Predicting Football Games Using Historic Results and XGBoost

By Archie Norman

***Abstract***

This project aims to develop a model capable of outperforming bookmakers' by placing bets on football matches with a positive expected value (EV). The data used spanned multiple football leagues worldwide, including the top-tier and lower leagues in England. The dataset, covering over 41,000 matches, included information on match outcomes, time, venue, and the teams involved. Extensive data cleaning and feature engineering were conducted, with key features identified as the home/away games-to-points ratio, goals scored, ELO ratings, and ELO differences between teams. The XGBoost algorithm was employed to predict the probabilities of match outcomes (win, loss, or draw), achieving an accuracy of 62.76%, with an ROC AUC of 0.78 and a log loss of 0.83, indicating reliable performance. These predictions were compared to bookmakers’ odds, and bets were placed where a positive EV was identified. Bankroll management was optimised using the Kelly criterion to adjust bet sizes based on predicted probabilities and model uncertainty. Back testing showed great potential for the model to consistently return profits, but this is yet to be applied in a real application.

**Keywords**: Positive EV, XGBoost, betting, sports betting, Kelly criterion.

**Table of Contents**

[**1 Introduction** 1](#_Toc187750659)

[1.1 How betting works 1](#_Toc187750660)

[1.2 How odds are set 2](#_Toc187750661)

[1.3 How it’s possible to make money 4](#_Toc187750662)

[1.4 The Comparison Between Betting and Trading 5](#_Toc187750663)

[**2 Data** 6](#_Toc187750664)

[2.1 Data sources 6](#_Toc187750665)

[2.2 Future matches 7](#_Toc187750666)

[2.3 Bookmakers odds 7](#_Toc187750667)

[2.4 Historical data 7](#_Toc187750668)

[2.5 Cleaning and Feature Engineering 7](#_Toc187750669)

[**3 Prediction** 8](#_Toc187750670)

[3.1 The model 8](#_Toc187750671)

[3.2 Model Summary 10](#_Toc187750672)

[3.3 Back testing 10](#_Toc187750673)

[3.4 Bet selection and risk management, bank roll management 11](#_Toc187750674)

[**4 Results** 11](#_Toc187750675)

[4.1 Model performance 11](#_Toc187750676)

[4.2 Back Testing 13](#_Toc187750677)

[**5 Discussion** 15](#_Toc187750678)

[5.1 Model Performance 15](#_Toc187750679)

[5.2 Back testing and Practical Application 16](#_Toc187750680)

[5.3 Future Improvements 16](#_Toc187750681)

[**6 Conclusion** 17](#_Toc187750682)

# **1 Introduction**

## 1.1 How betting works

Sports betting operates through two main platforms: bookmakers and exchanges. Both allow you to place bets on sporting events, but their operations and profit models differ significantly, which impacts how bets are made and how profits are earned.

**Odds and Probabilities**

Regardless of whether you're betting with a bookmaker or an exchange, the odds you see are essentially inferred probabilities of an event happening. These odds represent the likelihood of an outcome occurring, but each platform has a different way of calculating and offering those odds.

**The Bookmaker (Bookie)**

When betting with a bookmaker, you’re placing your bet directly with them. The bookmaker sets the odds and takes on all the risk. To make a profit, bookmakers add a built-in margin to the odds, often referred to as the “vig” or “juice.” This margin ensures that the total implied probabilities for all outcomes in an event exceed 100%, which reduces the payout compared to the true odds.

For example, if Team A and Team B have equal chances of winning (50% each), the true odds would be 2.0. However, a bookmaker might offer odds of 1.9, implying a probability of 52.6%. This slight reduction in the payout is how bookmakers ensure they make a profit.

Bookmakers manage their risk by “balancing the books.” This means they adjust the odds to encourage more bets on the less popular outcome. For example, if too many people bet on Team A to win, the bookmaker may lower Team A’s odds and increase Team B’s to attract more bets on Team B. This helps balance the betting pool and limits the bookmaker’s potential losses.

Additionally, bookmakers often use hedging strategies to reduce their exposure to high-risk bets. If there is heavy betting on a particular outcome, they might place a bet with another bookmaker or exchange to cover their risk. This process, known as "layoff betting," allows bookmakers to spread their risk and protect themselves against large losses.

**The Exchange**

Exchanges offer a different approach to betting. Instead of betting against the bookmaker, you place bets against other users. Exchanges act as intermediaries, matching bettors who want to "back" an event (bet that it will happen) with those who want to "lay" an event (bet that it will not happen). The exchange takes a commission on winning bets but does not assume any risk themselves.

One of the main advantages of exchanges is flexibility. Bettors can both back and lay bets, allowing them to "trade" positions. For example, if a bettor backs Team A to win and the odds change in their favour, they can "lay" the bet (bet against Team A) to lock in a profit before the match concludes.

However, exchanges depend heavily on liquidity—the availability of matching bets. For a bet to be placed, there needs to be someone willing to take the opposite side. In popular markets, such as top football leagues, liquidity is usually high, meaning bets can be easily matched at competitive odds. On the other hand, smaller or niche events may have less liquidity, making it harder to find a match for your bet or forcing you to accept less favourable odds.

Exchanges make money by charging a commission on winning bets, typically a percentage of the net winnings. Unlike bookmakers, exchanges do not inflate the odds to make a profit, so bettors are more likely to receive better odds. This model aligns the exchange’s interests with those of the bettors, encouraging high-volume trading and the success of its users.

**Summary**

While both bookmakers and exchanges offer ways to bet on sports, they operate under different models. Bookmakers set the odds and assume all the risk, making a profit through the built-in margin. Exchanges, on the other hand, allow bettors to wager against each other and charge a commission on winnings. Understanding these differences can help you make more informed decisions when choosing where to place your bets.

## 1.2 How odds are set

When developing a model to predict football outcomes and place bets, you're essentially competing against the sophisticated systems used by bookmakers. Understanding how they set odds is critical for gaining a competitive edge. However, this is not a simple task. The sports betting industry is highly secretive, with bookmakers investing heavily in proprietary technologies and methodologies to maintain an advantage. As a result, the specific mechanisms behind their odds-setting process are closely guarded. Nonetheless, certain key principles are publicly known, which offer valuable insights into how bookmakers set their odds.

**The Foundation of Odds-Setting**

At its core, odds-setting relies on advanced statistical modelling and the analysis of historical data. Bookmakers use complex algorithms and often incorporate machine learning techniques to estimate the probabilities of various outcomes. This allows them to make highly accurate predictions about match results. However, simply predicting outcomes isn't enough for bookmakers to maximise profits; they also need to manage risk and ensure balanced betting on all potential outcomes. This is why bookmakers continuously adjust their models to account for factors beyond pure statistical probabilities.

**Key Principles in Odds-Setting**

* **Historical Data and Statistical Models**: While historical data provides a foundation for creating models, it’s often a barrier to entry for others looking to develop similar algorithms. High-quality data goes beyond the final score; to create the best models, access to granular data such as timing for each goal, card, foul, and corner is essential. However, obtaining this level of detail is often difficult or expensive, giving established bookmakers an edge.
* **Public Perception and Market Sentiment**: The behaviour of bettors also plays a crucial role in shaping the odds. Bookmakers must anticipate how the public will bet to avoid imbalances that could expose them to risk. For example, in high-profile boxing matches, such as those between Tyson Fury and Deontay Wilder, bookmakers may lower the odds on the UK fighter (Fury) due to social bias among British bettors who favour him over the American (Wilder). This adjustment ensures that wagers on both sides are balanced and that the bookmaker can secure a profit, regardless of the result.
* **Expert Opinions and Contextual Information**: In addition to statistical data, expert analysis and real-time information are key for refining odds. Factors like player injuries, team form, and starting lineups are critical in adjusting probabilities. For instance, in international tournaments where squads may change significantly between matches, historical data may be less reliable. In these cases, expert opinions become even more valuable. A late injury to a key player can shift the probabilities, prompting bookmakers to revise the odds accordingly.
* **Profit Margins ("Vig" or "Juice")**: Unlike true probabilities, which sum to 100%, bookmakers intentionally skew the odds to include a built-in margin that ensures profitability. This margin, often called the "vig" or "juice," is a core component of the bookmaker’s strategy. To manage risk and maintain profitability, bookmakers continuously adjust their odds in response to betting volumes, ensuring a balanced book and protecting themselves from heavy losses on one outcome.

**Conclusion**

The process of setting odds is a complex balance of statistical modelling, expert knowledge, market sentiment, and profit management. While the specific techniques used by bookmakers remain a closely guarded secret, understanding these fundamental principles provides a foundation for bettors to build more effective models and make informed decisions. The interaction between historical data, public perception, expert analysis, and the bookmaker's margin is what ultimately shapes the odds you see, creating both opportunities and challenges for those looking to gain an edge in sports betting.

## 1.3 How it’s possible to make money

The primary objective of sports betting is to generate profit. However, this is no easy feat, given the sophisticated strategies employed by bookmakers to ensure their odds are favourable. While short-term gains may sometimes be attributed to luck, consistently making money over the long term requires a more strategic approach—one focused on identifying and exploiting value bets.

**The Concept of Value Bets**

A value bet occurs when the true probability of an event is higher than the implied probability derived from the bookmaker's odds. Understanding the difference between true and implied probabilities is essential to identifying value.

* **True Probability**: This represents the actual likelihood of an event occurring. For example, in a fair coin toss, the true probability of heads or tails is 50%. Even if the coin is flipped multiple times, the true probability remains unchanged, regardless of the outcome of each flip.
* **Implied Probability**: This is the probability reflected in the bookmaker’s odds. Bookmakers often adjust the odds to account for their profit margin, meaning that the implied probability will typically be higher than the true probability. As a bettor, your task is to identify situations where the bookmaker’s implied probability underestimates the actual likelihood of an outcome.

**Bankroll Management**

Effective bankroll management is another critical factor in achieving long-term success in sports betting. Even with a strategy focused on positive-value bets, poor bankroll management can lead to financial ruin.

Two popular approaches to managing stake sizes are **flat betting** and the **Kelly criterion.**

* **Flat Betting**: This approach involves wagering a fixed amount on each bet, regardless of the odds or perceived value. While this method is simple and easy to follow, it doesn’t consider the varying levels of risk or potential return associated with different bets. It’s a straightforward method, but it may not be the most efficient over the long run.
* **Kelly Criterion**: The Kelly criterion is a more sophisticated approach that calculates the optimal bet size based on the bettor's edge and the odds offered. This method adjusts for both the likelihood of success and the potential payoff, making it a more dynamic and conservative strategy. It helps bettors maximise growth while minimising risk. However, there are some limitations to this approach. For example, when placing many bets, the total stake could exceed the available bankroll. To address this, bettors can sum their total edge across all bets or scale down individual stakes to ensure they do not exceed a predetermined percentage of their total bankroll.

**Conclusion**

In conclusion, making money in sports betting requires more than just luck. By identifying positive-value bets, understanding the true and implied probabilities, and practicing effective bankroll management, bettors can increase their chances of long-term profitability. While bookmakers have the advantage in setting odds, bettors who focus on value and use sound strategies can still find opportunities to profit in the long run.

## 1.4 The Comparison Between Betting and Trading

The comparison between betting and trading is an important topic that highlights both the similarities and differences between these two profit-driven activities. Both involve predicting future outcomes with the goal of making money, yet key distinctions provide valuable insight into the challenges faced by bettors and the unique nature of the betting market.

**Market Maturity: Trading vs. Betting**

One of the most significant differences between the two lies in the maturity of the financial markets compared to betting markets. Financial markets have long been governed by the principles of the Efficient Market Hypothesis (EMH). According to EMH, market prices reflect all available information, making it extraordinarily difficult for individuals or organisations to consistently achieve returns above the market average. This is due to the vast number of participants, each employing advanced technologies, algorithms, and analytics to inform their decisions. The competition among traders drives prices toward their "true" value.

In contrast, sports betting markets are far less mature and efficient than traditional financial markets. Although bookmakers use sophisticated models and algorithms to set odds, these odds often fail to fully reflect true probabilities. External factors such as public perception, market manipulation, and strategic adjustments made by bookmakers to balance their books create inefficiencies in the betting market. These inefficiencies provide opportunities for knowledgeable bettors to exploit value bets.

**The Evolving Betting Market**

While sports betting markets are currently less efficient, their growing popularity and the increasing use of advanced predictive models by both individuals and professional syndicates suggest that these markets could become more efficient over time. If betting odds were to fully incorporate all available information and reflect true probabilities, achieving consistent profits would become significantly more challenging. Unlike financial markets, which tend to produce average returns over time due to economic growth and dividends, betting markets do not have an inherent long-term upward trend. Instead, profits in betting rely entirely on the ability to beat the odds, making it a more volatile and unpredictable environment.

**Betting vs. Trading for Profit**

The key takeaway is that if your primary goal is to make money, the stock market is a more reliable and sustainable option. While trading still requires skill, it offers a structured environment that historically provides steady returns. In contrast, sports betting is inherently more speculative, and achieving consistent profits is difficult, especially as markets evolve toward greater efficiency. Therefore, if you are looking to pursue betting as a means of making money, it should be approached more as a hobby or an interest rather than a primary source of income.

# **2 Data**

## 2.1 Data sources

|  |  |
| --- | --- |
| Data source | Description |
| <https://www.footballwebpages.co.uk/api>  <https://rapidapi.com/football-web-pages1-football-web-pages-default/api/football-web-pages1> | A free API that provides data mostly on English football with some first-tier European leagues.  First link is the providers webpage  Second link is the API website |
| <https://football-data.co.uk/> | Free historical data on football results with odds from multiple bookmakers. |

*Table 1. A list of the sources used to get data.*

## 2.2 Future matches

To obtain match data, I used the above-mentioned API instead of working with manually saved Excel sheets. I opted for the free tier to minimise costs and make the process accessible. The data retrieved from the API includes fixtures and results, which consist of the following fields:

* Date, Time, Venue
* Home Team, Home Score
* Away Team, Away Score
* Competition ID

For matches that haven't occurred yet, the scores are left blank. Since this data is limited, significant feature engineering was required to generate the necessary features for my model.

## 2.3 Bookmakers odds

The primary bookmaker I rely on is Bet365, but web scraping this data is particularly challenging due to the complexity of the website. As a result, odds data must be entered manually, which becomes a tedious process when dealing with large volumes of matches (sometimes hundreds per day). Alternative options, such as Betfair, are easier to scrape, but the main challenge remains that odds can change frequently, meaning that bets need to be placed almost immediately after data retrieval to secure the odds.

## 2.4 Historical data

Some historical data is available through the API, but it lacks odds information, which makes back testing difficult. To address this, I sourced additional data from the [Football Data website](https://football-data.co.uk/), which provides historical football results along with bookmaker odds. While the dataset is limited in size, it is sufficient for basic back testing. For odds, I specifically selected the closing odds from Bet365.

## 2.5 Cleaning and Feature Engineering

Feature engineering focused on metrics such as ELO ratings, goals scored, goals conceded, and specific ratios like points per game. For both teams involved in a match, the data was split into separate rows—one for each team—and relevant metrics were calculated before merging the data back together. This approach simplified the process of obtaining team-specific features. For a full list of features, you can look at the code in the GitHub.

The features were scaled, though XGBoost models are generally insensitive to scaling. This is because XGBoost relies on threshold-based conditions, unlike distance-based methods used by algorithms like KNN or SVM. As a result, no log transformations were necessary. Additionally, XGBoost does not assume equal class distributions, so no class balancing was required. XGBoost’s ensemble nature also mitigates issues related to class imbalance by focusing on correcting misclassifications in subsequent trees.

Finally, before making predictions, categorical features such as competition/league and team names were one-hot encoded. This encoding process converts categorical values into binary features by creating a separate column for each unique value in the original feature.

# **3 Prediction**

## 3.1 The model

For predicting the outcome of football matches, I utilised XGBoost (Extreme Gradient Boosting), an ensemble learning algorithm that builds multiple decision trees to make predictions. The model was trained on the data from the API and the historical data to predict upcoming fixtures. XGBoost was chosen due to its strong performance in various classification tasks, particularly for structured data such as sports statistics. It excels in handling complex relationships, including non-linear interactions between features, and offers several hyperparameters that can be tuned to optimise model performance.

**Model Evaluation**

The model's evaluation was set to log loss, the default metric for XGBoost. Log loss measures both the probability and accuracy of predictions, making it ideal for this task. It penalises incorrect predictions, especially when the model is confident but wrong, which is crucial for applications like sports betting.

**Parameter tuning**

Hyperparameter tuning plays a critical role in improving model performance. GridSearchCV was employed to tune the hyperparameters of the XGBoost classifier, exploring multiple combinations of parameters to find the optimal configuration and a 5 fold cross validation was used to help limit over fitting. The parameters tuned included:

* **max\_depth**: Tree depth
* **min\_child\_weight**: Minimum instance weight in a node
* **subsample**: Fraction of training data for each tree
* **colsample\_bytree**: Fraction of features per tree
* **learning\_rate**: Step size for each tree's contribution
* **n\_estimators**: Number of boosting rounds
* **gamma**: Minimum loss reduction for a split
* **alpha**: L1 regularisation
* **reg\_lambda**: L2 regularisation

These parameters control the model’s complexity and regularisation, balancing bias and variance. The grid search tests various values for each parameter to identify the best settings. However, excessive tuning can lead to overfitting, making careful optimisation essential. Given the resource-intensive nature of grid search, it was performed only when significant changes were made to the model. Its also important to note that the parameters were also checked for over fitting and then some had to be scaled back such as n\_estimators, gamma, and alpha.

**Calibration**

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated

*Figure 1. Calibration curve for uncalibrated (left) and calibrated (right) data.*

Model calibration adjusts raw probability predictions to reflect true likelihoods, enhancing the reliability of predictions. This is particularly important for making accurate betting decisions. Several calibration methods were considered:

* **Platt scaling**: Applies logistic regression to map predicted scores to probabilities, assuming a sigmoid relationship between predictions and actual outcomes. It works well for probabilistic outputs.
* **Isotonic regression**: A flexible, non-parametric method that fits a piecewise constant function to the predictions, ideal for non-linear relationships but requiring ample data to avoid overfitting.
* **Beta calibration**: Uses the Beta distribution to adjust probabilities, effective for skewed prediction distributions or imbalanced datasets.

Cross-validation with calibration ensured that the model generalised well. Calibration curves were used to visualise the calibration, with the ideal curve following the diagonal, indicating that predicted probabilities match observed frequencies. The calibration process revealed that Platt scaling was sufficient for the data, while Isotonic regression was not suitable due to the relatively small dataset. As figure 1 shows the model was already well calibrated and the calibrated only took some of the variation out of the model. The model struggles predicting draws and at high probabilities the calibration made it worse but if a draw is being predicted at a probability above 60% you might be suspicious of the prediction anyway.

### 3.2 Model Summary

Throughout the modelling process, I observed that simpler models performed better. The more I adjusted the distribution of features, the more the model struggled. Despite extensive tuning, the model performed well even with default parameters. Key improvements came from reducing noise by removing low-importance features and adding meaningful features, such as variance in performance.

The final parameters were:

* **eval\_metric**='logloss'
* **max\_depth**=4
* **min\_child\_weight**=10
* **subsample**=0.7
* **colsample\_bytree**=0.6
* **alpha**=5
* **gamma**=1
* **reg\_lambda**=10
* **n\_estimators**=200
* **learning\_rate**=0.05

## 3.3 Back testing

Back testing was performed using data from several European football leagues, including:

* **English** (Premier League, Divisions 1-3, Conference)
* **Scottish** (Premier League, Divisions 1-3)
* **German** (Bundesliga 1 & 2)
* **Italian** (Serie A & B)
* **Spanish** (La Liga Primera & Segunda)
* **French** (Ligue 1 & Div 2)
* **Netherlands** (Eredivisie)
* **Belgium** (Jupiler League)
* **Portugal** (Liga 1)
* **Turkey** (Super Lig)
* **Greece** (Super League)

The dataset included over 41,000 games, spanning the 2024/2025 season back to 2019/2020.

**Back testing Methodology**

A 70-30 split was used for training and testing with the training data being the oldest data and the test data being the newest data. This was done to mimic what would happening in real life and to limit any data leaks from the recent form features. For each game, a 1-unit bet was placed on matches with a positive expected value. Bet365 was chosen as the bookmaker for all the back testing scenarios, as it is the intended bookmaker for real-world bets.

One challenge with back testing is the higher volume of matches in the historical data due to the inclusion of multiple leagues. This is problematic for real-world bets, as the available sample size is smaller, leading to slower results another factor is the API doesn’t have data on these leagues making it unfeasible to bet on all these leagues.

## 3.4 Bet selection and risk management, bank roll management

Several bankroll management strategies were considered. Initially, a flat betting strategy was employed to check for consistency in profits. Later, I incorporated the Kelly Criterion, which is the most efficient bankroll management strategy when the true probabilities are known. However, if the true probabilities are not accurate, the Kelly Criterion can significantly be affected, and performance will be limited.

The Kelly Criterion utilised the predicted probabilities, accounting for the model’s standard error. This ensured that bets were only placed when the model was confident that the probability was higher than the estimated true probability, leading to more conservative bets while ensuring a positive expected value.

# **4 Results**

## 4.1 Model performance

The model’s performance was evaluated using cross-validation and various classification metrics to assess its predictive accuracy and calibration. Key results are summarised below:

**Model Performance**

* **Overall Accuracy**: 62.76%
* **Log Loss**: 0.828, reflecting the model's calibration and confidence in probabilistic predictions.
* **ROC AUC**: 0.783, indicating strong ability to distinguish between classes.

**Classification Metrics**

The model’s classification performance for predicting match outcomes (Loss, Draw, Win) is summarised below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-Score | Support |
| Loss | 0.63 | 0.68 | 0.65 | 688 |
| Draw | 0.44 | 0.25 | 0.32 | 542 |
| Win | 0.67 | 0.80 | 0.73 | 969 |

*Table 2. Performance metrics for the model.*

* **Weighted Average F1-Score**: 0.61

The model performed best in predicting wins, while draws were the most challenging to forecast, with lower recall and F1-score.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted Loss | Predicted Draw | Predicted Win |
| Actual Loss | 466 | 88 | 134 |
| Actual Draw | 163 | 134 | 245 |
| Actual Win | 108 | 81 | 780 |

*Table 3. Confusion matrix for the model*

This distribution indicates a strong performance in predicting wins but relatively weaker accuracy for draws, suggesting the need for model improvements to better distinguish between draws and other outcomes.

**Brier Scores**

* **Loss (Class 0)**: 0.146
* **Draw (Class 1)**: 0.175
* **Win (Class 2)**: 0.166

**Brier Score Explanation**  
The Brier score measures the accuracy of probabilistic predictions, with lower values indicating better performance. It is calculated as the mean squared difference between predicted probabilities and actual outcomes (1 if the event occurred, 0 otherwise). Each score corresponds to a specific class (Loss, Draw, Win), representing how well the model’s predicted probabilities align with actual match outcomes.

* A Brier score of 0 represents perfect probability predictions, while a score of 1 indicates the worst possible prediction.
* In this model, the Brier score is lowest for the "Loss" class, suggesting it provides the most reliable probability estimates for losses. The slightly higher scores for draws and wins reflect less precise probability estimates, particularly for draws, which aligns with the observed difficulty in accurately predicting this outcome.

## A graph of a graph Description automatically generated with medium confidence4.2 Back Testing

*Figure 2. profit over time for each competition.*

This shows the profitability of the of the model with it having varying levels of success in different leagues although some did see a larger profitability than others.

A graph with colored lines and numbers

Description automatically generated

*Figure 3. Calibration curve for the back testing test data.*

The calibration curve for the test data in the back testing seemed very close to true probability showing the low error of the model.

A graph of a function

Description automatically generated with medium confidence

*Figure 4. individual bet winnings by edge (difference between my predictions and the bookmakers) in blue with the distribution of losing bets divided by total bets in yellow.*

The losses quickly drop off when the percentage difference (edge) is larger showing I did find an edge not just luck with the mode

A graph of a number of blue bars

Description automatically generated

*Figure 5. Bookmakers’ odds by percentage of losses at certain odds.*

Essentially the same as a calibration curve as if the odds are 2 then the normalised count of losses needs to be below 0.5.

# **5 Discussion**

The goal of this project was to develop a predictive model capable of forecasting football match outcomes, which could be leveraged for strategic betting decisions. The approach centred on using an XGBoost classifier, a powerful machine learning algorithm, to handle the complex relationships in football data. The model was trained using historical data from a variety of European leagues, spanning multiple seasons, to ensure a comprehensive understanding of team performance patterns and match dynamics.

## 5.1 Model Performance

The results obtained from the model suggest a solid foundation for predicting match outcomes, with an overall accuracy of 62.76%. While this performance is promising, it highlights that there is still room for improvement, especially when it comes to predicting draws, which remained the most difficult outcome to forecast. The model performed particularly well with wins, demonstrating its ability to discern matches where one team is more likely to dominate.

The model’s ROC AUC score of 0.783 indicates that it has a good ability to distinguish between the different match outcomes, with a higher probability of success for wins. However, this metric also emphasises that the model struggles more when distinguishing between losses and draws, as reflected in the lower precision and recall values for the "Draw" class. The skewed distribution of match outcomes, where wins are more common than draws or losses, may contribute to these challenges.

**Calibration and Reliability**

The calibration of the model, as measured by the Brier scores, also provides valuable insights. A Brier score for the "Loss" class of 0.146 suggests that the model’s probability estimates for losses are relatively accurate. In contrast, the Brier scores for draws and wins (0.175 and 0.166, respectively) indicate that the model’s predictions for these outcomes are slightly less reliable. This highlights an opportunity for further improvement in the model’s ability to predict these less frequent events. The calibration curves suggested that the model was generally overestimating the probability of draws and underestimating the likelihood of wins, which might suggest that the model is slightly more conservative in predicting the less likely outcomes.

Given the inherent uncertainty in predicting match outcomes and the complexity of football, these calibration issues were expected. Further calibration methods, such as Platt scaling, were employed to adjust the predicted probabilities, and while they led to minor improvements, the model’s performance still faced challenges in predicting draws accurately. This underscores the importance of considering multiple calibration techniques and testing them thoroughly before drawing conclusions on their effectiveness.

## 5.2 Back testing and Practical Application

Back testing results, using a diverse set of European leagues, showed the model’s ability to identify matches with a positive expected value. However, the high volume of historical matches, spanning multiple leagues, introduces a limitation in real-world betting applications, where the available sample size is often much smaller. This discrepancy between historical data and actual betting conditions highlights the challenges of translating model predictions into real-world profits. The Back testing can be considered successful given it was profitable in all leagues saw accurate calibration but I’m still unsure on the practical application as there are some challenges with mainly in the data for the model as the back testing model had multiple seasons of data on each team before having to make predictions which the free API doesn’t provide due to it being the free tier. Another limitation of its practical use will be smaller sample size from the API with limit positive EV bets every day which slows progress.

## 5.3 Future Improvements

There are several areas where the model could be improved. Firstly, the inclusion of additional features, such as player-level data or more granular match statistics (e.g., possession, shots on target), could potentially provide more insights and improve the model's accuracy. The model could also benefit from exploring more complex machine learning algorithms or hybrid approaches, combining multiple models to increase robustness and predictive power.

Furthermore, implementing Markov Chain Monte Carlo (MCMC) methods could help improve model estimation, particularly when dealing with uncertainty in predictions. MCMC allows for better exploration of the parameter space, providing a more comprehensive understanding of the model’s underlying distributions and uncertainties. This could potentially enhance the model's performance by better capturing the randomness in football match outcomes.

Expanding the model’s scope by adding more leagues would further strengthen its predictive power. Including additional European leagues or even integrating international competitions could provide a richer dataset, potentially leading to more generalisable results across different teams and playing styles.

Additionally, hierarchical modelling could be explored to model data at multiple levels, accounting for both team-specific and league-specific factors. This approach could improve predictions by capturing the interactions between teams and the broader context of the league they play in, offering a more nuanced understanding of match dynamics.

An exciting opportunity lies in adapting the model for other types of sports, such as predicting over/under odds for goals scored in football. This would involve adjusting the model to predict the total number of goals in a match, which can provide valuable insights for betting on total goals over or under specific thresholds. Moreover, the model could be adapted for sports with less available data, such as darts. With careful feature selection and the application of techniques like transfer learning, the model could be trained on smaller datasets and still yield reliable predictions, making it versatile across different sports.

# **6 Conclusion**

This project focused on developing and evaluating a machine learning model for predicting football match outcomes, with the aim of identifying positive expected value for betting. By leveraging historical data from multiple European football leagues, the model was trained to predict the probability of match results, using XGBoost—a powerful and well-established ensemble learning algorithm.

Throughout the process, various techniques were explored, including hyperparameter tuning with GridSearchCV to optimise the model's performance and calibration methods to adjust predicted probabilities for more reliable betting decisions. While the model demonstrated reasonable performance, achieving an accuracy of 62.76% and a ROC AUC score of 0.783, it was clear that the dataset's size and the imbalance between match outcomes posed challenges, particularly in predicting less common events such as draws.

The back testing results confirmed the model’s utility in simulated betting scenarios but also highlighted the difficulty in achieving consistent results with a smaller sample size for real-world applications. Despite these challenges, the incorporation of bankroll management strategies like the Kelly Criterion added value, allowing for more informed and risk-managed betting decisions.

Looking ahead, there are numerous opportunities to enhance the model. By adding more leagues, incorporating richer features such as player-level data, and exploring advanced techniques like Markov Chain Monte Carlo (MCMC) for uncertainty estimation, the model’s predictive power could be further refined. Additionally, adopting hierarchical models could improve the model’s generalisability by accounting for both team-specific and league-specific factors. The model could also be easily adapted to predict other sports outcomes, such as over/under goals in football or darts, making it versatile and applicable to a wide range of domains with smaller datasets.

In conclusion, this project laid the foundation for a data-driven approach to football betting, demonstrating the potential of machine learning models in sports analytics. While there are limitations to address, the results are promising and provide a solid basis for future developments and refinements.