Preprints.org

Review

A Systematic Review of Machine Learning in Soccer Betting: Techniques, Challenges, and Future Directions

René Manassé GALEKWA [*] , Jean Marie TSHIMULA , Etienne Gael TAJEUNA , Kyamakya KYANDOGHERE [*]

Article

A Systematic Review of Machine Learning in Soccer Betting: Techniques, Challenges, and Future Directions

René Manassé Galekwa [1,2] ⬤, **Jean Marie Tshimula [2,3]**, **Etienne Gael Tajeuna [4]**
and Kyamakya Kyandoghere [1,*]

[1]   Institute of Smart Systems Technologies, University of Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria
[2]   Mathematics, Statistics and Computer Science Department, University of Kinshasa, Kinshasa, DRC
[3]   Department of Computer Science, Université de Sherbrooke, QC J1K 2R1, Canada
[4]   Department of Computer Science and Engineering, Université du Québec en Outaouais, QC J8X 3X7, Canada
*   Correspondence: kyandoghere.kyamakya@aau.at

Abstract: The sports betting industry has experienced rapid growth, driven largely by technological advancements and the proliferation of online platforms. Machine learning (ML) has played a pivotal role in the transformation of this sector by enabling more accurate predictions, dynamic odds-setting, and enhanced risk management for both bookmakers and bettors. This systematic review explores various ML techniques, including support vector machines, random forests, and neural networks, as applied specifically in soccer. These models utilize historical data, in-game statistics, and real-time information to optimize betting strategies and identify value bets, ultimately improving profitability. For bookmakers, ML facilitates dynamic odds adjustment and effective risk management, while bettors leverage data-driven insights to exploit market inefficiencies. This review also underscores the role of ML in fraud detection, where anomaly detection models are used to identify suspicious betting patterns. Despite these advancements, challenges such as data quality, real-time decision-making, and the inherent unpredictability of sports outcomes remain. Ethical concerns related to transparency and fairness are also of significant importance. Future research should focus on developing adaptive models that integrate multimodal data and manage risk in a manner akin to financial portfolios. This review provides a comprehensive examination of the current applications of ML in sports betting, and highlights both the potential and the limitations of these technologies.

Keywords: machine learning; sports betting; soccer; predictive analytics; value betting

---

1. Introduction

Sports betting, traditionally seen as a recreational activity, has become a significant financial sector driven by technological advancements and the proliferation of online betting platforms [1]. The industry allows bettors to place wagers on the outcomes of sporting events with odds that reflect the likelihood of various scenarios. As this market has grown, it has evolved from traditional betting shops to sophisticated online platforms that offer a wide range of betting options, including real-time and in-play bets. The accessibility and convenience of online betting have significantly contributed to the sector's rapid expansion, attracting a global audience and generating billions in revenue annually [2,3].

The growth of sports betting has been paralleled by an explosion of data generation, making it one of the most data-intensive industries [4]. This sector mirrors traditional financial markets, where odds and betting strategies are akin to stock market predictions. Bookmakers collect a large amount of data from various sources, including player statistics, team performance, live game data, and even social media sentiment [5]. This data-driven environment provides fertile ground for the application of machine learning (ML) techniques, which have become essential to managing the complexities of odds setting, risk assessment, and optimization of betting strategy. Machine learning models, particularly those that incorporate real-time data, are crucial to maintaining competitive odds that attract bettors

while ensuring profitability for bookmakers.

Machine learning has significantly impacted the sports betting landscape by improving both the accuracy of predictions and the efficiency of betting strategies. For bookmakers, ML models enable dynamic odds setting and sophisticated risk management, adjusting for new information as events unfold [6]. For bettors, ML provides the tools to develop data-driven strategies that improve the chances of success by identifying value bets and exploiting market inefficiencies [7,8]. As a result, the sports betting industry increasingly resembles a financial sector, with both bettors and bookmakers leveraging advanced predictive analytics to maximize returns. This growing reliance on ML underscores the need for ongoing research on new techniques and emphasizes the importance of ethical considerations, such as transparency and fairness, in the implementation of ML in sports betting. Machine learning, a subset of artificial intelligence, involves the use of algorithms and statistical models to identify patterns and make predictions from data. In the context of sports betting, machine learning techniques can be applied to vast amounts of historical data, including team statistics, player performance metrics, injuries, weather conditions, and even odds movements of bookmakers [9]. By analyzing these diverse data sources, machine learning models can uncover intricate relationships and trends that may not be apparent to human analysts. This leads to the research question (*RQ1*): *how can machine learning algorithms be leveraged to predict match outcomes and maximize profitability in sports betting?*

The application of machine learning in sports betting has garnered significant attention from researchers and industry professionals alike. Numerous studies have explored the use of various machine learning algorithms, such as support vector machines, random forests, neural networks, Bayesian and ensemble methods, to predict the outcomes of sporting events with greater accuracy [10]. These predictive models have the potential to outperform traditional analytical methods and provide valuable insights to bettors, enabling them to make more informed decisions and potentially increasing their profitability.

In addition, machine learning techniques have been employed to identify mispriced odds offered by bookmakers, presenting opportunities for savvy bettors to capitalize on these inefficiencies [11,12]. By developing models that can accurately predict match outcomes and compare them with the odds offered by bookmakers, bettors can identify instances where the odds are mispriced, allowing them to place bets with a positive expected value. Recently, anomaly detection models have been developed to identify suspicious betting patterns that may indicate match-fixing [11,13,14]. These models analyze a range of variables including sports results, team rankings, player data, and betting odds to detect abnormal behaviors that deviate from the expected. Classifying matches as normal, caution, danger, or abnormal based on ensemble model predictions, the system aims to ensure fairness and integrity in sports competitions [15,16].

Despite the promising potential of machine learning in sports betting, there are several challenges and limitations. Data availability and quality can be a significant hurdle, as some sports may have limited historical data or incomplete records. Furthermore, the dynamic nature of sports, with factors such as injuries and team dynamics, can introduce uncertainties that may not be fully captured by predictive models [17]. This raises another critical research question (*RQ2*): *what are the challenges and limitations associated with the application of machine learning in sports betting, and how can novel multimodal approaches be developed to address these issues and improve predictive performance?*

Building on this, a more holistic approach to risk management can be found by drawing parallels to financial portfolio management, where investments are balanced to maximize returns while minimizing risk. Similarly, adaptive betting portfolios could be developed using ML techniques to optimize returns for bettors. This introduces the third research question (*RQ3*): *How can machine learning be applied to create adaptive betting portfolios that optimize returns while minimizing risk, similar to financial portfolio management?* (Table 1).

This systematic review aims to synthesize the current state of research on the application of machine learning techniques in sports betting. By examining the existing literature, we seek to provide a comprehensive overview of the methodologies used, the challenges encountered, and the potential

benefits and limitations of using machine learning in this domain. Furthermore, we explore future directions and opportunities for further research, as the field of machine learning continues to evolve and offers new avenues for innovation in sports betting.

Table 1. Research questions and their purposes.

Research questions	Purpose
RQ1: How can machine learning algorithms be leveraged to predict match outcomes and maximize profitability in sports betting?	To explore the potential of machine learning models to enhance predictive accuracy and maximize profitability in sports betting.
RQ2: What challenges and limitations are associated with the application of machine learning in sports betting?	To identify the existing barriers and constraints that impact the performance and adoption of machine learning models within the sports betting industry.
RQ3: How can machine learning be utilized to develop adaptive betting portfolios that optimize returns while minimizing risk, in a manner analogous to financial portfolio management?	To address a significant gap in the literature by exploring how machine learning can be employed to apply principles from financial portfolio management to the domain of sports betting.

In this review, we limit our scope to soccer. The remainder of the paper is organized as follows. Section 2 presents the methodology employed in this study; Section 3 reviews the related work in the field of machine learning in sports betting; Section 4 delves into the various machine learning techniques applied to sports betting; Section 5 provides a detailed discussion of the findings; Section 6 evaluates the datasets and features used in the studies; Section 7 explores the machine learning platforms for betting tips; Section 8 outlines the challenges and limitations encountered; and finally, Section 9 presents future directions for research in this area.

2. Methodology

The primary objective of this systematic review is to explore the current challenges and advances in applying machine learning techniques to sports betting. The insights derived from this review will serve as a basis for future research in this rapidly evolving field. The research questions and objectives addressed in this study are described in Table 1. This systematic review follows the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to ensure a rigorous and transparent review process (Figure 1). PRISMA was used to structure the review process, including the formulation of research questions, the identification of relevant studies, the evaluation of the quality of the study, the extraction of data, and the synthesis of the findings.
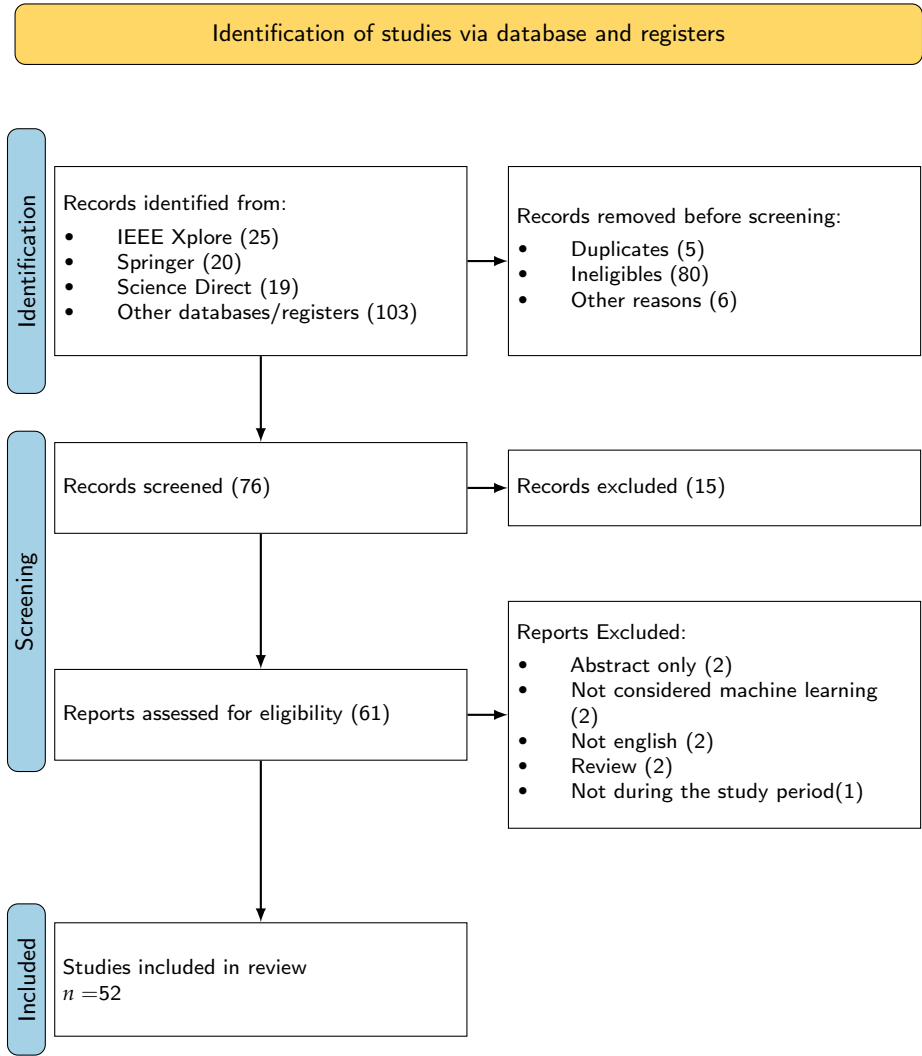
Figure 1. Preferred reporting items for systematic reviews and meta-analyses flowchart of article screening process.

2.1. Data Selection

A comprehensive search was conducted in multiple electronic databases, including IEEE Xplore, Springer, Science Direct, MDPI, arXiv, and Google Scholar. The search terms were carefully selected to capture all relevant studies and included combinations of keywords such as "machine learning", "sports betting", "predictive analytics", "odds estimation", "value betting", and "soccer".

2.2. Inclusion and Exclusion Criteria

The inclusion criteria for this review include studies published between January 2010 and July 2024 that apply machine learning techniques to predict outcomes in soccer betting. These studies consist of peer-reviewed articles, conference papers, and preprints that evaluate the effectiveness of predictive models. A total of 167 articles were initially identified through the search. After applying the inclusion criteria, 49 articles remained for further analysis (Table 2, Figure 2). Exclusion criteria include non-English publications, studies focusing on machine learning methods, and articles that do not contain empirical data, such as purely theoretical or opinion papers.

Table 2. Number of articles per journal/conference

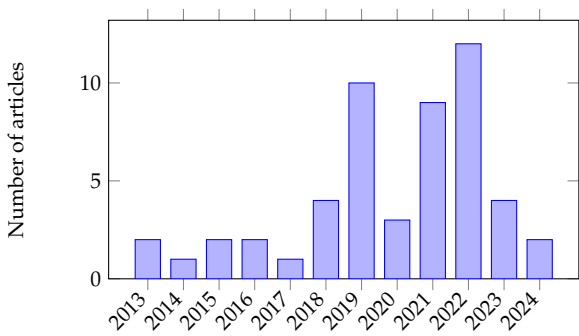| Journal/Conference	Number of Articles
Others	34
MDPI	7
IEEE	3
Journal of Quantitative Analysis in Sports	2
International Conference on Agents and Artificial Intelligence	2
Springer	2
Science Direct	2

Figure 2. Histogram showing the number of articles per year in soccer betting.

2.3. Data Extraction

Data were extracted from each included study, focusing on several key aspects: study details (including author, year, title and journal), the machine learning algorithms used, the characteristics of the dataset (such as size, type and characteristics), performance metrics (including accuracy, precision, recall and Ranked Probability Score), and key findings and limitations.

3. Related Work

The application of machine learning techniques to sports betting has gained significant attention in recent years, and researchers have explored various approaches to leverage data and predictive models to identify profitable betting opportunities.

Outcome prediction. A substantial portion of the literature focuses on the development of predictive models to accurately forecast the results of sporting events. Bunker and Thabtah [18] proposed a machine learning framework for the prediction of sports results, evaluating the performance of several algorithms, including support vector machines (SVMs), decision trees and neural networks, on various sports datasets. Their findings suggest that advanced feature engineering, incorporating factors such as team form, head-to-head records, and home advantage, can improve predictive performance compared to using basic box-score statistics.

Miljković et al. [19] explored the use of data mining techniques, such as decision trees and neural networks, to predict the outcomes of basketball matches. Their results demonstrated the potential of these methods to outperform traditional statistical models, particularly when incorporating diverse features such as team rankings, player statistics, and betting odds. Horvat and Job [7] conducted an initial review of machine learning techniques in the literature on sports betting, examining more than 100 studies on predicting outcomes. They identified neural networks and SVMs as the most common models and highlighted the importance of feature extraction and selection to enhance prediction accuracy. The review also pointed out the lack of standardized datasets and the need to include contextual factors such as player injuries and psychological states.

Kollár [20] discussed the use of Artificial Neural Networks (ANN), Markov chains, and SVMs to handle complex patterns and enormous data sets to forecast sports results. They emphasized the

dynamic nature of sports events, which poses challenges such as inconsistent data and the need for frequent model retraining. Feature selection and extraction were found to be crucial in improving model performance and accuracy.

Odds estimation and value betting. Another line of research investigates the use of machine learning models to estimate the true probability of outcomes, which can then be compared to bookmakers' odds to identify value bets–examples where the model's predicted probability differs significantly from the implied odds. Franck et al. [21] examined inter-market arbitrage opportunities in betting markets, highlighting the potential to exploit inefficiencies and biases in bookmakers' odds-setting processes.

Betting strategies and risk management. Walsh and Joshi [10] conducted a comprehensive study on the importance of model calibration in sports betting. They showed that optimizing predictive models for calibration, rather than accuracy, leads to significantly higher betting profits, with a calibration-optimized model generating 69.86% higher average returns compared to an accuracy-optimized model. Building on this idea of model optimization in the sports betting market, Arscott [22] showed that illegal bookmakers engage in risk management activities 6.5 times more frequently than legal bookmakers. Offshore bookmakers employ pricing policies to balance their books, reducing cash flow variance but also decreasing profits, with illegal bookmakers adjusting commissions on 39% of their prices, compared to 6% for legal bookmakers. This distinction highlights the differing priorities between illegal and legal bookmakers, where illegal firms prioritize risk management due to their limited access to external financing.

Continuing the examination of sports betting strategies, Matej et al. [23] conducted an experimental review of the most popular betting approaches using modern portfolio theory and the Kelly criterion. They demonstrated that formal investment strategies, when applied with risk control modifications, significantly enhance profitability. Their adaptive fractional Kelly method was especially effective in different sports, highlighting the practical importance of mitigating the unrealistic assumptions inherent in pure mathematical strategies. Testing in horse racing, basketball, and soccer confirmed the necessity of these risk control methods to achieve optimal results.

Technological advancements in sports analytics. Keys et al. [24] conducted a systematic review investigating innovative techniques to monitor training loads for the prediction of injury and performance. They highlighted the use of Global Positioning System (GPS), accelerometers, and Rated Perceived Exertion (RPE) to track and predict athlete performance and injury risk. Machine learning was noted for its potential to identify important predictive features, though standardized methods and more research are needed to optimize its application in sports.

Naik et al. [25] conducted an extensive investigation into computer vision in sports, concluding with significant improvements in video analysis. This review included recognition of sports players and balls, tracking, trajectory prediction, and event classification. It emphasized how AI and machine learning improve accuracy and efficiency while tackling challenges such as occlusions, low-resolution video, and real-time analysis. For sports video analysis, the review suggested the use of complex algorithms, standard datasets, and GPU-based workstations and embedded platforms.

4. Machine Learning in Soccer

Machine learning techniques have been extensively applied in various sports betting scenarios, demonstrating their potential to improve prediction accuracy and profitability. Research has demonstrated the effectiveness of models, including artificial neural networks, support vector machines, and ensemble methods in sports (soccer, basketball, tennis, cricket, American football, baseball, horse racing, rugby, golf, and hockey). These models leverage vast datasets, including historical match data, player statistics, and betting odds, to uncover patterns and trends that inform betting strategies. For instance, in soccer, methods such as the Rank Probability Score (RPS) and Principal Component Analysis (PCA) are utilized to identify betting inefficiencies and predict the outcomes of matches (Table 3).

Figure 3 illustrates a comprehensive system for predicting sports betting outcomes using machine learning. The construction of such machine learning models includes multiple data sources, including external factors (for example, weather and match location), historical performance data, real-time betting odds, expert analysis, and general public sentiment from social media.
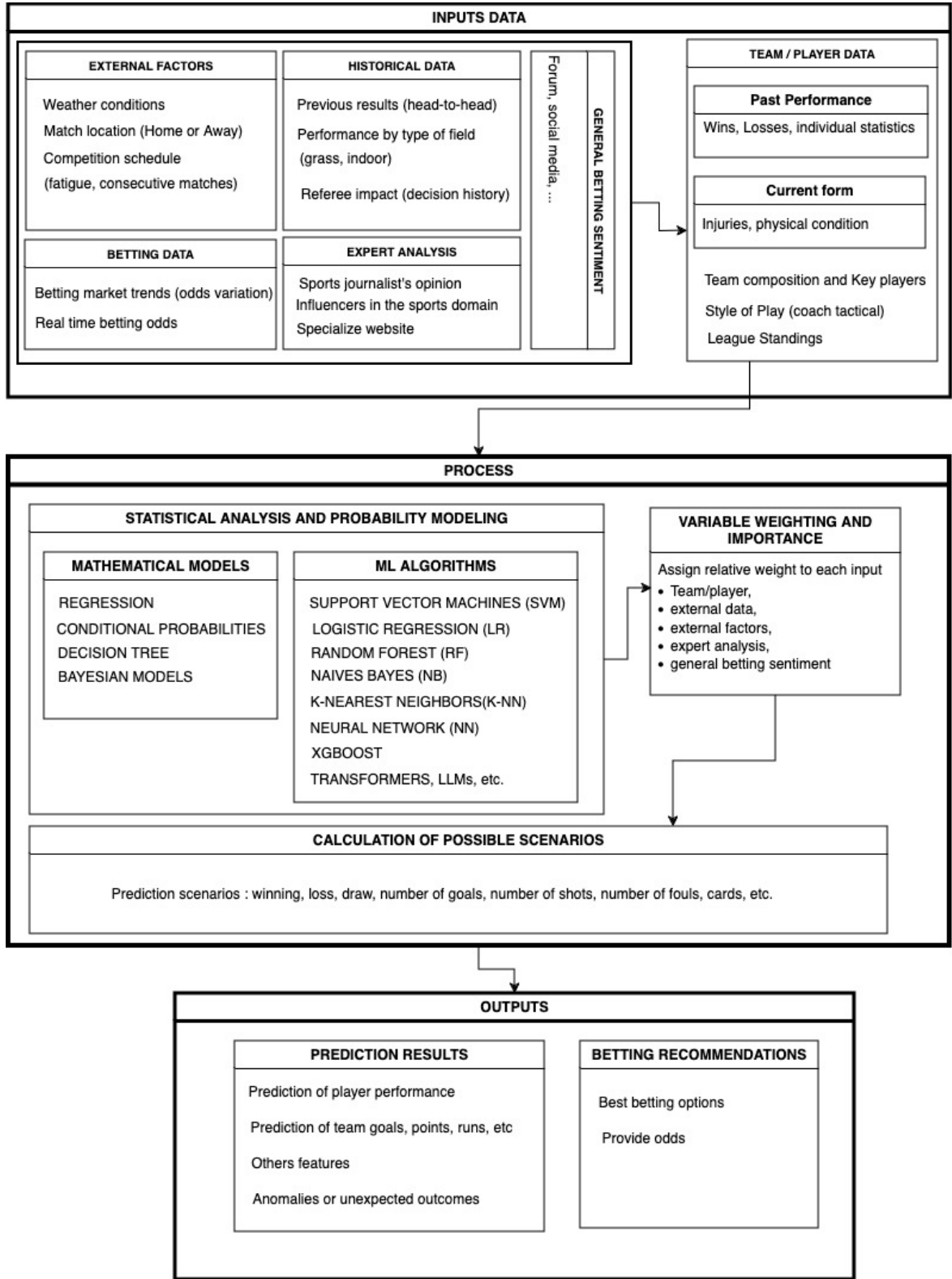
Figure 3. General overview of how to construct machine learning models for predicting sports outcomes, calculating odds, recommending betting options, and more.

The system generates outputs such as probability-based predictions for match outcomes (e.g., win, loss, or draw), detailed performance metrics like goals scored, shots taken, and cards received, and betting recommendations including Bet Boosts. By integrating historical data, real-time statistics, expert insights, and betting market trends, the model enhances prediction accuracy and offers bettors and stakeholders informed betting options and odds. This progression aligns with the broader advancements in soccer analytics, where numerous studies have explored methodologies for predicting match outcomes, player performance, and tactical strategies. The following sections review key contributions in the field, focusing on methodologies, datasets, and results, as summarized in Table 3 and illustrated in Figures 2 and 4.

Table 3. Summary of soccer studies grouped by methodological families

| Family	Model	Work	Performance	Metrics	Features	Datasets
Traditional and predictive methods	Logistic regression, linear regression, bayesian networks, SVM, dimensionality reduction	Tax and Joustra [26], Hervert-Escobar et al. [27], Li et al. [28], Andrews et al. [29], Shin and Gasparyan [30], Elmiligi and Saad [31]	Accuracy 56.054%, RPS 0.2176, R² up to 0.948, AUC 80%	Prediction accuracy, AIC, BIC, RMSE, RPS, AUC	Dimensionality reduction, historical patterns, team ranking, player attributes	www.football-data.co.uk, Open International Soccer Database, FIFA Soccer Rankings, FIFA World Cup datasets, EPL matches
Ensemble methods	Random forest, CatBoost, XGBoost	Hubáček et al. [32], Toda et al. [33], Joseph [34], Malamatinos et al. [35]	Accuracy 67.73%, RPS 0.197, MAE 0.62	Accuracy, RPS, cross-entropy, MAE	Classifier combinations, ELO ratings, team rankings	Open International Soccer Database, EPL matches, 1993/1994 to 2020/2021, Belgian Pro League
Deep learning-based methods	Deep neural networks (DNN), graph neural networks (GNN), LSTM, CNN	Wang et al. [36], Goka et al. [37], Deng and Zhong [38], Lee and Jung [39]	AP 0.967, accuracy 99%, F1 score 0.914	Accuracy, F1 score, AP	Spatio-temporal trajectory data, player profiles, game styles, formations	Liverpool FC, Kaggle, Premier League seasons, 11 European countries
Hybrid approaches	Hybrid ensembles, model stacking	Rahman [40], Shin and Gasparyan [30]	Accuracy 58.9%, RPS 0.2176	Prediction accuracy, RPS	Features from multiple families combined, historical match data, ELO ratings	FIFA Soccer Rankings, Kaggle, custom datasets
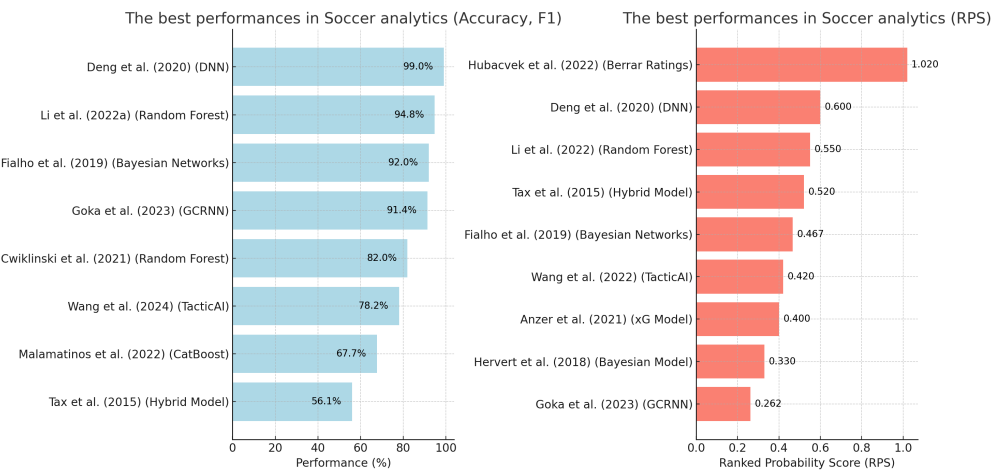
Figure 4. The best performances in Soccer analytics based on accuracy, F1 and RPS

4.1. Analyzing Inefficiencies and Predictive Models in Football Betting Markets

Research on inefficiencies in the online European football gambling market, as conducted by Constantinou and Fenton [41], spanned seven seasons (2005/06 to 2011/12) and examined data from 14 European leagues to identify arbitrage opportunities and biases in bookmakers' odds. By analyzing profit margins and assessing bookmakers' accuracy using the Rank Probability Score (RPS), the study uncovered significant arbitrage opportunities arising from discrepancies in profit margins across bookmakers. Notably, no improvement in the accuracy of odds was observed over time, suggesting persistent inefficiencies. The dataset for this analysis was sourced from www.football-data.co.uk.

In the domain of match outcome prediction, Tax and Joustra [26] investigated results in the Dutch Eredivisie across thirteen seasons (2000–2013). Employing dimensionality reduction techniques like PCA and classification algorithms such as Naïve Bayes, Multilayer Perceptron, and LogitBoost, the study achieved a maximum prediction accuracy of 56.054% using a hybrid model that combined public data with betting odds. Similarly, Hervert-Escobar et al. [27] utilized a Bayesian framework enriched with historical match data and triangular distributions to predict match outcomes, including the 2018 FIFA World Cup group stage. The model, trained on over 200,000 match results, achieved an RPS of 0.2620, showcasing high accuracy. Both studies utilized diverse datasets, including www.football-data.co.uk, www.elfvoetbal.nl, www.transfermarkt.co.uk, and www.fcupdate.nl.

4.2. ML for Soccer Tactics and Event Prediction

Recent advancements in AI have significantly contributed to optimizing soccer tactics and predicting key match events. For instance, TacticAI, developed by Wang et al. [36], leverages geometric deep learning on spatio-temporal player tracking data to enhance football tactics, particularly corner kick strategies. This tool was validated on 7,176 corner kicks from the 2020–2021 Premier League seasons, demonstrating high prediction accuracy and garnering expert approval in 90

Similarly, Goka et al. [37] introduced a novel method for predicting shooting events based on players' spatial-temporal relations modeled through complete bipartite graphs. By combining Mask R-CNN for player detection with a graph convolutional recurring neural network (GCRNN), their approach yielded an impressive AP of 0.967 and an F1 score of 0.914. The study relied on 400 video clips from the 2019 and 2020 Japan J1 League seasons. In a different domain, Anzer and Bauer [42] advanced expected goals (xG) modeling using an extreme gradient boosting algorithm to analyze 105,627 shots from the German Bundesliga. The model excelled with an RPS of 0.197, utilizing data from ChyronHego's TRACAB system alongside Bundesliga event data.

4.3. ML Applications in Player Valuation and Match Outcome Predictions

Machine learning also plays a crucial role in player valuation and match outcome predictions. Li et al. [28] employed machine learning techniques to assess football players' market value using data from FBREF and CAPOLOGY. Comparing multiple linear regression and random forest models, the latter stood out with an R^2 value of 0.948, supported by metrics like AIC, BIC, RMSE, and R^2 values. Additionally, Peters and Pacheco [43] investigated the impact of player lineups on football score predictions using machine learning models, including SVR, which outperformed other techniques in predicting final scores. Their dataset encompassed 680 English Premier League matches (2020–2022), sourced from online repositories like FixtureDownload and FBRef.

In another example, Deng and Zhong [38] analyzed soccer match data from 11 European countries spanning 2008 to 2016. By applying logistic regression, decision trees, random forests, and deep neural networks (DNN), they achieved a peak accuracy of 99% with DNN models. Similarly, Ćwiklinski et al. [44] used machine learning algorithms like Random Forest, Naïve Bayes, and AdaBoost to support team-building and player transfer decisions. Their work, based on Sofascore data covering 4,700 players across four seasons, achieved an accuracy of 0.82 and an F1 score of 0.83. Extending these applications to betting strategies, Stübinger et al. [45] employed ensemble machine learning techniques to predict match outcomes for 47,856 matches in Europe's top five leagues, resulting in a 1.58

4.4. Predictive Modeling Across Leagues and Contexts

Predictive models have been applied across various soccer leagues and contexts, yielding valuable insights. Ganesan and Harini [46] explored the English Premier League (EPL), using SVM, XGBoost, and logistic regression, with XGBoost showing superior performance. Similarly, Andrews et al. [29] leveraged the CRISP-DM framework to predict EPL match outcomes, finding that logistic regression yielded the best F1 score of 0.6119. The use of advanced data sources has also been pivotal; Hubáček et al. [32] evaluated statistical models and rating systems for soccer matches, with Berrar ratings and Double Weibull models excelling in predictions based on 218,916 match results from 52 leagues.

Expanding to international contexts, Mattera [47] employed score-driven models, including the generalized autoregressive score model (GAS), to predict binary outcomes in soccer matches from the EPL and Serie A, analyzing 13 seasons of data. Meanwhile, Malamatinos et al. [35] applied CatBoost to predict outcomes in the Greek Super League, achieving 67.73% accuracy with data spanning six seasons (2014–2020). Finally, Geurkink et al. [48] focused on Belgian professional soccer, where Extreme Gradient Boosting identified critical match outcome variables with 89.6% accuracy using a SportVU-tracked dataset.

5. Discussion

This systematic review demonstrates the transformative impact of machine learning (ML) on soccer analytics, highlighting its applications, limitations, and potential for future development. The diverse methodologies employed across studies indicate the growing sophistication and variety of ML approaches tailored to specific objectives, ranging from match outcome predictions to player performance evaluations and tactical analysis.

5.1. Advances in Machine Learning for Soccer Analytics

The studies reviewed emphasize the versatility of ML in addressing complex problems in soccer. Bayesian approaches, as utilized by Hervert-Escobar et al. [27] and Fialho et al. [49], showcase robust probabilistic frameworks for predicting match outcomes, with significant performance metrics such as the Rank Probability Score (RPS) and prediction accuracy. These approaches provide valuable insights into the uncertainties inherent in soccer matches, particularly in tournaments like the FIFA World Cup.

The integration of deep learning models and spatio-temporal data, as seen in Wang et al. [36] and Goka et al. [37], represents a significant leap forward. These models utilize geometric deep learning and graph convolutional neural networks to capture dynamic player interactions and spatial configurations. Such techniques are invaluable for tactical analysis, particularly in optimizing strategies for set pieces like corner kicks.

Moreover, gradient-boosted algorithms, such as those employed by Anzer and Bauer [42] and Geurkink et al. [48], underline the importance of feature-rich datasets in enhancing predictive accuracy. These studies demonstrate that incorporating synchronized positional data, event streams, and player actions can lead to substantial improvements in expected goals (xG) modeling and overall match analysis.

5.2. Observations from Results

The breadth of ML applications in soccer analytics underscores the significant progress made in this field. For instance, models that integrate historical match data with betting odds, as employed by Tax and Joustra [26], provide practical tools for stakeholders aiming to maximize predictive performance. The study's highest accuracy of 56.054% using a hybrid model highlights the challenges and potential of combining diverse data sources effectively.

Similarly, spatio-temporal tracking data has opened new frontiers in understanding player and team dynamics. Wang et al. [36] demonstrated how geometric deep learning could optimize corner kick strategies, with expert assessments validating the practical utility of these models. This approach bridges the gap between theoretical model performance and actionable insights on the field.

The use of ensemble methods and extreme gradient boosting (Hubáček et al. [32], Geurkink et al. [48]) further illustrates the advantages of integrating multiple ML techniques. Ensemble models have shown superior predictive capabilities by leveraging the strengths of individual algorithms, as evidenced by consistent performance improvements across datasets from elite leagues such as the Bundesliga and Premier League.

Additionally, ML has been instrumental in player evaluation and injury prediction. Studies like Rico-González et al. [50] highlight the potential to enhance decision-making processes for team management and medical staff, using injury prediction models with accuracies exceeding 66%. By analyzing situational and player-specific variables, these models enable proactive interventions to minimize risks and optimize performance.

A notable observation is the role of feature engineering in enhancing model accuracy. For example, Li et al. [28] achieved an R^2 value of 0.948 in player valuation by incorporating comprehensive datasets that included age, goals, and transfer market values. This underscores the necessity of curated features that capture critical aspects of player and match dynamics.

Lastly, the integration of AI tools like TacticAI (Wang et al. [36]) demonstrates the potential for ML to influence tactical decisions directly. High prediction accuracies and favorable expert assessments point to a future where AI-driven insights could become integral to coaching strategies, offering nuanced analyses of gameplay that were previously unattainable.

6. Datasets, Features, and Evaluation Metrics in Sports Analytics

This section provides a comprehensive overview of the datasets, features, and metrics used in various sports prediction models as summarized in Table 3. Table 3 details approaches, works, performance metrics, features, and datasets in soccer. These elements help to understand the methodologies and tools used to predict outcomes in different sports disciplines (**RQ1**).

6.1. Datasets

Various sources have been utilized in the literature, ranging from publicly available repositories to proprietary systems. Public datasets, such as those from www.football-data.co.uk, www.transfermarkt.co.uk, and www.whoscored.com, provide accessible match statistics, player attributes, and historical results. These datasets have been instrumental in predicting match outcomes and analyzing team performance (e.g., Tax and Joustra [26], Andrews et al. [29]).

Proprietary datasets, such as ChyronHego's TRACAB system and SportVU tracking data, offer granular spatio-temporal data, including player trajectories, ball movement, and event streams. These datasets have facilitated advanced analyses, such as geometric deep learning applications in tactical optimization (Wang et al. [36], Goka et al. [37]). However, their restricted access limits widespread use.

The Open International Soccer Database and similar initiatives provide standardized datasets for researchers, covering extensive match histories from multiple leagues and tournaments. These datasets have been essential for studies exploring Bayesian methods, ensemble learning, and expected goals modeling (e.g., Hubáček et al. [32], Anzer and Bauer [42]).

6.2. Features

Features used in predictive models include dimensionality reduction, classifier combinations, historical patterns, team rankings, player attributes, spatio-temporal trajectory frames, event stream data, and player profiles. Common features include team-level statistics, such as possession percentage, shots on target, goals scored, and defensive actions, which are widely used to evaluate team performance and predict match outcomes (Andrews et al. [29], Malamatinos et al. [35]). Player-level attributes, including age, speed, passing accuracy, and goal-scoring ability, are critical for player evaluation and valuation models (Li et al. [28], Ćwiklinski et al. [44]).

Spatio-temporal features, such as data on player positions, ball movement, and event timing, enable analyses of tactical strategies and game dynamics (Wang et al. [36], Goka et al. [37]). External

factors, including weather conditions, home/away status, and historical rivalries, add contextual depth to models (Palinggi [51], Toda et al. [33]).

Advanced studies also incorporate engineered features, such as expected goals (xG) metrics and player valuation indices, derived through complex algorithms and domain expertise (Anzer and Bauer [42], Li et al. [28]). The inclusion of such features has consistently improved predictive accuracy and interpretability.

6.3. Metrics

Evaluation metrics are fundamental in assessing the performance of ML models in soccer analytics. These metrics vary depending on the specific application, such as match outcome prediction, player performance analysis, or tactical optimization. Prediction accuracy, widely used for classification tasks, measures the proportion of correctly predicted outcomes (Tax and Joustra [26], Andrews et al. [29]). The Rank Probability Score (RPS), a probabilistic metric, evaluates the accuracy of predicted outcome distributions and is extensively used in Bayesian and ensemble models (Hervert-Escobar et al. [27], Hubáček et al. [32]).

F1 Score and AUC balance precision and recall, providing a nuanced evaluation of model performance, particularly in imbalanced datasets (Wang et al. [36], Goka et al. [37]). Root Mean Square Error (RMSE), commonly employed in regression tasks, quantifies prediction errors in continuous variables, such as player valuations (Li et al. [28]). Cross-Entropy and Log-Loss evaluate probabilistic models by penalizing inaccurate probability estimates (Hubáček et al. [32]).

In addition to traditional metrics, studies increasingly leverage human expert assessments and domain-specific indicators, such as tactical success rates and injury prevention effectiveness, to evaluate the practical utility of ML models in real-world scenarios (Wang et al. [36], Rico-González et al. [50]).

7. Machine Learning Platforms for Betting Tips

Machine learning platforms specializing in the commercialization of predictive analytics and insights offer bettors valuable tools to build their bet slips and parlays. **Sports AI** presents itself as a comprehensive AI-driven solution for sports bettors, leveraging machine learning to identify value bets and potentially profitable betting opportunities in various sports and bookmakers (https://www.sports-ai.dev/). Similarly, **DeepBetting**, a French startup, sells betting tips based on machine learning and deep learning algorithms trained on historical sports data, covering major football leagues (https://deepbetting.io/). **BetIdeas** analyzes statistics from more than 500 leagues around the world to provide free AI betting tips, exact score predictions, both teams to score tips, and a "bet of the day" feature (https://betideas.com/). Another notable platform, **1x2AI**, quantifies the confidence level of the predictions, allowing users to filter the tips based on how certain the algorithms are of the outcome (https://1x2.ai/tips/).

Leans.ai covers multiple sports such as soccer, tennis, basketball, and esports, offering a free trial and a 60-day money-back guarantee (https://leans.ai/). **FindYourBettingTips** focuses on AI sports betting predictions for football matches and publishes articles explaining the rationale behind each tip, supported by a Telegram community of over 5,000 members (https://findyourbettingtips.com/). **PredictBet** provides AI betting tips for matches up to two weeks in advance, together with a blog on sports news and betting insights (https://predictbet.ai/). **BetQL** offers detailed match predictions, expected value calculations, smart money tracking, and more using AI models (https://betql.co/). **WinnerOdds** employs AI algorithms to estimate real probabilities of match outcomes and find value bets, including tools like odds comparison and variable staking plans (https://winnerodds.com/). **BettorView** offers an AI betting tips free trial and boasts AI models that have achieved a 60% ROI for NFL picks since 2016 (https://www.bettorview.com/). Lastly, **Scaleo** is an affiliate marketing platform that provides machine learning solutions for user segmentation, predictions of betting behavior, risk management and compliance checks in the gambling industry (https://www.scaleo.io/).

8. Challenges and Limitations

The application of machine learning in sports betting presents several challenges and limitations that researchers and practitioners must navigate to enhance predictive accuracy and operational effectiveness. This section discusses key issues, including data availability, the dynamic nature of sports, model overfitting, feature selection, ethical concerns, computational resources, and regulatory challenges (**RQ2**).

8.1. Data Availability and Quality

Data availability and quality represent a significant challenge in soccer analytics. Despite the increasing accessibility of publicly available datasets, many studies rely on proprietary data sources such as player tracking systems and event data providers (e.g., ChyronHego's TRACAB system, SportVU, and Wyscout). Public datasets often lack granularity and completeness, hindering the development of highly accurate models. Furthermore, inconsistencies in data collection methods and annotation standards can lead to noisy or biased datasets, as seen in Hubáček et al. [32], who highlighted variations in dataset quality across leagues. Addressing these challenges requires collaboration between data providers, clubs, and researchers to establish standardized data collection protocols.

In addition, the dynamic nature of data generation in soccer further complicates these issues. Player performance, team dynamics, and even environmental factors like weather are recorded differently across leagues and seasons, creating disparities that challenge the reproducibility of research findings. Several studies have noted that the proprietary nature of high-quality datasets often excludes researchers from smaller institutions or regions with limited resources, thus limiting innovation and diversity in research approaches. Comprehensive, accessible, and standardized datasets are imperative to fostering progress in soccer analytics.

8.2. Dynamic Nature of Sports

The dynamic and unpredictable nature of soccer poses inherent challenges for modeling and prediction. Soccer involves numerous factors such as tactical adjustments, player injuries, and referee decisions that are difficult to capture in static datasets. Models like TacticAI [36], which incorporate spatio-temporal tracking, attempt to address these complexities, but the high variability in match outcomes and player behavior remains a limitation. As highlighted by Goka et al. [37], even advanced methods struggle to account for rare events such as penalties or last-minute goals, which significantly influence match results.

Moreover, the temporal evolution of tactics and player roles adds another layer of complexity to predictive modeling. For example, the increasing use of high-press strategies in modern soccer has shifted traditional metrics of success, such as ball possession, to include more nuanced indicators like pressing efficiency. Models need to evolve continuously to capture these tactical shifts accurately. Without real-time data integration and adaptability, models risk becoming outdated, thus diminishing their predictive reliability.

8.3. Model Overfitting and Generalization

Overfitting is a common issue in machine learning models, particularly when training on small or homogeneous datasets. Models like deep neural networks [38], which achieve high accuracy on training data, often fail to generalize to new or unseen matches. Ensemble methods [32] and regularization techniques have been employed to mitigate overfitting, but the trade-off between model complexity and generalization remains an open research question. Cross-validation and the use of diverse datasets spanning multiple leagues and seasons are essential to improve generalizability.

Despite these efforts, the challenge of generalization persists, particularly when applying models across leagues with varying play styles and data quality. For instance, a model trained on the English Premier League may underperform when applied to the Japanese J1 League due to differences in tactical preferences and player characteristics. Addressing these discrepancies requires not only diverse

datasets but also the integration of transfer learning techniques to enable cross-league adaptability. This ensures models are robust and applicable in varied contexts.

8.4. Feature Selection and Engineering

Feature selection and engineering are critical for developing robust soccer analytics models. Many studies, such as Berrar et al. [52] and Rico-González et al. [50], emphasize the importance of domain knowledge in identifying relevant features. However, the inclusion of redundant or irrelevant features can lead to overfitting and reduced model performance. Methods like dimensionality reduction [26] and feature importance analysis [42] are commonly used to address this challenge. The dynamic nature of soccer also necessitates adaptive feature engineering to account for evolving tactics and player roles.

A deeper challenge lies in the interpretability of engineered features, especially when using complex methods like neural networks. Black-box models often make predictions based on intricate feature interactions that are difficult to explain, which may hinder their acceptance among coaches and analysts. Transparent feature selection processes and explainable AI techniques are crucial to bridge the gap between advanced analytics and actionable insights, ensuring models are not only accurate but also interpretable and trustworthy.

8.5. Ethical and Integrity Concerns

The application of machine learning in soccer raises ethical and integrity concerns, particularly regarding data privacy and the potential misuse of predictive models. For instance, player performance analytics may lead to biased scouting or contract decisions if not transparently implemented. Additionally, models predicting match outcomes [53] could be exploited for gambling purposes, raising questions about their impact on the integrity of the sport. Researchers must ensure ethical data usage and consider the broader implications of their work.

Furthermore, the increasing reliance on machine learning could exacerbate inequalities in access to technology and resources among teams and leagues. Wealthier organizations with access to advanced analytics may gain a disproportionate advantage, potentially widening the competitive gap in soccer. Ethical frameworks and regulatory oversight are needed to ensure fair play, protect player rights, and prevent the misuse of analytical tools, fostering a more equitable application of technology in the sport.

8.6. Computational Resources

Advanced machine learning models, such as geometric deep learning [36] and extreme gradient boosting [48], often require substantial computational resources. These requirements pose barriers for smaller research teams and organizations with limited budgets. Cloud-based platforms and collaborative initiatives can help address these limitations by providing access to high-performance computing resources. However, the computational cost of processing large-scale datasets and training complex models remains a significant challenge.

In addition to hardware limitations, the energy consumption associated with training large models raises environmental concerns. As machine learning becomes increasingly central to soccer analytics, researchers must explore energy-efficient algorithms and computational frameworks. Initiatives that balance model performance with sustainability are critical to ensuring the long-term viability of analytics-driven approaches in soccer.

8.7. Regulatory and Legal Challenges

The use of machine learning in soccer is subject to various regulatory and legal challenges, particularly concerning data ownership and intellectual property rights. Many datasets used in research, such as those from Li et al. [28] and Hubáček et al. [32], are sourced from proprietary platforms, raising questions about licensing and data sharing. Furthermore, the development of predictive models for gambling purposes may conflict with regulations in certain jurisdictions. Addressing these challenges requires clear legal frameworks and ethical guidelines to govern the use of machine learning in soccer analytics.

Moreover, cross-border collaborations in soccer analytics often encounter legal discrepancies, as data privacy regulations such as GDPR in Europe differ significantly from laws in other regions. Researchers and organizations must navigate these complexities to ensure compliance while fostering international partnerships. Establishing global standards for data usage and model deployment can facilitate collaboration and innovation while addressing legal uncertainties.

9. Future Directions

In traditional financial markets, portfolio optimization is a well-established strategy in which investors allocate assets in a way that balances risk and return, with the aim of maximizing profitability. This concept, rooted in Modern Portfolio Theory [54], involves the selection of a mix of stocks, bonds, or other financial instruments that together provide the best possible return for a given level of risk. Portfolio management relies heavily on data analysis, predictive modeling, and optimization techniques to dynamically adjust asset allocations based on market conditions and investor goals. ML has further revolutionized this field by improving predictive capabilities and enabling real-time adjustments to portfolios that significantly improve decision-making in finance [55].

When we make a parallel with sports betting, the concept of a 'betting portfolio' is similar to financial portfolio management, aiming to optimize bet combinations to maximize returns and minimize risk. In such a context, ML can play a key role by analyzing vast datasets, including game results, player statistics, odds, and external factors like weather and team morale. Hence, ML models can be exploited to design diversified betting portfolios that adapt dynamically to game conditions, much like financial portfolios adjust to market changes (**RQ3**). Beyond win-loss predictions, ML could also be designed to enable sophisticated portfolio management, treating bets as assets that affect overall risk and return [56].

As the use of ML in betting grows, there is also a critical need for transparency in the models being deployed. In the financial world, transparent models are increasingly valued for their ability to provide interpretable insights into how investment decisions are made, which is essential to maintain investor trust and regulatory compliance. Similarly, in sports betting, the adoption of Explainable AI techniques, such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), can demystify complex ML predictions and provide bettors and stakeholders with clear, understandable explanations of why certain bets are favored or why odds are set in a particular way [57]. This transparency is not only vital for ethical considerations, but also helps bettors make more informed decisions, ultimately contributing to a more fair and accountable betting environment [58].

10. Conclusions

We have explored the impact of machine learning on soccer and have highlighted its potential to be leveraged as a key component of a financial portfolio. The integration of ML into sports betting marks a major shift, transforming the industry into a data-driven sector with strong parallels to traditional financial markets. With machine learning, diverse datasets can be analyzed, such as historical game data, real-time player statistics, and social media sentiment. This capability improves predictive accuracy and optimizes betting strategies. By treating bets as assets within a 'betting portfolio', similar to financial portfolio management, machine learning enables dynamic adjustment of strategies, improving overall risk and return for bettors and bookmakers as conditions evolve.

As machine learning models such as deep learning and reinforcement learning advance, they offer opportunities to elevate sports betting into a sophisticated investment strategy similar to stock trading. These models facilitate the creation of adaptive betting portfolios that optimize returns and manage risks, driving profitability in a competitive market. The emphasis on transparency and explainability will be essential for maintaining ethical standards and regulatory compliance. By fully embracing these technologies, sports betting can evolve from a game of chance into a strategic financial activity, unlocking new growth opportunities and positioning itself alongside traditional financial sectors.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/1010000/s1.

References

1. Williams, R.; Rehm, J.; Stevens, R. The social and economic impacts of gambling. Technical report, Faculty of Health Sciences, 2011.
2. Gainsbury, S.; Tobias-Webb, J.; Slonim, R. Behavioral economics and gambling: A new paradigm for approaching harm-minimization. *Gaming Law Review* **2018**, *22*, 608–617.
3. Galekwa, R.M.; Tshimula, J.M.; Tajeuna, E.G.; Kyandoghere, K. A Systematic Review of Machine Learning in Sports Betting: Techniques, Challenges, and Future Directions. *arXiv preprint arXiv:2410.21484* **2024**.
4. Forrest, D.; Simmons, R. Sentiment in the betting market on Spanish football. *Applied Economics* **2008**, *40*, 119–126.
5. Haghighat, M.; Rastegari, H.; Nourafza, N.; Branch, N.; Esfahan, I. A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal* **2013**, *2*, 7–12.
6. Thabtah, F.; Zhang, L.; Abdelhamid, N. NBA game result prediction using feature analysis and machine learning. *Annals of Data Science* **2019**, *6*, 103–116.
7. Horvat, T.; Job, J. The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2020**, *10*, e1380.
8. Haruna, U.; Maitama, J.; Mohammed, M.; Raj, R. Predicting the outcomes of football matches using machine learning approach. In Proceedings of the International Conference on Informatics and Intelligent Applications. Springer, 2021, pp. 92–104.
9. Hubáček, O.; Šourek, G.; Železný, F. Exploiting sports-betting market using machine learning. *International Journal of Forecasting* **2019**, *35*, 783–796.
10. Walsh, C.; Joshi, A. Machine learning for sports betting: Should model selection be based on accuracy or calibration? *Machine Learning with Applications* **2024**, p. 100539.
11. Ramirez, P.; Reade, J.J.; Singleton, C. Betting on a buzz: Mispricing and inefficiency in online sportsbooks. *International Journal of Forecasting* **2023**, *39*, 1413–1423.
12. Clegg, L.; Cartlidge, J. Not feeling the buzz: Correction study of mispricing and inefficiency in online sportsbooks. *arXiv preprint arXiv:2306.01740* **2023**.
13. Kim, C.; Park, J.H.; Lee, J.Y. AI-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling. *Scientific Reports* **2024**, *14*, 6470.
14. Mravec, L. Match-fixing as a Threat to Sport: Ethical and Legal Perspectives. *Studia sportiva* **2021**, *15*, 37–48.
15. Deutscher, C.; Dimant, E.; Humphreys, B.R. Match fixing and sports betting in football: Empirical evidence from the German Bundesliga. *Available at SSRN 2910662* **2017**.
16. Ibrahim, L. Integrity issues in competitive sports. *Journal of Sports and Physical Education* **2016**, *3*, 67–72.
17. Taber, C.B.; Sharma, S.; Raval, M.S.; Senbel, S.; Keefe, A.; Shah, J.; Patterson, E.; Nolan, J.; Sertac Artan, N.; Kaya, T. A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. *Scientific Reports* **2024**, *14*, 1162.
18. Bunker, R.; Thabtah, F. A machine learning framework for sport result prediction. *Applied computing and informatics* **2019**, *15*, 27–33.
19. Miljković, D.; Gajić, L.; Kovačević, A.; Konjović, Z. The use of data mining for basketball matches outcomes prediction. In Proceedings of the IEEE 8th international symposium on intelligent systems and informatics. IEEE, 2010, pp. 309–312.
20. Kollár, A. Betting models using AI: A review on ANN, SVM, and Markov Chain **2021**.
21. Franck, E.; Verbeek, E.; Nüesch, S. Inter-market arbitrage in betting. *Economica* **2013**, *80*, 300–325.
22. Arscott, R. Risk management in the shadow economy: Evidence from the sport betting market. *Journal of Corporate Finance* **2022**, *77*, 102307.
23. Matej, U.; Gustav, S.; Ondřej, H.; Filip, Z. Optimal sports betting strategies in practice: an experimental review. *IMA Journal of Management Mathematics* **2021**, *32*, 465–489.
24. Keys, G.; Ryan, L.; Faulkner, M.; McCann, M. Workload Monitoring Tools in Field-Based Team Sports, the Emerging Technology and Analytics used for Performance and Injury Prediction: A Systematic Review. *International Journal of Computer Science in Sport* **2023**, *22*, 26–48.
25. Naik, B.T.; Hashmi, M.F.; Bokde, N.D. A comprehensive review of computer vision in sports: Open issues, future trends and research directions. *Applied Sciences* **2022**, *12*, 4429.

doi:10.20944/preprints202501.2060.v1

26. Tax, N.; Joustra, Y. Predicting the Dutch football competition using public data: A machine learning approach. *Transactions on knowledge and data engineering* **2015**, *10*, 1–13.

27. Hervert-Escobar, L.; Matis, T.I.; Hernandez-Gress, N. Prediction learning model for soccer matches outcomes. In Proceedings of the 2018 Seventeenth Mexican International Conference on Artificial Intelligence (MICAI). IEEE, 2018, pp. 63–69.

28. Li, C.; Kampakis, S.; Treleaven, P. Machine learning modeling to evaluate the value of football players. *arXiv preprint arXiv:2207.11361* **2022**.

29. Andrews, S.K.; Narayanan, K.L.; Balasubadra, K.; Josephine, M.S. Analysis on sports data match result prediction using machine learning libraries. In Proceedings of the Journal of Physics: Conference Series. IOP Publishing, 2021, Vol. 1964, p. 042085.

30. Shin, J.; Gasparyan, R. A novel way to soccer match prediction. *Stanford University: Department of Computer Science* **2014**.

31. Elmiligi, H.; Saad, S. Predicting the outcome of soccer matches using machine learning and statistical analysis. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2022, pp. 1–8.

32. Hubáček, O.; Šourek, G.; Železnỳ, F. Forty years of score-based soccer match outcome prediction: an experimental review. *IMA Journal of Management Mathematics* **2022**, *33*, 1–18.

33. Toda, K.; Teranishi, M.; Kushiro, K.; Fujii, K. Evaluation of soccer team defense based on prediction models of ball recovery and being attacked: A pilot study. *Plos one* **2022**, *17*, e0263051.

34. Joseph, L.D. Time series approaches to predict soccer match outcome. PhD thesis, Dublin, National College of Ireland, 2022.

35. Malamatinos, M.; Vrochidou, E.; Papakostas, G.A. On predicting soccer outcomes in the greek league using machine learning. *Computers* **2022**, *11*, 133.

36. Wang, Z.; Veličković, P.; Hennes, D.; Tomašev, N.; Prince, L.; Kaisers, M.; Bachrach, Y.; Elie, R.; Wenliang, L.K.; Piccinini, F.; et al. TacticAI: an AI assistant for football tactics. *Nature communications* **2024**, *15*, 1–13.

37. Goka, R.; Moroto, Y.; Maeda, K.; Ogawa, T.; Haseyama, M. Prediction of shooting events in soccer videos using complete bipartite graphs and players' spatial-temporal relations. *Sensors* **2023**, *23*, 4506.

38. Deng, W.; Zhong, E. Analysis and prediction of soccer games: an application to the kaggle european soccer database. *Insight-Statistics* **2020**, *3*, 1–6.

39. Lee, G.J.; Jung, J.J. DNN-based multi-output model for predicting soccer team tactics. *PeerJ Computer Science* **2022**, *8*, e853.

40. Rahman, M.A. A deep learning framework for football match prediction. *SN Applied Sciences* **2020**, *2*, 165.

41. Constantinou, A.C.; Fenton, N.E. Profiting from arbitrage and odds biases of the European football gambling market. *The Journal of Gambling Business and Economics* **2013**, *7*, 41–70.

42. Anzer, G.; Bauer, P. A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Frontiers in Sports and Active Living* **2021**, *3*, 624475.

43. Peters, G.; Pacheco, D. Betting the system: Using lineups to predict football scores. *arXiv preprint arXiv:2210.06327* **2022**.

44. Ćwiklinski, B.; Giełczyk, A.; Choraś, M. Who will score? a machine learning approach to supporting football team building and transfers. *Entropy* **2021**, *23*, 90.

45. Stübinger, J.; Mangold, B.; Knoll, J. Machine learning in football betting: Prediction of match results based on player characteristics. *Applied Sciences* **2019**, *10*, 46.

46. Ganesan, A.; Harini, M. English football prediction using machine learning classifiers. *International Journal of Pure and Applied Mathematics* **2018**.

47. Mattera, R. Forecasting binary outcomes in soccer. *Annals of Operations Research* **2023**, *325*, 115–134.

48. Geurkink, Y.; Boone, J.; Verstockt, S.; Bourgois, J.G. Machine learning-based identification of the strongest predictive variables of winning and losing in Belgian professional soccer. *Applied Sciences* **2021**, *11*, 2378.

49. Fialho, G.; Manhães, A.; Teixeira, J.P. Predicting sports results with artificial intelligence–a proposal framework for soccer games. *Procedia computer science* **2019**, *164*, 131–136.

50. Rico-González, M.; Pino-Ortega, J.; Méndez, A.; Clemente, F.; Baca, A. Machine learning application in soccer: a systematic review. *Biology of sport* **2023**, *40*, 249–263.

51. Palinggi, D. Predicting soccer outcome with machine learning based on weather condition **2019**.

52. Berrar, D.; Lopes, P.; Dubitzky, W. Incorporating domain knowledge in machine learning for soccer outcome prediction. *Machine learning* **2019**, *108*, 97–126.

53. Stenerud, S.G. A study on soccer prediction using goals and shots on target. Master's thesis, NTNU, 2015.

54. Du Plessis, A.; Ward, M. A note on applying the Markowitz portfolio selection model as a passive investment strategy on the JSE. *Investment Analysts Journal* **2009**, *38*, 39–45.

55. Bartram, S.; Branke, J.; De Rossi, G.; Motahari, M. Machine learning for active portfolio management. *Journal of Financial Data Science* **2021**, *3*, 9–30.

56. Abinzano, I.; Campion, M.; Muga, L.; Raventós-Pujol, A. Sports Betting and The Black-Litterman Model: A New Portfolio-Management Perspective. *International Journal of Sport Finance* **2021**, *16*, 184–195.

57. Ribeiro, M.; Singh, S.; Guestrin, C. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

58. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* **2019**, *1*, 206–215.