Sports Match Outcome Prediction Modeling

MSc in Business Analytics, Central European University

Shahana Ayobi

June 2023

Table of Contents

1. Introduction

This project is conducted in collaboration with BeDedicated, an emerging sports business intelligence consultancy specializing in football (soccer). BeDedicated aims to become the leading provider of comprehensive analysis, insights, and benchmarking services for football clubs, leagues, federations, and other stakeholders in the industry. By delivering a range of services and B2B products, BeDedicated assists clients in making informed decisions, optimizing their operations, and driving performance on and off the field.

As more football fans show interest in football betting, the need for accurate match outcome predictions becomes crucial for betting providers aiming to stay competitive in the market. According to a report by Forbes, the global sports betting market is projected to reach $155 billion by 2024. By collaborating with BeDedicated, this project seeks to develop precise prediction models that consider potential financial risks associated with skewed odds. These models will enable betting companies to attract more customers, ensure balanced odds, and increase their profits. As football plays a significant role in driving revenue, with the online betting sector in the UK alone generating £446 million, accurate predictions give bookkeepers a competitive advantage, allowing them to provide reliable information about match outcomes.

The primary objective of this project is to build a model that accurately predicts match outcomes for the 20 English Premier League teams across three seasons. By leveraging BeDedicated's expertise and industry insights, the project aims to identify the predictors that have the most significant influence on match outcomes. The predictors considered for this analysis are home and away team expected goals, team total market value, home and away team attacking strength, audience attendance, B365 betting odds, and time of the match.

Every football match typically results in one of the three outcomes: a home team win, a draw, or an away team win. Therefore, the feature selection for this analysis is organized such that, the factors affecting each of these outcomes are considered. Thus, both home and away predictors are included to examine their impact on match results.

Figure 1 illustrates the average number of goals scored by home and away teams in each season. It is apparent that a home team advantage exists among the teams across these three seasons; however, this analysis dives deeper into identifying the predictors that can significantly influence match outcomes. By understanding these factors, they can be used to predict future match results with better accuracy and provide better insights for the bookmakers to make strategic decisions.

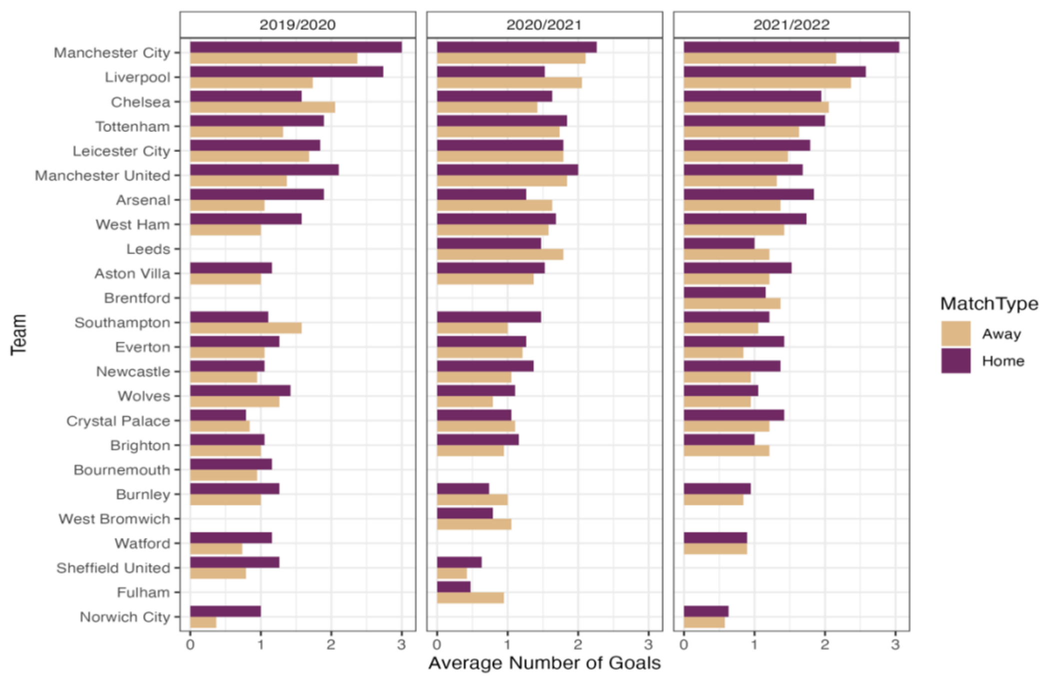

Figure 1: Average Goals Scored by Teams

2. Performance Metrics

In order to evaluate the performance of the machine learning algorithms employed, a performance metrics should be used to assess the prediction models' effectiveness in determining match outcomes. These metric will provide objective measures to quantitatively analyze the models' predictive capabilities and their ability to minimize errors.

One fundamental performance metric employed is accuracy. It is a widely used measure that evaluates the overall correctness of the model's predictions. It quantifies the percentage of correct predictions made across all three outcome classes: home team win, draw, and away team win. A higher accuracy score indicates a greater degree of concordance between the model's predictions and the actual match outcomes.

The accuracy of a model is computed using the following formula:

$$Accuracy = (True\ Positives + True\ Negatives) / (True\ Positives + False\ Positives + True\ Negatives + False\ Negatives)$$

It serves as objective benchmark to measure the models' predictive accuracy and their ability to minimize erroneous predictions. The findings derived from the analysis of this metric contributes valuable insights to the optimization of decision-making processes within the betting industry, ensuring the provision of reliable information to the betting providers.

3. Data Processing and Engineering Pipeline

This section covers the data sources used for this analysis, and elaborates on the data munging techniques to make it ready for modeling.

3.1 Datasets

The data is extracted from three widely used data sources for football match outcome analysis and considers three Premier league seasons; 2019/2020, 2020/2021, and 2021/2022. Each team plays 38 matches, hence each season contains 380 matches, where in total would be 1140 matches for the three seasons. The first source is from Football.co.uk and includes variables such as location of the team, betting odds by different betting providers, and match scores. It contains a total of 106 variables, and 380 rows. This table is cross section match information for each specific season, and the rows represent a single match played between two teams, home and away.

The second dataset comes from FRref and includes information on form of the team and expected goals. It contains a total of 14 variables, and 380 rows for each season. It is extracted from the website by applying specific filters including selection of clubs, the Premier league, and scores and fixtures sequentially.

The third dataset is obtained from Transfermarkt includes team average and total market values for each season in the English Premier League. It consists of 6 variables and 20 rows per season, representing participating teams. To extract the market values for each season, the 'rvest' package in R was employed. The process involved reading the HTML content of the webpage and extracting the specific table that contains the market values for each team. In this case, table number two on the webpage was used and processed by removing empty rows and irrelevant columns. The cleaned data for each season, including team name, squad size, average age, foreign players, average market value, and total market value, was saved in CSV format for further analysis. This step ensures that the data remains accessible for further analysis and avoids the time-consuming process of scraping the data again.

```
# Scrape the market values for each season
url_2019 <- "https://www.transfermarkt.com/premier-league/startseite/wettbewerb/GB1/plus/?saison_id=2019"
  # Read HTML table from URL
  webpage <- read_html(url_2019)
  table <- html_table(webpage, fill = TRUE)[[2]]
  # Filter out empty rows and column headers
  table <- table[-c(1, 1), -c(1, 8)]
  # Rename columns
  colnames(table) <- c("Team", "Squad", "AvgAge", "Foreigners", "AvgMV", "TotalMVH")
  write_csv(table, "market_value_2019.csv")our code here
```

Chunk 1: Scraping the Market Values

3.2 Data Preparation

Dropping Empty Rows

To ensure consistency and equalize the datasets, empty rows were dropped. These rows often served as match week separators and were not relevant to the analysis.

Removing Symbols

In the columns representing average market value (AvgMV) and total market value (TotalMV), symbols and abbreviations were removed to maintain data consistency. This involved removing the Euro symbol and abbreviations for billion (bn) and million (m).

Ensuring measurement Uniformity

To achieve uniformity in measurement units, rows representing market values in billions were converted to millions. This step ensures consistent measurement across the datasets, allowing for accurate analysis and comparison.

Standardizing Team Names

To ensure data consistency and facilitate merging of the datasets obtained from different sources, several discrepancies need to be addressed. One such discrepancy is variations in team names, abbreviations, and team changes/relegations that occurred during the seasons. Resolving these discrepancies is essential for accurate data analysis and interpretation.

The table below illustrates an example of how the team name "Manchester United" was represented differently in the datasets obtained from three different websites:

Stats	XG Scores	Market Values
Man United	Manchester Utd	Manchester United
Man City	Manchester City	Manchester City

Table 1: Differing Team Names Across Datasets

To standardize the names across datasets, a fuzzy matching approach was employed by using Jaro-Winkler method to first calculate the string distances to enable the identification of the closest match. Then each team name was replaced with the predefined standardized names to establish a unified framework. By standardizing the team names, data consistency and accuracy is insured and the datasets are easily merged.

```r
# Create a vector of standard team names
standard_names <- c("Brentford", "Manchester United", "Leicester City", "Burnley", "Chelsea", "Watford", "Everton", "Norwich City",
"Newcastle", "Tottenham", "Liverpool", "Aston Villa", "Manchester City", "Leeds", "Crystal Palace", "Brighton", "Wolves", "Southampton",
"Arsenal", "West Ham", "Sheffield United", "West Bromwich", "Bournemouth", "Fulham")

# Function to fuzzy match a single name with the standard names
fuzzy_match <- function(name, standard_names) {
  distances <- stringdistmatrix(name, standard_names, useNames = FALSE, method = "jw")
  closest_index <- which.min(distances)
  closest_name <- standard_names[closest_index]
  return(closest_name)}

# Standardize team names in each data frame
stats_2019$HomeTeam <- sapply(stats_2019$HomeTeam, fuzzy_match, standard_names = standard_names)
xg_2019$Home <- sapply(xg_2019$Home, fuzzy_match, standard_names = standard_names)
stats_2019$AwayTeam <- sapply(stats_2019$AwayTeam, fuzzy_match, standard_names = standard_names)
xg_2019$Away <- sapply(xg_2019$Away, fuzzy_match, standard_names = standard_names)
market_values_2019$Team <- sapply(market_values_2019$Team, fuzzy_match, standard_names = standard_names)
```

Chunk 2: Team Names Standardization

Merging the Data

To make sure the order of the matches are preserved, a new column, called 'teams', was created to concatenate the names of the home and away team, serving as a unique identifier for the matches. The datasets are then merged using this column for each year individually, and then the merged datasets are combined together using the 'rbind' function to create the final dataset called 'merged_data', which eventually contains 1140 matches as rows, and 51 variables.

4. Feature Engineering

Taking into account the nature of football, the match can result in one of the three possible outcomes: a win, draw, or loss. Thus, the outcome variable is multiclass that takes the value of 1 for a home team win, 0.5 for a draw, and 0 for an away team win. By limiting the outcome variable between 0 and 1, it allows to view the outcome as the likelihood of each event occurring. With this probabilistic interpretation, the relative likelihood of different match outcomes can be assessed, enabling informed predictions based on the calculated probabilities.

As for predictors, In football, the home team typically holds an edge and one of the factors considered is crowd support. The crowd turnout for home team is usually larger which can sometimes overwhelm the opponent side. Figure 2 shows the distribution of attendance for each match outcome category across three seasons. The box plots show a positive relationship between attendance and the likelihood of a home team win. As attendance increases, there is a wider spread of values, indicating greater variability in the effect of crowd support on match outcomes. This indicates that higher attendance levels may contribute to a higher probability of a home team victory.

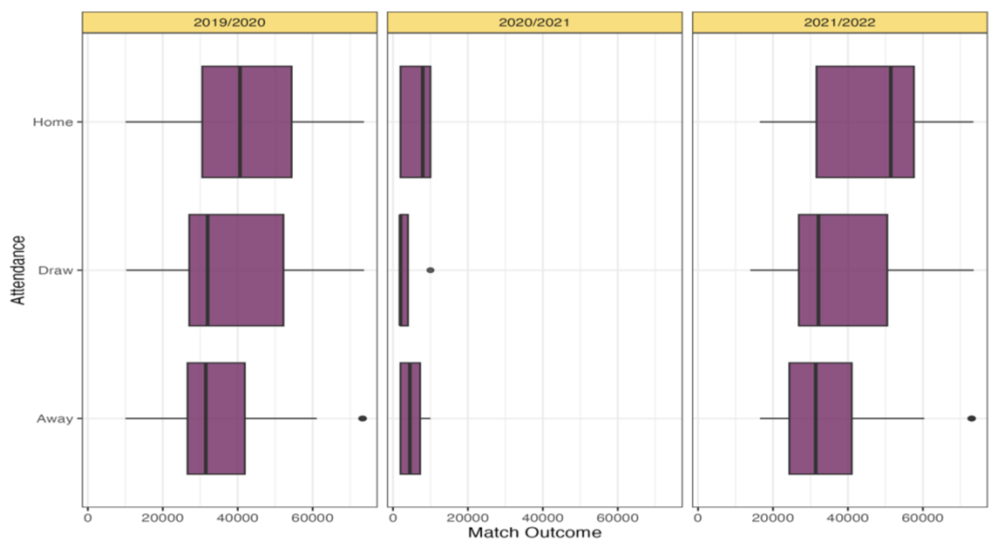

Figure 2: Attendance Distribution for Each Match Outcome

However, adding the attendance directly into the model would drop most of the observations for the 2020/2021 season since there are many missing values for this column in that season because of COVID-19 restrictions. Average value imputation is also not considered a suitable approach in this case since it would not accurately reflect the actual attendance patterns during those matches.

For this analysis, those missing values have been imputed with zeros assuming that no spectators joined those matches. To account for data imputation for those missing values, a flag variable is created for the attendance variable to indicate when the attendance is equal to 0. This flag variable helps distinguish between genuine zero attendances and missing values. By incorporating this flag variable, the analysis takes into account the specific impact of zero attendances and ensures the transparency of the data reporting process.

Table 2 shows the summary statistics of attendees before value imputation for all three seasons, the spectator patterns joining the matches look the same for the seasons 2019/2020 and 2021/2022, while 2020/2021 only match attendees of 10,000 or less in a total of 32 matches out of 380.

Season		Mean	SD	Min	Max	P25	P75	N
2019/2020	Attendance	39314.2	16125.4	10020	73737	27073.5	53324.5	288
2020/2021		5487.2	3399.8	2000	10000	2000	8475.2	32
2021/2022		39543.3	15469.1	13933	73564	27022.7	52966.5	380

Table 2: Summary Statistics for Attendance

Another independent variable for predicting the match outcome is the attack strength of the team. It is derived by dividing each team's average number of goals by the league's average (The Punters Page, 2023). Incorporating the home and away team's attack strength provides an insight into the team's attacking ability, and offers a relative measure of the team's offensive performance (Baio & Blangiardo, 2010).

```
# calculating the home attack strength of the teams
calculate_attack_strength <- function(data) { league_avg_goals <- mean(data$FTHG)
  data <- data %>%
    group_by(HomeTeam) %>%
    mutate(AvgGoalsH = mean(FTHG),
    HAS = AvgGoalsH / league_avg_goals)
    return(data)}

# applying the function to the dataset
stats_2019 <- calculate_attack_strength(stats_2019)
stats_2020 <- calculate_attack_strength(stats_2020)
stats_2021 <- calculate_attack_strength(stats_2021)
```

Chunk 3: Attack Strength Calculation

A key predictor for this analysis is the expected goals (xG) of a team. It is a statistical measure that estimates the likelihood of a shot resulting in a goal depending on parameters such as pass type, shot location, attack type, and others. It is an important predictor to include in the model since it assesses the quality of goal-scoring opportunities provided by a team based on the team's previous performance.

5. Exploratory Data Analysis

Before proceeding with the modeling process, it is essential to assess variable correlations to address concerns related to multicollinearity and ensure the robustness of the model. After investigating the correlations among the odds provided by eight online bookmakers, it was found that these odds exhibited a high degree of correlation. Their high correlation is expected because bookmakers aim to set similar odds to minimize their risk and ensure balanced betting. Therefore, only the odds by B365 are added into the model that would be representative of the rest of 7 bookmakers' odds. This choice captures the broader market trends while minimizing the impact of multicollinearity resulting from correlated odds.

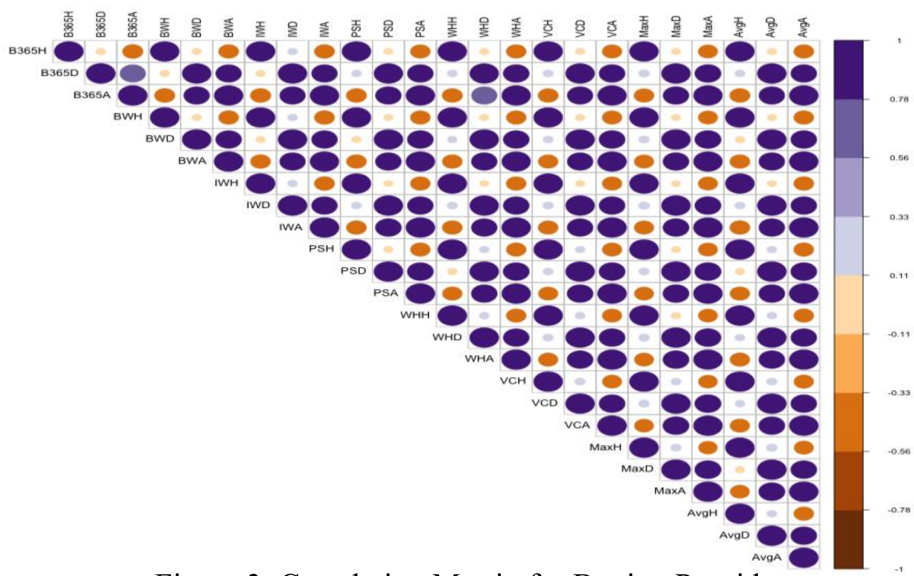

Figure 3: Correlation Matrix for Betting Providers

By focusing on the B365 odds as a representative indicator, the model maintains robustness and interpretability while ensuring consistency in the analysis. This approach avoids potential multicollinearity issues and provides reliable insights into the factors influencing match outcomes.

To gain further insights on match outcomes for each class before moving into modeling, Table 1 shows that on average there is an overall home team advantage with a probability of 42% considering all three seasons. However, for the season 2020/2021, an Away team advantage is seen with 40% win probability.

Season	Home Win Prob	Draw Prob	Away Win Prob
2019/2020	0.4526	0.2421	0.3053
2020/2021	0.3789	0.2184	0.4026
2021/2022	0.4289	0.2316	0.3395
Overall	0.4202	0.2307	0.3491

Table 3: Match Outcome Probability Per Season

6. Modeling

To ensure accurate predictions, the data is split into training and test sets using a 20% to 80% ratio. This allows the model to train on a larger portion of the data and evaluate its performance on unseen test data, preventing overfitting and assessing its generalization capabilities. The training data consists of the first 912 matches, while the test data comprises the remaining 228 matches.

To ensure comprehensive analysis, two types of models are employed: a simple model and a complex model. In the simple model, only the variables of attendance and attendance flag are included as predictors. This allows for a focused examination of their individual impact on match outcomes. In contrast, the complex model incorporates additional predictors and by including these variables, the complex model offers a more comprehensive understanding of the factors influencing match outcomes.

Models	Predictors
M1	attendance + attendance flag
M2	M1 + home expected goals + away expected goals + home total market values + away total market values + home attack strength + away attack strength + Bet365 odds + time category

Table 4: Models Summary

6.1. Multinomial Logistic Regression

Since each match has three probable outcomes, a win, a loss, or a draw, utilizing a Multinomial Logistic regression will make sure to predict these multiple outcome classes simultaneously. This means that the model optimizes its parameters to improve the alignment between predicted probabilities and true value labels, leading to increased classification accuracy. Additionally, multinomial logistic regression provides interpretable coefficients, helping in the interpretation of the model's impact on the outcome classes.

6.2. Random Forest

Random Forest is a type of ensemble learning algorithm that combines multiple decision trees to make predictions. It is relevant to this project because the algorithm is efficient, scalable, and robust against overfitting. By randomly selecting the data and features, random forest creates a collection of trees that would collectively provide more accurate predictions.

For this analysis, the random forest model was fitted to predict the match outcome using the same predictors as the previous logit model. The model was trained with 500 trees, minimum node size of 20, and maximum depth of 5. The model was then predicted on the test set to check its performance on the live data. creating many trees will enable the model to benefit from the combined predictions of those trees resulting in improved predictions. Additionally, the stopping rule is 20 observations in the terminal nodes, this ensures that each tree has a sufficient number of observations to make reliable predictions, while the maximum depth limits the complexity of each tree, promoting better generalization and preventing overfitting on the training data

6.3. GBM Model

The GBM (Gradient Boosting Machine) algorithm is the final model used for match outcome prediction. This algorithm combines predictions from multiple decision trees to generate the final prediction accuracy. GBM learns from previous trees and reduces the usage of unimportant variables in subsequent trees. The tuning parameters for the GBM model include the number of trees set to 500, an interaction depth of 2, a shrinkage value of 0.01, and a minimum number of observations in each terminal node set to 20 which are set to optimize model performance. GBM requires more computational power and time compared to other models due to its iterative nature; however, it is an effective algorithm for capturing complex relationships in the data and improving prediction accuracy.

7. Result Analysis

The table presents the results of the six models used for predicting football match outcomes. The first three models are simple, considering only the attendance as an explanatory variable, while the second set of models includes all other predictors mentioned earlier. Surprisingly, the multinomial logit model with more predictors achieved the highest accuracy of 64% on the test dataset. This means that the model was able to correctly classify the match outcomes 64% of the time.

This result is unexpected for a simple model like multinomial logit, but it highlights the importance of considering the right set of predictors. Sometimes, a simpler model can give better results, as more complex models have a higher risk of overfitting or underfitting the data. When a model becomes overly specific to the training data, it may fail to generalize well to new or unseen data.

These findings align with a study done by Beal, Norman, and Ramchurn in 2019, they showed that the accuracy of bookmakers' predictions vary across different sports. Football (soccer) has a relatively lower accuracy rate of 54%, indicating the challenging nature of predicting match outcomes in this sport. Factors such as the low-scoring nature of football, the occurrence of surprising results, and a higher number of draw outcomes contribute to the complexity of making accurate predictions. Additionally, uncertainties such as team configurations, player health, match location, weather conditions, and team strategies further add to the computational challenges involved in predicting football match outcomes.

Model	Accuracy Train	Accuracy Test
Multinomial Logit	0.4145	0.4386
Random Forest	0.4178	0.4342
GBM	0.4638	0.4956
Multinomial Logit Complex	0.5866	0.6404
Random Forest Complex	0.6502	0.5921
GBM Complex	0.6656	0.5746

Table 5: Model Accuracy Results

7.1. Variable significance and Importance Plots

Since the last three models provided with the best results. It is important to see which variables have the most effect on the probability of each match outcome. Table 2 in the Appendix shows the result of the complex multinomial logit model. It is apparent that xGH and xGA have the strongest influence on the predicted probabilities of home and away team. This means that a one unit increase in expected goals scored by the home team leads to an increase in the log-odds of the home team winning and a decrease in the log-odds of the away team winning taking all other predictors constant. The opposite is true for xGA. Market values of the teams are also significant at 1% significance level, but the magnitude are relatively small compared to other predictors. Home and away attack strength predictors have significant influence on the predicted probability of the football match outcome, the signs are expected, a unit increase in home attack strength will increase the probability of the home team winning, ceteris paribus. Another interesting finding is that, teams playing in the evening, after 5 p.m. have higher probability of winning the match. Reasons can vary from increased crowd support in the evenings to players having higher energy levels in the evenings as compared to the afternoon. The following variable importance plots show the most important variables influencing the probabilities of match outcomes.

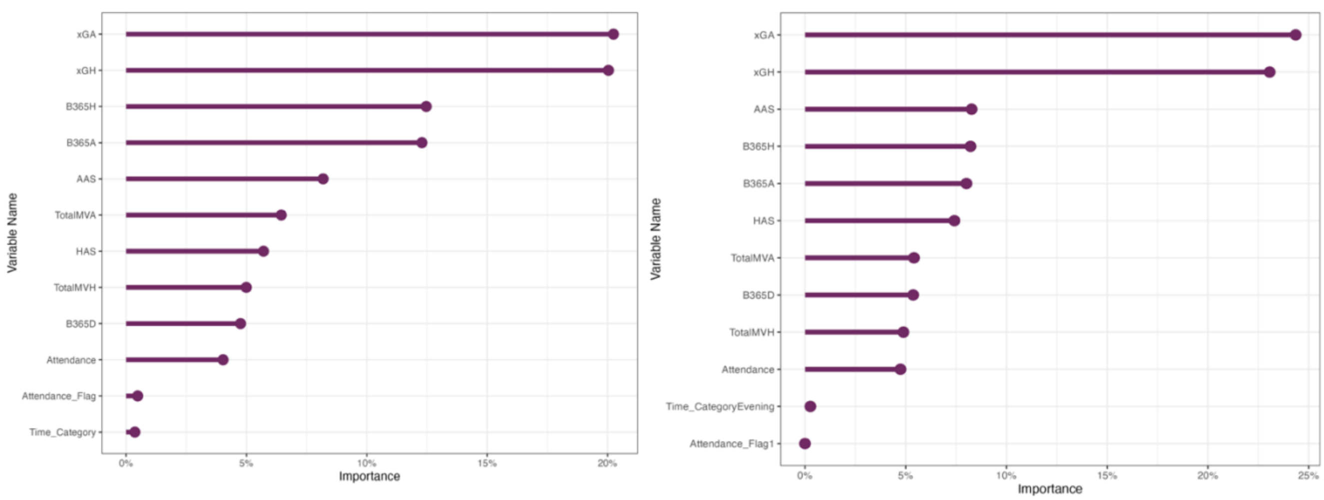

Figure 2: Variable Importance Plots for Random Forest and GBM

The importance plots helps to visualize the importance of each explanatory variable influencing model performance. Figure show the importance plots for the complex random forest and GBM models respectively. The results are relatively consistent for both models with xGA and xGH explaining more than 40% of model performance and followed by attacking strength, betting odds, market values, audience attendance, and time of the match.

7.2. Model Selection and Predicted Odds

The best model chosen for this project is the complex multinomial logit model that yields the highest accuracy, 64% for the test set. This model can generalize well to the unseen data, and can provide interpretable coefficients. It is also simpler and requires less computational power to run. Therefore, the next step is compare the predicted odds received from the best model with the odds provided by the B365 bookmaker. The odds are calculated by first getting the predicted probabilities of the test set for the multinomial logit model, and the dividing 1 by the probabilities to get the betting odds. Table compares a few specific odds for the matches in 2021/2022 season, it is apparent that the model can predict the home odds reasonably comparable for the B365 home betting odds. However, when it comes to the betting odds for the away team, there are some discrepancies seen. One reason could be the fact that the model underestimates the away team win. This is also because the model does not include current market dynamics through which betting providers like B365 adjust their bets. The simple multinomial logit still performs well while predicting home and draw odds despite the fact that B365 usually uses extensive resources and expertise.

Date	HomeTeam	AwayTeam	B365H	B365D	B365A	Home Odds	Draw Odds	Away Odds
11/12/2021	Chelsea	Leeds	1.25	5.75	12	1.39	7.84	6.44
11/12/2021	Liverpool	Aston Villa	1.2	7	13	1.22	6.8	32.16
11/12/2021	Norwich City	Manchester United	7	4.5	1.45	7.39	5.02	1.5
12/12/2021	Burnley	West Ham	3.6	3.6	2	3.73	2.64	2.83
12/12/2021	Leicester City	Newcastle	1.66	4.1	4.75	1.27	5.58	30.58
12/12/2021	Crystal Palace	Everton	2.3	3.2	3.3	1.77	3.37	7.24

Table 6: Betting Odds for Multinomial Logit Model

8. Project Limitations

- Data quality and processing issues: One of the main limitations associated with this project is that the data is gathered from three different sources. This introduces challenges such as inconsistent team names and abbreviations, team changes during the seasons, and team relegations. These variations can make it difficult to clean and preprocess the data consistently, potentially affecting the accuracy and reliability of the models' predictions, and the cleaning process cannot be automated properly.
- Time constraint: The scope of this project is broader than the time limit given for this project, therefore, the bets estimation can be further improved through experiments with machine learning algorithms and adding new relevant predictors.
- Influence of unforeseen circumstances and external factors: Football matches are subject to various unforeseen circumstances and external factors such as weather conditions, referee decisions, injuries, and unexpected events may not be captured by the available data and predictors used in the models. These factors can significantly impact match outcomes and invalidate the model's predictions, leading to imbalanced betting, and potential losses for betting providers.
- Limited generalizability to other leagues: While the models developed in this project focus on the English Premier League, and it achieves an impressive accuracy of 64%, applying the same models to other leagues and clubs might not predict their match outcomes this accurate. Different leagues have unique characteristics, team dynamics, and playing styles, which may require different sets of predictors and models to accurately predict match outcomes.

- Market dynamics and model-bookmaker discrepancies: The odds provided by bookmakers, such as B365, may not perfectly align with the predictions of the developed models. This is because online bookmakers take into account market dynamics, public sentiment, and other factors that affect the odds.

9. Recommendations

- Leverage Natural Language Processing (NLP) or Fuzzy matching techniques for team name standardization to automate the process. This will better address the challenges relating to inconsistent team names and abbreviations and will ensure data consistency and facilitate the merging of data from multiple sources.
- Allocate time for thorough evaluation since sufficient amount of time and expertise in the field is needed to evaluate the models performance. This includes conducting comprehensive testing and validation procedures to assess the models' effectiveness in predicting match outcomes.
- Additional data sources must be explored that capture relevant information to address the impact of unforeseen circumstances and external factors. This includes weather data, injury reports, referee statistics, or other factors that can impact match outcomes. Adding more data to the model from previous seasons will also better capture the patterns of match outcomes.
- Conduct comparative analysis to investigate the generalizability of the models to other leagues, it is suggested to conduct a comparative analysis of different league structures, playing styles, and team dynamics. Identify similarities and differences between the English Premier League and other target leagues. This will make sure to see if the current model or analysis is generalizable to other leagues as well. If not, then it is suggested to

identify league-specific predictors such as team formation, historical performance, or league characteristics. This will ensure that each unique league dynamics are captured.

10. Conclusion

In conclusion, this project aimed to develop precise model predictions for football match outcomes, specifically focusing on the English Premier League. By considering various predictors such as expected goals, market values, attack strength, audience attendance, and betting odds, the models were able to predict match results. The multinomial logistic regression, random forest, and GBM models were evaluated, with the multinomial logistic regression model achieving the highest accuracy of 64% on the test dataset. These models provide valuable insights for betting providers, assisting them in attracting more customers, ensuring balanced odds, and increasing profits.

It is important to note that predicting match outcomes in sports, particularly in football, is a complex task due to various factors and uncertainties involved. Despite the limitations and challenges faced in this project, further research and refinement of the models could potentially lead to improved accuracy and performance.

11. Appendix

The following table shows the results and coefficient estimates of the simple multinomial logit model.

	Dependent variable:	
	Away	Home
	(1)	(2)
Attendance	-0.00001***	0.00001***
	(0.00000)	(0.00000)
Attendance_Flag1	0.025***	0.375***
	(0.000)	(0.000)
Constant	0.522***	0.166***
	(0.000)	(0.000)
Akaike Inf. Crit.	1,952.501	1,952.501
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 1: Model Results for Simple Multinomial Logit Model

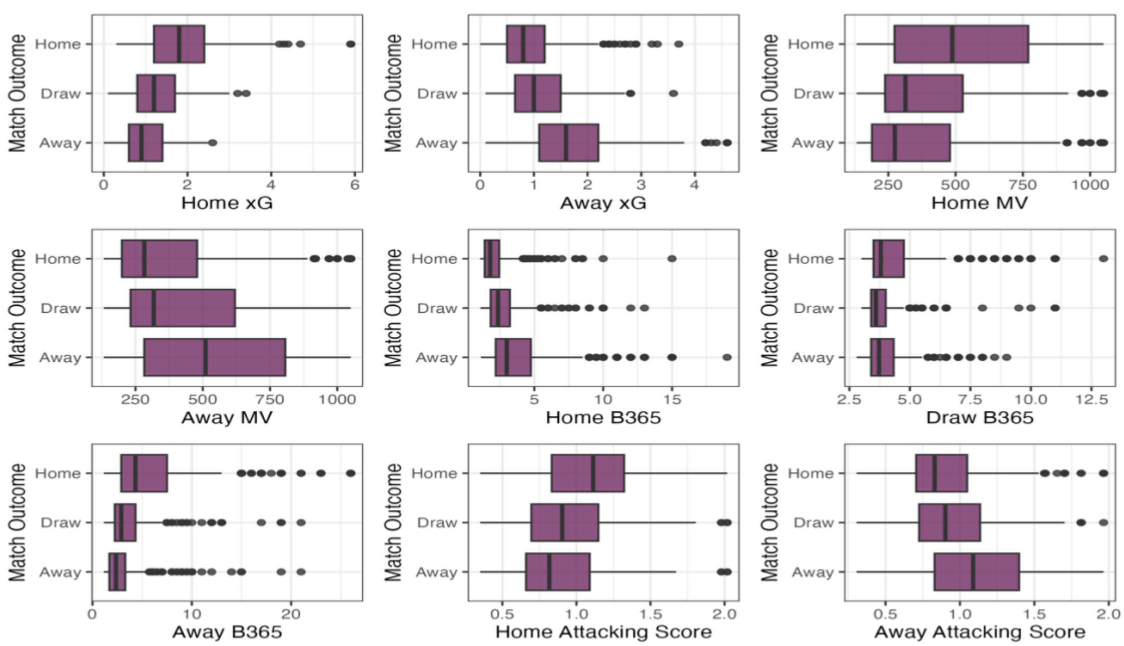

Figure 1: Distribution of Predictor Variables for Each Match outcome

	Dependent variable:	
	Away	Home
	(1)	(2)
xGH	-0.608***	0.657***
	(0.00000)	(0.00000)
xGA	0.998***	-0.163***
	(0.00000)	(0.00000)
TotalMVH	-0.0001	0.001**
	(0.0003)	(0.0003)
TotalMVA	0.0005*	-0.001***
	(0.0003)	(0.0003)
HAS	-0.546***	1.150***
	(0.00000)	(0.00000)
AAS	1.116***	-0.276***
	(0.00000)	(0.00000)
B365H	0.233***	0.176***
	(0.00000)	(0.00000)
B365D	-0.628***	-0.227***
	(0.00000)	(0.00000)
B365A	0.351***	0.078***
	(0.00000)	(0.00000)
Attendance	-0.00001***	-0.00001*
	(0.00000)	(0.00000)
Attendance_Flag	-0.122***	-0.397***
	(0.00000)	(0.00000)
Time_CategoryEvening	0.176***	0.208***
	(0.00000)	(0.00000)
Constant	-0.485***	-0.752***
	(0.00000)	(0.00000)
Akaike Inf. Crit.	1,615.527	1,615.527
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Complex Multinomial Model Result

12. References

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37(2), 253-264.

Beal, R., Norman, T. J., & Ramchurn, S. D. (2019). Artificial intelligence for team sports: a survey. *The Knowledge Engineering Review*, *34*, e28.

Ranjan, A., Kumar, V., Malhotra, D., Jain, R., & Nagrath, P. (2021). Predicting the Result of English Premier League Matches. In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020 (pp. 435-446). Springer Singapore. Retrieved from: https://link.springer.com/chapter/10.1007/978-981-16-0733-2_30

The Punters Page. (2023). Poisson Distribution in Sports Betting – A Step by Step Guide. Retrieved from https: https://www.thepunterspage.com/poisson-distribution-betting/