. Phys. Educ. 9 (2019) 215–222.
[25] A.S. Randrianasolo, Using convolutional neural network to predict EURO games, in: 2024 IEEE 14th Annual Computing and Communication Workshop and Conference, CCWC, IEEE, 2024, pp. 285–290, http://dx.doi.org/10.1109/CCWC60891.2024.10427667.
[26] A. Groll, C. Ley, G. Schauberger, H. Van Eetvelde, A. Zeileis, Hybrid machine learning forecasts for the FIFA women's world cup 2019, 2019, http://dx.doi.org/10.48550/arXiv.1906.01131, arXiv preprint arXiv:1906.01131.
[27] J.D. Rose, M.K. Vijaykumar, U. Sakthi, P. Nithya, Comparison of football results using machine learning algorithms, in: 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES, IEEE, 2022, pp. 1–7, http://dx.doi.org/10.1109/ICSES55317.2022.9914265.
[28] A. Zimmermann, S. Moorthy, Z. Shi, Predicting college basketball match outcomes using machine learning techniques: some results and lessons learned, 2013, http://dx.doi.org/10.48550/arXiv.1310.3607, arXiv preprint arXiv:1310.3607.
[29] D. Iskandaryan, F. Ramos, D.A. Palinggi, S. Trilles, The effect of weather in Soccer results: an approach using machine learning techniques, Appl. Sci. 10 (2020) 6750, http://dx.doi.org/10.3390/app10196750.
[30] S. Kampakis, A. Adamides, Using Twitter to predict football outcomes, 2014, http://dx.doi.org/10.48550/arXiv.1411.1243, arXiv preprint arXiv:1411.1243.
[31] M.P. Schwellnus, E.W. Derman, Jet lag and environmental conditions that may influence exercise performance during the 2010 FIFA world cup in South Africa: CPD, South Afr. Family Pract. 52 (2010) 198–205, http://dx.doi.org/10.1080/20786204.2010.10873971.
[32] M. Mohr, L. Nybo, J. Grantham, S. Racinais, Physiological responses and physical performance during football in the heat, PLoS One 7 (2012) e39202, http://dx.doi.org/10.1371/journal.pone.0039202.
[33] F. Owramipur, P. Eskandarian, F.S. Mozneb, Football result prediction with Bayesian network in Spanish league-Barcelona team, Int. J. Comput. Theory Eng. (2013) 812–815, http://dx.doi.org/10.7763/ijcte.2013.v5.802.
[34] J.W. Orchard, M. Waldén, M. Hägglund, J.J. Orchard, I. Chivers, H. Seward, J. Ekstrand, Comparison of injury incidences between football teams playing in different climatic regions, Open Access J. Sports Med. (2013) 251–260, http://dx.doi.org/10.2147/OAJSM.S52417.

[35] G.P. Nassis, J. Brito, J. Dvorak, H. Chalabi, S. Racinais, The association of environmental heat stress with performance: analysis of the 2014 FIFA world cup Brazil, Br. J. Sports Med. 49 (2015) 609–613, http://dx.doi.org/10.1136/bjsports-2014-094449.

[36] S. Landset, M.F. Bergeron, T.M. Khoshgoftaar, Using weather and playing surface to predict the occurrence of injury in major league soccer games: A case study, in: 2017 IEEE International Conference on Information Reuse and Integration, IRI, IEEE, 2017, pp. 366–371, http://dx.doi.org/10.1109/IRI.2017.86.

[37] R.L. Lucena, E.T. Steinke, C. Pacheco, L.L. Vieira, M.O. Betancour, V.A. Steinke, The Brazilian world cup: too hot for Soccer? Int. J. Biometeorol. 61 (2017) 2195–2203, http://dx.doi.org/10.1007/s00484-017-1425-y.

[38] R. Beal, S.E. Middleton, T.J. Norman, S.D. Ramchurn, Combining machine learning and human experts to predict match outcomes in football: A baseline model, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 15447–15451.

[39] M. Gifford, T. Bayrak, A predictive analytics model for forecasting outcomes in the national football league games using decision tree and logistic regression, Decis. Anal. J. 8 (2023) 100296.

[40] Y.F. Roumani, Sports analytics in the NFL: classifying the winner of the superbowl, Ann. Oper. Res. 325 (2023) 715–730, http://dx.doi.org/10.1007/s10479-022-05063-x.

[41] O. Hub´a˘cek, G. Sourek, F. Zelezny´, Exploiting sports-betting market using machine learning, Int. J. Forecast. 35 (2019) 783–796, http://dx.doi.org/10.1016/j.ijforecast.2019.01.001.

[42] J. Liu, Predicting United States national basketball games using machine learning techniques, 2020, Available at SSRN 3827474 URL: https://ssrn.com/abstract=3827474.

[43] W.-J. Chen, M.-J. Jhou, T.-S. Lee, C.-J. Lu, Hybrid basketball game outcome prediction model by integrating data mining methods for the national basketball association, Entropy 23 (2021) 477, http://dx.doi.org/10.3390/e23040477.

[44] R.P. Alonso, M.B. Babac, Machine learning approach to predicting a basketball game outcome, Int. J. Data Sci. 7 (2022) 60–77, http://dx.doi.org/10.1504/IJDS.2022.124356.

[45] J. Liu, D. Jayasuriya, K. Dow, Predicting United States National Basketball Game Spreads using Machine Learning Techniques, Technical Report, The University of Auckland Business School Research Paper Series, 2024, http://dx.doi.org/10.2139/ssrn.4766044.

[46] B. Baumer, A. Zimbalist, The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball, University of Pennsylvania Press, 2014, https://www.jstor.org/stable/j.ctt13x1njg.

[47] A.Y. Cui, Forecasting Outcomes of Major League Baseball Games using Machine Learning, University of Pennsylvania, Philadelphia, PA, USA, 2020, https://fisher.wharton.upenn.edu/wp-content/uploads/2020/09/Thesis_Andrew-Cui.pdf.

[48] M.-L. Huang, Y.-Z. Li, Use of machine learning and deep learning to predict the outcomes of major league baseball matches, Appl. Sci. 11 (2021) 4499, http://dx.doi.org/10.3390/app11104499.

[49] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics 21 (2020) http://dx.doi.org/10.1186/s12864-019-6413-7.

[50] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage Lysozyme, Biochim. Biophys. Acta (BBA)-Protein Struct. 405 (1975) 442–451, http://dx.doi.org/10.1016/0005-2795(75)90109-9.

[51] Q. Zhu, On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset, Pattern Recognit. Lett. 136 (2020) 71–80, http://dx.doi.org/10.1016/j.patrec.2020.03.030.

[52] football Data.co.uk, Football-data.com.uk, 2024, URL: https://www.football-data.co.uk/englandm.php.

[53] OpenWeatherMap.org, Open weather, 2024, URL: https://openweathermap.org/.

[54] G. Draper, P. Chesterton, M.D. Wright, Is travel associated with match performance in elite North American professional soccer? An exploratory study, Int. J. Sports Physiol. Perform. 19 (2024) 1021–1029.

[55] G.M. Steeger, J.L. Dulin, G.O. Gonzalez, Winning and losing streaks in the national hockey league: Are teams experiencing momentum or are games a sequence of random events? J. Quant. Anal. Sports 17 (2021) 155–170, http://dx.doi.org/10.1515/jqas-2020-0077.

[56] L. Weimer, Z.C. Steinert-Threlkeld, K. Coltin, A causal approach for detecting team-level momentum in NBA games, J. Sports Anal. 9 (2023) 117–132.

[57] M. Zhong, Z. Liu, P. Liu, M. Zhai, Searching for the effects of momentum in tennis and its applications, Procedia Comput. Sci. 242 (2024) 192–199.

[58] A. Kolla´, Betting models using AI: A review on ANN, SVM, and Markov chain, MPRA Paper (2021) 106821, URL: https://mpra.ub.uni-muenchen.de/id/eprint/106821.

[59] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, Adv. Neural Inform. Process. Syst. 30 (2017) URL: https://dl.acm.org/doi/10.5555/3294996.3295074.

[60] Y. Hu, F. Ma, M. Hu, B. Shi, D. Pan, J. Ren, Development and validation of a machine learning model to predict the risk of readmission within one year in HFpEF patients: Short title: Prediction of HFpEF readmission, Int. J. Med. Inform. (2024) 105703.

[61] V. Chang, S. Sivakulasingam, H. Wang, S.T. Wong, M.A. Ganatra, J. Luo, Credit risk prediction using machine learning and deep learning: A study on credit card customers, Risks 12 (2024) 174.

[62] R. Theisen, H. Kim, Y. Yang, L. Hodgkinson, M.W. Mahoney, When are ensembles really effective? Adv. Neural Inf. Process. Syst. 36 (2024).

[63] O. Jaiyeoba, E. Ogbuju, O.T. Yomi, F. Oladipo, Development of a model to classify skin diseases using stacking ensemble machine learning techniques, J. Comput. Theories Appl. 2 (2024) 22–38, http://dx.doi.org/10.62411/jcta.10488.

[64] M. Gollapalli, A. Alansari, H. Alkhorasani, M. Alsubaii, R. Sakloua, R. Alzahrani, M. AlHariri, M. Alfares, D. AlKhafaji, R. Al Argan, A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: pre-diabetes, T1dm, and T2DM, Comput. Biol. Med. 147 (2022) 105757, http://dx.doi.org/10.1016/j.compbiomed.2022.105757.

[65] M.H.I. Bijoy, A. Pramanik, M.S. Rahman, M. Hasan, S.A. Akhi, M.M. Rahman, MKRF stacking-voting: a data mining technique for predicting educational satisfaction level of Bangladeshis student during pandemic, in: 2022 IEEE 7th International Conference for Convergence in Technology, I2CT, IEEE, 2022, pp. 1–7, http://dx.doi.org/10.1109/I2CT54291.2022.9824357.

[66] S.B. Atitallah, M. Driss, I. Almomani, A novel detection and multi-classification approach for IoT-malware using random forest voting of fine-tuning convolutional neural networks, Sensors 22 (2022) 4302, http://dx.doi.org/10.3390/s22114302.

[67] R. Kumar, A. Chaturvedi, Software bug prediction using reward-based weighted majority voting ensemble technique, IEEE Trans. Reliab. (2023) http://dx.doi.org/10.1109/TR.2023.3295598.

[68] A.H. Farooqi, S. Akhtar, H. Rahman, T. Sadiq, W. Abbass, Enhancing network intrusion detection using an ensemble voting classifier for Internet of Things, Sensors 24 (2023) 127, http://dx.doi.org/10.3390/s24010127.

[69] Scikit-Learn.org, Stacking classifiers, 2024, URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.StackingClassifier.html.

[70] Scikit-Learn.org, Voting classifier, 2024, URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html.