Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 1, 2026, 71-81.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

User Bias and Algorithmic Accuracy in Football Outcome Prediction: Evidence from the English Premier League

Fateme Ghasemi[1], Ameneh Khadivar[*2], Leila Moslehi[2] and Fatemeh Abbasi[3]

1. Department of Management, Faculty of Social Sciences and Economics,Alzahra University,Iran.
2. Department of Sport Management, Faculty of Sport Sciences, Alzahra University, Tehran, Iran.
3. Department of Industrial Management and Information Technology, Shahid Beheshti University, Tehrani, Iran.

Article Info

Article History:
Received 26 August 2025
Revised 01 October 2025
Accepted 30 October 2025

DOI:10.22044/jadm.2025.16736.2805

Keywords:
Football Match Prediction, Sports Data Mining, Machine Learning Algorithms, prediction.

**Corresponding author: a.khadivar@alzahra.ac.ir (A. Khadivar).*

Abstract

This study investigates the effectiveness of machine learning algorithms, including Neural Networks, Bayesian Networks, Support Vector Machines, and Random Forests, in predicting football match outcomes using data from the English Premier League (2018–2022). By incorporating user-generated probabilities for home win, away win, and draw alongside conventional features, the models were evaluated under binary and multi-class classification scenarios. The Support Vector Machine achieved the highest accuracy (69%) in the win-loss scenario, while the Neural Network reached 51% in the win-draw-loss scenario. Results indicate that user-derived features enhance predictive performance, although user predictions exhibit a bias toward home teams, especially in uncertain cases. These findings highlight the potential of integrating user perspectives into predictive modeling and underscore the importance of addressing cognitive bias in sports analytics.

1. Introduction

In today's digital era, vast volumes of data are generated across domains such as the Internet of Things, business, social media, and healthcare [1]. Among these, sports stand out as a dynamic field producing extensive data, including match statistics, player performance, news articles, and expert reviews. For instance, football matches alone yield rich datasets on teams, players, and seasons [2]. When analyzed with purpose, such data can be transformed into valuable insights.

Artificial Intelligence (AI) plays a pivotal role in intelligent data analysis and the development of smart, automated applications [1,3]. The synergy between AI and big data offers numerous benefits to the sports industry [4]. Within AI, machine learning (ML) has emerged as a leading approach, enabling accurate predictions and trend analysis through techniques such as data mining and association [5,6,7]. The prediction process typically involves data collection, processing, and modeling, often expressed through mathematical frameworks [8].

Predicting sports outcomes has long intrigued fans, coaches, players, media professionals, and gamblers alike [2,9]. It also presents a compelling research challenge due to its inherent complexity [10]. Historically, decisions in sports were driven by intuition and personal experience. Today, data-driven methodologies have revolutionized predictive strategies [11]. Modern sports prediction involves drawing conclusions from historical data, knowledge, and contextual information [12]. In professional sports, where financial stakes are high, accurate forecasting is essential for strategic planning [13]. The growing interest in match prediction has also fueled the expansion of the betting industry, with the football betting market estimated to be worth between $700 billion and $1 trillion [14].

Football, played between two teams of eleven players, results in one of three outcomes: win, draw, or loss [15]. Over the past century, it has become one of the most influential and commercially successful sports worldwide. In recent decades, football has evolved into a major industry, impacting sponsorships, broadcasting rights, and player transfers [14,16]. Its economic significance is particularly notable in Western countries, where football contributes substantially to national economies.

With the proliferation of global tournaments such as the FIFA World Cup, UEFA Champions League, and AFC Champions League, predicting match outcomes has gained prominence in academic research [17,18].

However, football outcomes are influenced by a multitude of subjective and objective factors, including player skill, team dynamics, injuries, rankings, and game context—making quantitative prediction highly complex [19,20]. Traditional statistical methods often fall short in handling the volume and variability of sports data, especially when implicit rules and patterns are difficult to extract [21].

In contrast, ML algorithms offer adaptive and intelligent learning capabilities, particularly when applied to time-series data. These models can outperform traditional approaches in forecasting future outcomes [22], making them especially valuable in professional leagues with global influence.

This study focuses on the English Premier League (EPL), the most-watched sports league globally, reaching 643 million households across 212 territories and a potential audience of 4.7 billion [14]. Predicting EPL match outcomes is not only of interest to fans but also critical for managers and stakeholders due to its financial implications. While early research predominantly utilized artificial neural networks, recent literature suggests the need for broader algorithmic comparisons [2]. ML has been widely applied to football prediction tasks [9,23,24], with deep learning (DL) offering advanced capabilities through multi-layered data abstraction [1,25].

Although neural networks have shown promise, other classifiers, such as decision trees, random forests, support vector machines, and Bayesian networks, have occasionally demonstrated superior accuracy [2].

In this research, we apply and compare several ML algorithms, including multilayer perceptrons, random forests, support vector machines, and Bayesian networks, using data extracted from a sports news website. The predicted outcomes are then evaluated against user opinions to assess performance and reliability.

2. Methodology

This applied research adopts the CRISP-DM methodology to guide the data mining process. CRISP-DM (Cross-Industry Standard Process for Data Mining) is a widely recognized, domain-independent framework for structuring data mining projects [26,27].

In alignment with its first phase, business understanding, it was essential to gain a comprehensive grasp of the research problem. Accordingly, data were sourced from Varzesh3, a prominent Iranian sports news platform that has been active since 2010. Varzesh3 offers 24-hour coverage of sports-related content and is currently ranked as the third most visited website in Iran and the leading news website in the country, according to official Alexa statistics.

One of Varzesh3's most popular features is its football match statistics section, which includes detailed information such as team names, match schedules, and historical performance metrics. Additionally, the platform allows users to submit predictions for upcoming matches, making it a valuable resource for analyzing fan behavior and forecasting outcomes. Users can submit their predictions up until the official start time of the match, after which the submission window automatically closes.

In the second phase of CRISP-DM—data understanding and collection—the focus shifted to user predictions related to the English Premier League. Data were extracted from multiple databases associated with the Varzesh3 website. A custom console application was developed using C# within the ASP.NET framework, comprising two main components. The first component interfaced directly with Varzesh3's relational database to retrieve in-game event statistics. The second component employed a web crawler to navigate the prediction section of each match and collect user-submitted predictions. A total of 55 features were selected for analysis (see Table 1), encompassing both match-related attributes and user prediction data.

Table 1. Selected features.

Number	Features	N	Features	N	Features
1	Match Id	20	Guest First Half Yellow Card Count	38	Host Offside
2	Host Id	21	Guest Second Half Yellow Card Count	39	Guest Offside
3	Host Name	22	Red Card Count	40	Host Possession
4	Guest Id	23	Host Red Card Count	41	Guest Possession
5	Guest Name	24	Host First Half Red Card Count	42	Host Shots On
6	Scheduled Start On	25	Host Second Half Red Card Count	43	Guest Shots On
7	Host Goals	26	Guest Red Card Count	44	Host Saves
8	Guest Goals	27	Guest First Half Red Card Count	45	Guest Saves
9	Status	28	Guest Second Half Red Card Count	46	Host Formation
10	Time Status	29	Winner Id	47	Guest Formation
11	Stage Id	30	Host Corners	48	Host Average Age
12	Round	31	Corners Guest	49	Guest Average Age
13	Stadium	32	Yellow Half Second Host Count Card	50	Winner Of Number Host
14	Referee	33	Guest Yellow Card Count	51	Winner Of Number Guest
15	Count Card Yellow	34	Kicks Free Host	52	Host Average Score in Last Five Matches
16	Cards Yellow Host	35	Kicks Free Guest	53	Guest Average Score in Last Five Matches
17	Yellow Half First Host Cards	36	Attempts Goal Host	54	Host Average Score in Last Five Head-to-Head Matches
18	Fouls Host	37	Attempts Goal Guest	55	Guest Average Score in Last Five Head-to-Head Matches
19	Fouls Guest				

2.1. Data Preprocessing and Modeling

In data preprocessing, missing values such as incomplete player statistics or absent possession data were handled by imputing with the median value of the corresponding feature across the last five games of the team. All numerical features were normalized to a [0,1] range using min-max scaling to ensure comparability across features. Features reflecting match performance (e.g., shots on goal, fouls) were aggregated as rolling averages of each team's last five matches prior to the target match, following established practices to preserve the predictive nature of historical data.

Following the initial phases of the CRISP-DM methodology, data preprocessing was conducted to prepare the dataset for modeling. This stage began with a thorough review, indexing, and preliminary processing of the raw data, followed by data cleaning to resolve errors and improve overall data quality. A key aspect of data cleaning involved the removal of duplicate and irrelevant variables, which typically do not contribute meaningfully to model performance. Reducing the number of features not only enhances model efficiency but also improves interpretability and reduces the likelihood of overfitting.

During this phase, variables were individually assessed using various techniques to identify those with the greatest impact on predictive performance. For instance, features such as HostGoals and GuestGoals were excluded, as they directly reveal match outcomes and would compromise the integrity of the prediction model. Similarly, the Status variable—indicating whether a match was completed—was removed, since only finalized matches were considered in this study.

Additional challenges arose with variables that are determined during or after the match, which are unsuitable for retrospective prediction. To address this, variables numbered 15 to 45, primarily influenced by in-game events, were averaged based on each team's last five matches. This approach ensured that only historical data was used, preserving the predictive nature of the model. Ultimately, 36 features were retained, encompassing 1,901 records from five English Premier League seasons spanning 2018 to 2022.

2.2. Modeling and Evaluation

The fourth phase of the CRISP-DM framework involves model development and evaluation. This step is critical, as a model may perform well under one metric but poorly under others. In this study, four widely accepted evaluation metrics were employed: accuracy, precision, recall, and F1-score. These metrics were calculated using the following formulas:

$$\begin{aligned}
&\bullet Precision = TP \,/\, (TP + FP) \\
&\bullet Accuracy = (TP + TN) \,/\, (TP + FP + FN + TN) \\
&\bullet Recall = TP \,/\, (TP + FN) \\
&\bullet F1\,Score = (2 \times Precision \times Recall) \,/\, (Precision + Recall)
\end{aligned} \quad (1)$$

Hyperparameters for each algorithm were selected using a manual grid search over a set of commonly recommended configurations based on prior studies [28,29]. For instance, the multilayer perceptron was tested with 1–3 hidden layers and varying neuron counts (10–50 per layer), using ReLU activation and the Adam optimizer. The random forest was optimized by varying the number of estimators (50–200) and maximum depth (5–20). SVM models were evaluated with both linear and RBF kernels, with grid search on the regularization parameter C (0.1–10). Final choices were those yielding the highest validation accuracy in preliminary experiments.

4.3. Algorithm Implementation

For algorithm development, the Python programming language was utilized due to its versatility, scalability, and widespread adoption in data science and machine learning. The following libraries were employed:

- pandas for data manipulation and retrieval
- matplotlib and seaborn for visualizing the correlation matrix
- sklearn.neural_network for implementing neural network models
- sklearn.metrics for evaluating model performance
- sklearn.ensemble for executing the random forest algorithm
- sklearn.svm for support vector machine implementation
- sklearn.naive_bayes for applying Bayesian classification

5. Results

The modeling diagram can be seen in Figure 1.

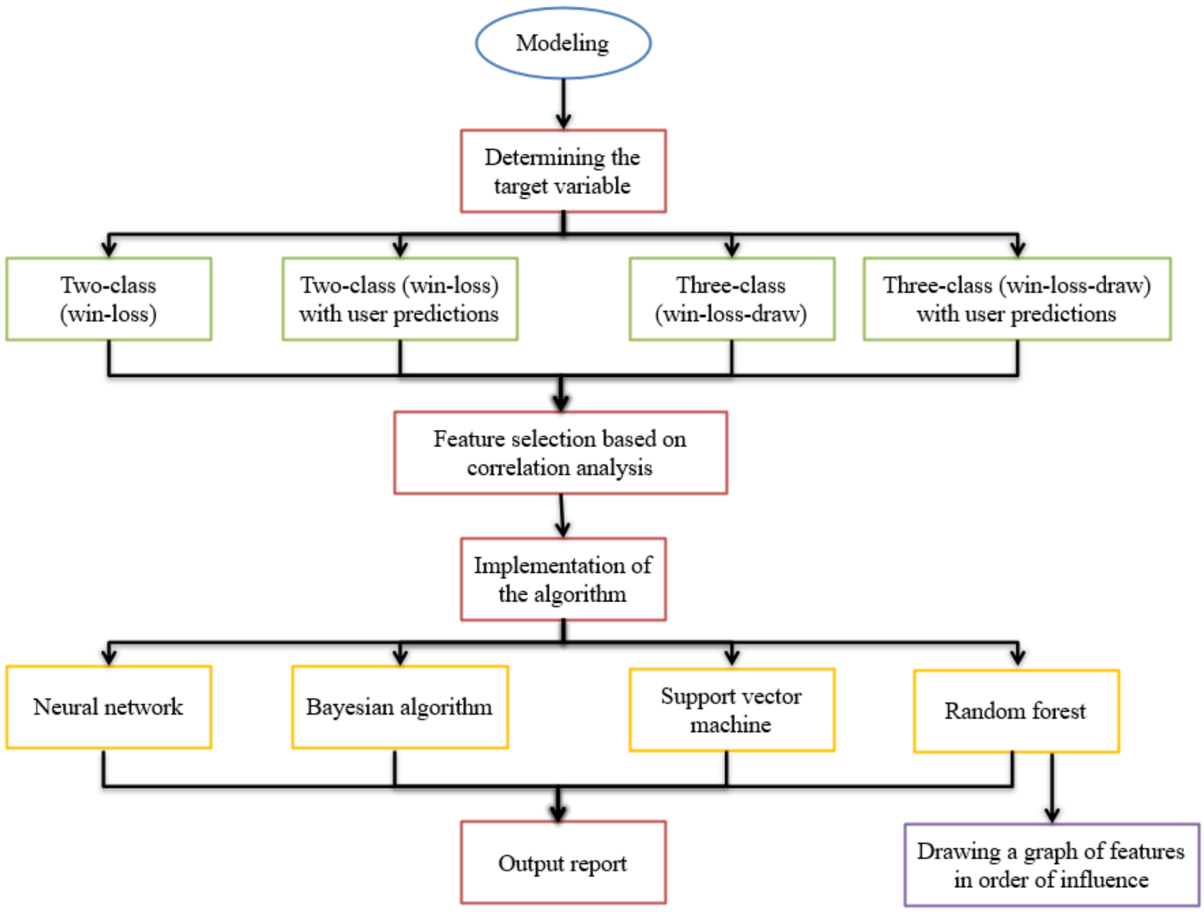

Figure 1. Modeling Diagram.

By defining the target variable in two classes, win-loss, and three classes, win-loss-draw, and considering user predictions as well as without considering them, two additional scenarios are created. Ultimately, a correlation matrix will be run for all four scenarios. If necessary, features will be removed based on the correlation percentage with the variables and the target variable. Subsequently four algorithms—neural network, Bayesian, support vector machine, and random forest—will be executed along with feature importance charts according to their effectiveness. Finally, the output of each algorithm will be reported.

- **First scenario: Two-class (win-loss)**

In this model, matches that resulted in a draw, meaning the winning team ID was zero, were removed from the data, reducing the dataset to 1,467 records. The target variable is used to label or predict the target(s) for each data sample in classification or prediction tasks. This variable can have binary, multiclass, or continuous values, depending on the problem type. A column was created as the target, categorizing the host team if the winning team ID was equal to the host team ID; otherwise, the guest team was chosen. 833 records were assigned to the host category, and 634 records to the guest category. A comparison of the accuracy and F1-score of the algorithms in this scenario is provided in Table 2.

Table 2. Comparison of Accuracy and F1-Score of Algorithms in the First Scenario.

Algorithm	Accuracy	F1-Score
MLP	0.68	0.67
NB	0.65	0.65
Svm	0.65	0.64
RF	0.66	0.66

- **Second scenario: Three-class (win-loss-draw)**

In this model, the draw scenario was considered, and the number of records remained unchanged. A column was created as the target, categorized as the host team if the winning team ID was equal to the host team ID; if the result was a draw, it was labeled as ‘none’; if the guest team won, it was categorized as ‘guest’. 833 records were classified under the host category, 634 under the guest category, and 433 under the "none" category, indicating a draw. A comparison of the accuracy and F1-score of the algorithms in this scenario is provided in Table 3.

Table 3. Comparison of Accuracy and F1-Score of Algorithms in the Second Scenario.

Algorithm	Accuracy	F1-Score
MLP	0.50	0.42
NB	0.48	0.40
Svm	0.47	0.43
RF	0.47	0.42

- **Third scenario: Two-class (win-loss) with user predictions**

Initially, the users' prediction data was merged with the match-related data, resulting in a total of 1,081 records. Three features were added: (user prediction ratio of the host team winning, user prediction ratio of the guest team winning, and user prediction ratio of a draw). In this model, matches ending in a draw, where the winning team ID was zero, were removed from the data, reducing it to 834 records. The target variable was created as the host team if the winning team ID was equal to the host team ID; otherwise, it was categorized as the guest team. 455 records were classified under the host category, and 379 under the guest category. A comparison of the accuracy and F1-score metrics of the algorithms in this scenario is provided in Table 4.

Table 4. Comparison of Accuracy and F1-Score of Algorithms in the Third Scenario.

Algorithm	Accuracy	F1-Score
MLP	0.64	0.64
NB	0.67	0.67
Svm	0.69	0.69
RF	0.67	0.67

- **Fourth scenario: Three-class (win-loss-draw) with user predictions**

Initially, the users' prediction data was merged with the match-related data, resulting in a total of 1,081 records. Three features were added: (user prediction ratio of the host team winning, user prediction ratio of the guest team winning, and user prediction ratio of a draw). In this model, the draw scenario was considered, and the number of records remained unchanged. A column was created as the target, categorized as the host team if the winning team ID was equal to the host team ID; if the result was a draw, it was labeled as ‘none’; if the guest team won, it was categorized as ‘guest’. 455 records were classified under the host category, 379 under the guest category, and 274 under the "none" category, indicating a draw. A comparison of the accuracy and F1-score of the algorithms in this scenario is provided in Table 5.

Table 5. Comparison of Accuracy and F1-Score of Algorithms in the Fourth Scenario

Algorithm	Accuracy	F1-Score
MLP	0.51	0.41
NB	0.50	0.45
Svm	0.48	0.45
RF	0.49	0.44

- **Guest Win Probability with AUC Metric Compared to SVM Algorithm**

The comparative AUC analysis highlights that while the Support Vector Machine demonstrated balanced discrimination ability for both host and guest win predictions (AUC ≈ 0.75), user predictions showed strong alignment with host wins but were significantly weaker in correctly identifying guest wins (AUC ≈ 0.24). This asymmetry underscores the presence of a home-team bias among users and validates the added value of algorithmic approaches for capturing guest victory scenarios.

The user prediction ratio density distribution of the model's predictions for the guest label category, as well as the user prediction ratio density distribution for the host win percentage, which was extracted from the input data X, can be seen in Figure (2).

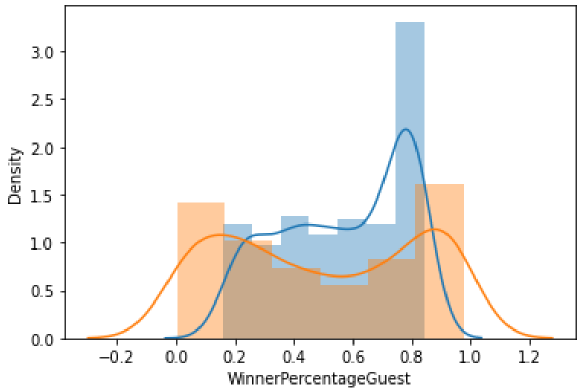

Figure 2. AUC Metric Comparison for Guest Win Probability. The orange colour represents SVM model predictions, while blue represents user prediction ratios.

According to Figure (2), in matches where the guest team won, a distribution that is more to the right is better, as it indicates that the user prediction ratio should be closer to one.

Table 6. AUC Metric Output for Guest Win Probability.

	AUC Score
Support Vector Machine	0.75
User Predictions	0.24

According to Table 6, the AUC score for the guest win user prediction ratio is 0.24, compared to the AUC score of 0.75 for the Support Vector Machine (SVM) algorithm.

- **Host Win Probability with AUC Metric Compared to SVM Algorithm**

The user prediction ratio density distribution of the model's predictions for the host label category, as well as the user prediction ratio density distribution for the host win percentage, which was extracted from the input data X, can be seen in Figure 3.

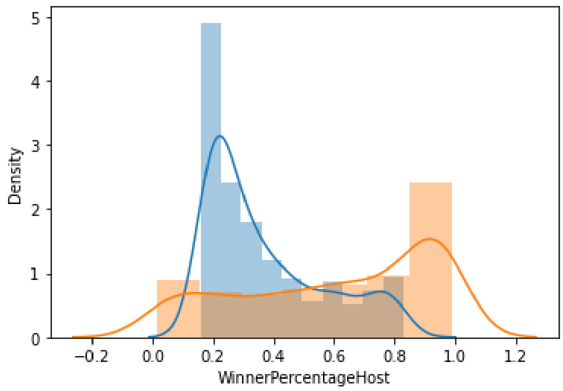

Figure 3. AUC Metric Comparison for Host Win Probability The orange colour represents SVM model predictions, and blue represents user prediction ratios.

According to Figure 3, in matches where the host team won, a distribution that is more to the left is better, as it indicates that the user prediction ratio should be closer to zero.

Table 7. AUC Metric Output for Host Win Probability.

	AUC Score
Support Vector Machine	0.75
User Predictions	0.75

According to Table 7, the AUC score for the host win user prediction ratio is 0.75, compared to the AUC score of 0.75 for the Support Vector Machine (SVM) algorithm. Both probabilities are equal.

Evaluation metrics, including accuracy and F1-score obtained from the algorithms, are presented side by side in Table 8 for easy comparison of their values.

Table 8. Algorithm Performance Results Based on Accuracy and F1-Score.

Algorithm	**User-2-class 1467**		**User-3-class 1900**		**Binary-Classification 1467**		**3-Classification 1900**	
	F1-score	accuracy	F1-score	accuracy	F1-score	accuracy	F1-score	**accuracy**
MLP	0.66	0.67	0.41	0.52	0.67	0.68	0.39	**0.53**
NB	0.68	0.68	0.45	0.50	0.67	0.67	0.45	**0.50**
Svm	0.69	0.69	0.45	0.47	0.68	0.69	0.47	**0.49**
RF	0.68	0.68	0.43	0.47	0.68	0.69	0.46	**0.49**

As shown in the results of Table 8, the algorithms perform better in the two-class scenario.

In comparing the confusion matrices of the algorithms, it can generally be said that a model with a higher number of TP (True Positives) and TN (True Negatives) and a lower number of FP (False Positives) and FN (False Negatives) performs better, as shown in Table 9. Classes are defined as follows: First = Host Win, Second = Guest Win, Third = Draw. Most misclassifications occur between Host and Draw outcomes due to home-team bias and close match statistics.

In interpreting the confusion matrix for the Random Forest algorithm in the three-class scenario without user predictions:

- In the classification of the first class, 337 samples were correctly classified, 133 samples were incorrectly classified as the second class, and 164 samples were incorrectly classified as the third class.
- In the classification of the second class, 469 samples were correctly classified, 161 samples were incorrectly classified as the first class, and 203 samples were incorrectly classified as the third class.
- In the classification of the third class, 127 samples were correctly classified, 147 samples were incorrectly classified as the first class, and 159 samples were incorrectly classified as the second class.

Table 9. Confusion Matrix of the Algorithms

Algorithm	User-2-class 1467	User-3-class 1900	Binary-Classification 1467	3-Classification 1900
	Confusion Matrix	Confusion Matrix	Confusion Matrix	Confusion Matrix
MLP	[[233 146] [126 329]]	[[248 126 5] [127 313 15] [121 119 7]]	[224410]] [[605228]	[[383 251 0] [195 637 1] [169 264 0]]
NB	[[253 126] [137 318]]	[[227 106 46] [107 277 71] [102 101 44]]	[[408 226] [245 588]]	[[359 174 101] [196 507 130] [171 177 85]]
Svm	[[277 102] [151 304]]	[[207 56 116] [94 219 142] [94 66 87]]	[[360 274] [170 663]]	[[361 107 166] [172 433 228] [139 144 150]]
RF	[[262 117] [147 308]]	[[216 75 88] [116 252 87] [106 93 48]]	[[410 224] [228 605]]	[[337 133 164] [161 469 203] [147 159 127]]

Therefore, it can be concluded that the model performs better in classifying the second class compared to the first and third classes. Based on the confusion matrices in Table 8, it can be inferred that the MLP model performed with higher accuracy for the three-class problem without user predictions and the two-class problem with user predictions compared to other models. Additionally, for the two-class problem without user predictions, both the MLP and Naive Bayes (NB) models performed better than the SVM and Random Forest (RF) models. In interpreting the confusion matrix for the MLP algorithm in the two-class scenario without user predictions, it can be stated that the algorithm correctly classified 1,037 samples (410 + 605) with an accuracy of 0.68, but misclassified 452 samples (228 + 224).

The influential variables according to the Random Forest algorithm are shown in Figure 3. The orange color represents SVM model predictions, and blue represents user prediction ratios.

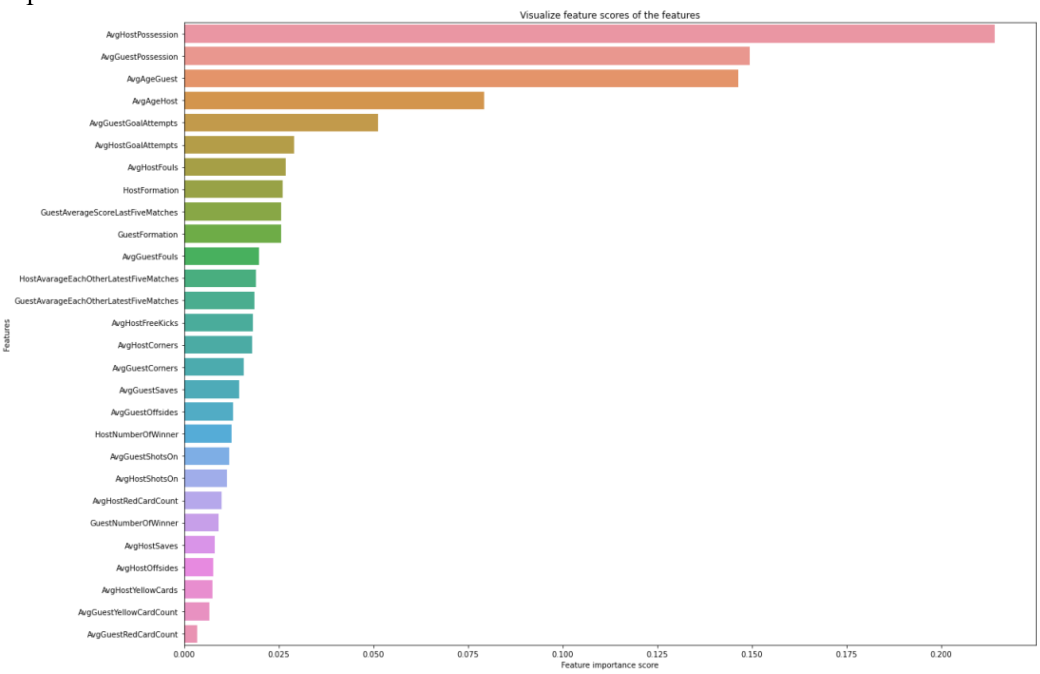

Figure 4. Features in order of effect.

According to Figure 4, the five most influential features, in order, are: AvgHostPossession, AvgGuestPossession, GuestAverageAge, AvgHostAge, and HostGoalAttempts.

6. Discussion and Conclusion

Added for Reviewer: While prior studies employ different datasets and league contexts, including them in the comparison helps to situate our results within the broader field of sports outcome prediction. The performance trends and algorithmic comparisons remain informative for understanding relative advances, despite variations in data sources. Added for Reviewer: The importance of features such as the average team age and possession rates reflects their strong correlation with match dominance and the strategic maturity of players. An older average age often signals higher experience, which in high-stakes matches can translate into better game management and increased win probability. Possession-related features indicate control over the match tempo and are thus naturally predictive of outcomes.

Added for Reviewer: Although a simple train-test split was used in the reported results to align with similar prior studies and maintain reproducibility, we acknowledge that cross-validation (e.g., 5-fold CV) could provide more robust performance estimates. This limitation of the present work will be addressed in future studies, which will incorporate cross-validation to enhance the reliability of the reported metrics.

This study implemented four machine learning algorithms—Neural Networks, Bayesian Networks, Support Vector Machines (SVM), and Random Forests—across four distinct scenarios: with and without the inclusion of draw outcomes, and with and without user-generated predictions. The results revealed that all algorithms performed more effectively in the binary classification (win-loss) scenario, particularly in terms of accuracy and F1-score.

Among the tested models, SVM and Random Forest consistently achieved the highest performance in the two-class scenario, both with and without user predictions, reaching an accuracy of 69% and an F1-score of 68%. The Bayesian algorithm also demonstrated competitive results in the two-class scenario with user predictions, matching the F1-score and accuracy of 68%, and outperforming the neural network model.

In the two-class win-loss scenario with user predictions, SVM delivered the best overall performance. Without user predictions, the neural network model achieved an accuracy of 68%. In contrast, the three-class scenario (win-draw-loss) showed reduced performance: the neural network achieved 43% accuracy without user predictions and 51% accuracy with them. Notably, in the three-class scenario with user predictions, SVM recorded the highest F1-score at 45%. The highest accuracy in both classification scenarios was observed when three user-derived features—predicted probabilities of home win, guest win, and draw—were included. This suggests that user predictions contributed positively to model performance. In scenarios without user predictions, the most influential features included average team age, possession percentage, team formation, average scores from the last five matches, and average goal attempts. When user predictions were incorporated, the most impactful features were the predicted probabilities and average team age.

Conversely, the least influential features across both scenarios were the average number of yellow and red cards, indicating limited predictive value.

A comparison with previous studies (see Table 10) shows that the present research achieved higher accuracy than most reviewed works. These findings underscore the effectiveness of incorporating user predictions and carefully selected features in improving model performance.

Table 10. Comparison of Prior Studies on Football Match Outcome Prediction.

Title	Year	Method	Sample	Features (N)	The highest accuracy	Matches (N)
Reed & O'Donoghue [30]	2005	Neural Network, Regression	English Premier League	7	57.9%	498
Joseph, Fenton, & Neil [31]	2006	Neural Networks, Decision Trees, Bayesian Networks	Tottenham Hotspur Football Club	4	59.2%	76
McCabe & Trevathan [32]	2008	Multilayer Perceptron Neural Networks	World Cup	19	54.6% (3)	-
Odachowski & Grekow [33]	2012	Bayesian Networks, Support Vector Machine	Baseball Seasons from 1967 to 2006	320	46%	1116
Danisik, Lacko, & Farkas [34]	2018	Regression	Video Game Called FIFA	139	52.5%	1520
Berrar, Lopes, & Dubitzky [35]	2019	Neural Network	Results of 21,6743 Football Matches	66	52.4%	216743
Yao, Wang, Zhu, Cao & Zeng [36]	2020	ML Prediction Models	NBA League	-	-	-
Rudrapal, Boro, Srivastava, & Singh [37]	2020	Neural Networks	UEFA Champions League 2015-2016	20	73.57%	11400
Jain, Quamer, & Pamula [38]	2021	DM Algorithms	Indian Premier League	-	70.03%	-
Rodrigues & Pinto [9]	2022	Neural Networks, Random Forest, Support Vector Machine, CS5, Bayesian Networks	English Premier League	-	63%	600

Note: The table has been retitled as 'Comparison of Prior Studies on Football Match Outcome Prediction'. A column has been added to distinguish between 2-class and 3-class classification approaches. Rudrapal et al. [37] used deep learning with image-based spatial features, while the present study relies on structured numerical and user-derived features.

Among the reviewed studies, the highest reported accuracy was 73%, achieved by Rudrapal et al. [37] in a UEFA Champions League prediction model using 40 features. However, the feature selection in that study differs significantly from the present research. The current study demonstrates superior performance compared to most of the existing literature, with only one other study reporting comparable results.

Further analysis using the Area Under the Curve (AUC) metric for the Support Vector Machine algorithm revealed a consistent pattern in user predictions. When the home team won, user predictions closely aligned with the algorithm's output. In contrast, when the guest team won, user predictions were notably less accurate. This suggests a bias among users toward favoring the home team, particularly in cases of uncertainty or lack of clear indicators. The primary innovation of this study lies in the integration of user-generated match predictions with traditional match statistics to improve predictive performance. Unlike prior works focusing solely on historical match features, our approach investigates the cognitive bias of users (favoring home teams) and quantifies its effect on algorithmic accuracy, which has not been previously explored in the literature.

One limitation of this study is its inability to incorporate unpredictable factors that can significantly influence match outcomes. Elements such as player injuries, tactical changes, team composition adjustments, and match-day conditions are dynamic and often difficult to model. Future research should aim to integrate real-time and contextual data to improve predictive accuracy. Another key challenge is the generalizability of machine learning models. Algorithms trained on historical data may not perform reliably when applied to new or evolving scenarios. To address this, future studies should explore techniques such as transfer learning, adaptive modeling, and continuous data integration to enhance robustness and flexibility.

References

[1] I. H. Sarker. "Machine Learning: Algorithms, Real-World Applications and Research Directions", *SN Computer Science*., vol 2, no. 3, pp. 160, March 2021.

[2] R. Bunker, T. Susnjak. "The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review", *The Journal of Artificial Intelligence Research*, vol. 73. Apr 2022.

[3] R. Pugliese, S. Regondi, R. Marini. "Machine learning-based approach: global trends, research directions, and regulatory standpoints". *Data Science and Management,* vol. 4, pp. 19-29. December 2021.

[4] X. Tan. "Enhanced Sports Predictions: A Comprehensive Analysis of the Role and Performance of Predictive Analytics in the Sports Sector". *Wireless Personal Communications*, vol. 132, no. 3, pp. 1613-1636. September 2023.

[5] S. M. Arabzad, M. E. Tayebi Araghi, S. Sadi-Nezhad, N. Ghofrani. "Football Match Results Prediction Using Artificial Neural Networks; The Case of Iran Pro League", *Journal of Applied Research on Industrial Engineering,* vol. 1, no. 3, pp.159-179. Summer 2014.

[6] C. S. Bojer. "Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities". *International Journal of Forecasting,* vol. *38,* no. 4, pp. 1555-1561. October–December 2022.

[7] D. Cica, B. Sredanovic, S. Tesic, D. Kramar. "Predictive modeling of turning operations under different cooling/lubricating conditions for sustainable manufacturing with machine learning techniques", *Applied Computing and Informatics,* vol. 20, no. 1/2, pp. 162-180. 2024.

[8] F. Thabtah, D. Peebles. "A new machine learning model based on induction of rules for autism detection", *Health Informatics*, vol. 26, no. 1, pp. 264-286. 2020.

[9] F. Rodrigues, Â. Pinto. "Prediction of football match results with Machine Learning". *Procedia Computer Science,* vol. *204*, pp. 463-470. 2022.

[10] B. Min, J. Kim, C. Choe, H. Eom, R. I. McKay. "A compound framework for sports results prediction: A football case study". *Knowledge-Based Systems*, vol. 21, no. 7, pp. 551-562. October 2008.

[11] R. Ievoli, L. Palazzo, G. Ragozini. "On the use of passing network indicators to predict football outcomes". *Knowledge-Based Systems*, vol. 222, 106997. June 2021.

[12] I. B. da Costa, L. B. Marinho, C. E. S. Pires. "Forecasting football results and exploiting betting markets: The case of "both teams to score". *International Journal of Forecasting*, vol. 38, no. 3, pp. 895-909. July–September 2022.

[13] W. Gu, K. Foster, J. Shang, L. Wei. "A game-predicting expert system using big data and machine learning". *Expert Systems with Applications*, vol. 130, pp. 293-305. September 2019

[14] R. Baboota, H. Kaur. "Predictive analysis and modelling football results using machine learning approach for English Premier League". *International Journal of Forecasting*, vol. 35, no. 2, pp. 741-755. April–June 2019.

[15] R. G. Martins, A. S. Martins, L. A. Neves, L. V. Lima, E. L. Flores, M. Z. do Nascimento. “Exploring polynomial classifier to predict match results in football championships”. *Expert Systems with Applications*, vol. 83, pp. 79-93. October 2017.

[16] A. Ati, P. Bouchet, R. Ben Jeddou. “Using multi-criteria decision-making and machine learning for football player selection and performance prediction: a systematic review”. *Data Science and Management*, vol. 7, no. 2, pp. 79-88. June 2024.

[17] S. J. Koopman, R. Lit. “Forecasting football match results in national league competitions using score-driven time series models”. *International Journal of Forecasting*, vol. 35, no.2, pp. 797-809. April–June 2019.

[18] Q. Zhang, X. Zhang, H. Hu, C. Li, Y. Lin, R. Ma. “Sports match prediction model for training and exercise using attention-based LSTM network”, *Digital Communications and Networks*, vol. 8, no. 4, pp. 508-515. August 2022.

[19] M. Herold, F. Goes, S. Nopp, P. Bauer, C. Thompson, T. Meyer. “Machine learning in men’s professional football: Current applications and future directions for improving attacking play”. *International Journal of Sports Science & Coaching*, vol. 14, no. 6, pp. 798-817. 2019.

[20] M. Settembre, M. Buchheit, K. Hader, R. Hamill, A. Tarascon, R. Verheijen, D. McHugh. “Factors associated with match outcomes in elite European football – insights from machine learning models”. *Journal of Sports Analytics*, vol. 10, pp. 1-16. 2024.

[21] C. K. Leung, K. W. Joseph. “Sports Data Mining: Predicting Results for the College Football Games”. *Procedia Computer Science*, vol. 35, pp. 710-719. 2014.

[22] Y. Men. “Intelligent sports prediction analysis system based on improved Gaussian fuzzy algorithm”. *Alexandria Engineering Journal*, vol. 61, no. 7, pp. 5351-5359. July 2022.

[23] J. Knoll, J. Stübinger. “Machine-Learning-Based Statistical Arbitrage Football Betting”. *KI - Künstliche Intelligenz*, vol. 34, no. 1, pp. 69-80. March 2020.

[24] P. Pugsee, P. Pattawong. “Football Match Result Prediction Using the Random Forest Classifier”. *Paper presented at the Proceedings of the 2nd International Conference on Big Data Technologies*, Jinan, China. August 2019.

[25] O. B. Sezer, M. U. Gudelek, A. M. Ozbayoglu. “Financial time series forecasting with deep learning: A systematic literature review: 2005–2019”. *Applied Soft Computing*, vol. 90, 106181. May 2020.

[26] C. Schröer, F. Kruse, J. M. Gómez. “A Systematic Literature Review on Applying CRISP-DM Process Model”. *Procedia Computer Science*, vol. 181, pp. 526-534. 2021.

[27] A. Khadivar, F. Abbasi, S. A. Mohseni, R. Basraei. “Utilizing Deep Learning for Aspect-Based Sentiment Analysis in Restaurant Reviews”. *Journal of Information Technology Management*, vol. 17. no. 2, pp. 50-68. 2025.

[28] S. Abdian, M. Hosseinzadeh Shahri, A. Khadivar. “A Bibliometric Analysis of Research on Big Data and Its Potential to Value Creation and Capture”. *Interdisciplinary Journal of Management Studies*, vol. 16, no. 1, pp. 1-24. January 2023.

[29] A. Khadivar, N. Hadianfar, A. Aliyari Boroujeni. “Predicting Mobile Marketing Acceptance Factors: A Hybrid Approach of Structural Equation Modeling and Artificial Neural Network (SEM-ANN)”. *Journal of Business Management*, vol. 17, no. 3, pp. 732-769. 2025.

[30] D. Reed, P. O’Donoghue. “Development and application of computer-based prediction methods”. *International Journal of Performance Analysis in Sport,* vol. 5, no. 3, pp. 12-28. 2005.

[31] A. Joseph, N. E. Fenton, M. Neil. “Predicting football results using Bayesian nets and other machine learning techniques”. *Knowledge-Based Systems, vol.* 19, no. 7, pp. 544-553. November 2006.

[32] A. McCabe, J. Trevathan. “Artificial Intelligence in Sports Prediction”. *Fifth International Conference on Information Technology: New Generations*, Las Vegas. April 2008

[33] K. Odachowski, J. Grekow. “Using Bookmaker Odds to Predict the Final Result of Football Matches”. In: M. Graña, C. Toro, R.J. Howlett, L.C. Jain (eds). Knowledge Engineering, *Machine Learning and Lattice Computing with Applications*. KES 2012. Lecture Notes in Computer Science, vol, 7828. Springer, Berlin, Heidelberg. 2013.

[34] N. Danisik, P. Lacko, M. Farkas. “Football Match Prediction Using Players Attributes”. *World Symposium on Digital Intelligence for Systems and Machines (DISA)*. Slovakia. August 2018.

[35] D. Berrar, P. Lopes, W. Dubitzky. “Incorporating domain knowledge in machine learning for soccer outcome prediction”. *Machine Learning, vol. 108*, no. 1, pp. 97-126. January 2019.

[36] W. Yao, Y. Wang, M. Zhu, Y. Cao, D. Zeng. “Goal or Miss? A Bernoulli Distribution for In-Game Outcome Prediction in Soccer”. *Entropy,* vol. 24, no. 7. July 2022.

[37] D. Rudrapal, S. Boro, J. Srivastava, S. Singh. “A Deep Learning Approach to Predict Football Match Result”. In: H. Behera, J. Nayak, B. Naik, D. Pelusi, (eds) Computational Intelligence in Data Mining. Advances in Intelligent Systems and Computing, vol 990. Springer, Singapore. 2020.

[38] P. K. Jain, W. Quamer, R. Pamula. "Sports result prediction using data mining techniques in comparison with base line model". *OPSEARCH,* vol. *58*, no, 1, pp. 54-70. March 2021.

سوگیری کاربر و دقت الگوریتمی در پیش‌بینی نتایج فوتبال: شواهدی از لیگ برتر انگلستان

فاطمه قاسمی[۱]، آمنه خدیور[۱،*]، لیلا مصلحی[۲] و فاطمه عباسی[۳]

[۱] گروه مدیریت، دانشکده علوم اجتماعی و اقتصاد، دانشگاه الزهرا، تهران، ایران.

[۲] گروه مدیریت ورزشی، دانشکده علوم ورزشی، دانشگاه الزهرا، تهران، ایران.

[۳] گروه مدیریت صنعتی و فناوری اطلاعات، دانشگاه شهید بهشتی، تهران، ایران.

ارسال ۲۰۲۵/۰۸/۲۶؛ بازنگری ۲۰۲۵/۱۰/۰۱؛ پذیرش ۲۰۲۵/۱۰/۳۰

چکیده:

این مطالعه اثربخشی الگوریتم‌های یادگیری ماشین، از جمله شبکه‌های عصبی، شبکه‌های بیزی، ماشین‌های بردار پشتیبان و جنگل‌های تصادفی را در پیش‌بینی نتایج مسابقات فوتبال با استفاده از داده‌های لیگ برتر انگلستان (۲۰۱۸ تا ۲۰۲۲) بررسی می‌کند. با ترکیب احتمالات تولیدشده توسط کاربران برای برد میزبان، برد میهمان و تساوی در کنار ویژگی‌های متعارف، مدل‌ها در دو حالت طبقه‌بندی دوتایی و چندکلاسه مورد ارزیابی قرار گرفتند. در سناریوی برد-باخت، الگوریتم ماشین بردار پشتیبان بیشترین دقت (۶۹٪) را به دست آورد، در حالی که شبکه عصبی در سناریوی برد-تساوی-باخت به دقت ۵۱٪ رسید. نتایج نشان می‌دهد ویژگی‌های استخراج‌شده از داده‌های کاربران عملکرد پیش‌بینی را بهبود می‌بخشند، هرچند پیش‌بینی‌های کاربران به‌ویژه در شرایط عدم اطمینان تمایل به سوگیری به نفع تیم میزبان دارند. این یافته‌ها ظرفیت ادغام دیدگاه کاربران در مدل‌سازی پیش‌بینی را نشان داده و بر اهمیت توجه به سوگیری‌های شناختی در تحلیل‌های ورزشی تأکید می‌کنند.

کلمات کلیدی: پیش‌بینی مسابقات فوتبال، داده‌کاوی ورزشی، الگوریتم‌های یادگیری ماشین، پیش‌بینی.