BEAT THE BOOKIE

A PREPRINT

Group Name: `Group I`
Department of Computer Science
University College London
London, WC1E 6BT

January 2, 2025

1 Introduction

In this report, we present the results of an attempt to create a predictive algorithm for the English Premier(EPL) Matches that are due to happen on February 1st 2025. Four different Multi-class classification Machine Learning models have been tested, with the XGBoost model being the most successful, with the accuracy of 54%. The model is trained to evaluate the strengths of different teams based on prior performance and environmental factors and come up with a prediction of a match outcome, which can be either a Home Win (H), an Away Win (A) or a Draw.

2 Data Transformation & Exploration

2.1 Data Transformation

2.1.1 Initial Dataset Cleaning and Preparation

The provided EPL training dataset underwent an initial cleaning process to ensure data integrity and consistency. Duplicate rows, which were identified using a combination of team names and match dates, were removed. The cleaned dataset was stored as a baseline for subsequent processing.

2.1.2 Multivariate Imputation and Missing Values

Missing data points in the combined dataset, such as possession percentages and set-piece efficiency, were handled carefully. Where possible, different techniques were used to impute missing values. Examples of this can be seen in the code, where Random Forest Regression has been used to impute Team Set Piece Efficiency (ASPE/HSPE) and Team Penalty Efficiency (HPE/APE), along with XGBoost Regressor being used to impute Team Market Value (HTV/ATV) and Posession Averages (HTPos_avg, aTPos_avg). These models used to impute data considered other values of the dataset which were most closely linked to the imputed value as features. This ensured that incomplete records did not adversely impact the model's training process.

2.1.3 Data Integration

External data sources were merged with the cleaned training dataset using unique identifiers, such as team names and match dates and the data was extracted using web scraping. For instance, team market values were matched by team names and seasons, while possession and set-piece data were aligned similarly. The merging process ensured that the final dataset captured both historical match statistics and additional contextual information, such as pre-match metrics and seasonal averages. The main sites used were: 'FBREF Champion League Website' to compute Strictness and Standings from various parameters, 'Transfer Markt' for team market value and 'WhoScored.com' for set piece efficiency and possession.

2.2 Data Exploration and Feature Selection

In order to determine which factors influence the model the most, we have visualised some of the data parameters from the existing, as well as additional data. For example in Figures 1 2 we can see the effect of having a higher Goal Rate, which is the proportions of shots that were successful, on the result of the match.

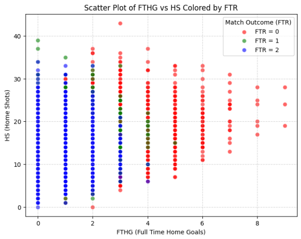


Figure 1: Home Team Goal Rate effect

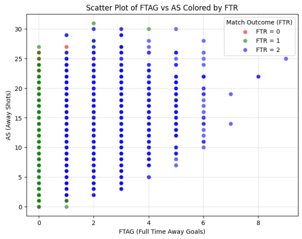


Figure 2: Home Team Goal Rate effect

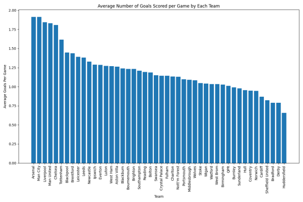


Figure 3: Average number of goals a team scores

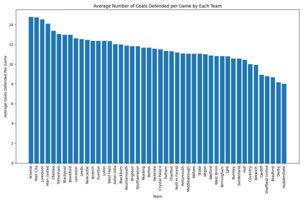


Figure 4: Defence quality of each team

From this, we can conclude that the average rate at which the teams score successfully is likely to affect the result of the game. This allows us to create a new metric which is a derivative of the data points given to us (Goal Rate = $\frac{FTHG}{HS}$ for home team).

In our data exploration, we have made an effort to focus on some of the features that can be traced back to the team. Figure 3 shows the average number of goals each team scores per game, which an intuitive way to assess the strength of the team since it includes the metrics both for each team's winning and losing matches.

In a similar way, we can assess the defence of each team by calculating how many of the opponents shots a team has been able to defend, on average (Figure 4).

Following an insight that the way the team balances the effort throughout the game (in the first or second half of the game), we have found if valuable to assess the percentage of goals a team tends to score in the first half-time, as can be seen in Figure 5. This correspondence can help determine the game outcome if, for example, one team tends to be more focused in the first half-time, and the other - in the second. This information can be used in conjunction with the statistic that 60% of matches half the same outcome as the half-time outcome (Figure 6).

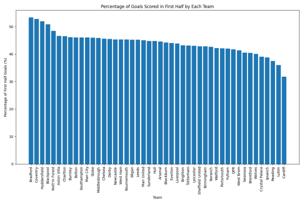


Figure 5: Percentage of goals scored in the first HT

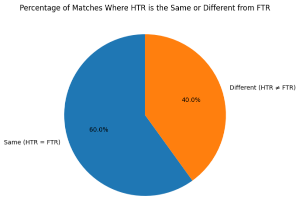


Figure 6: Half-time vs. Full time results

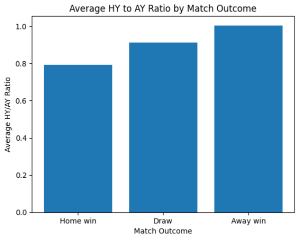


Figure 7: Yellow cards' effect of the game outcome

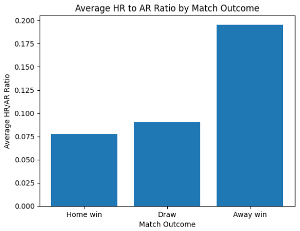


Figure 8: Red cards' effect of the game outcome

Other metrics included in the data that were the number of yellow and red cards that the contestant team received. Through our exploration, we have determined that yellow cards bear almost no effect of the game's outcome(Figure 7), while red cards numbers tend to show a slight tendency for the team with a lesser number to win (Figure 8).

Other useful metrics defining the different complimentary - or contrasting - game practices factored into the research included studying whether higher possession duration - duration of time when then team was in control of the ball, had any effect on their victory probability. This could allow us to estimate what the team's behaviour would be with regards to possession in future games (Figure 12).

Environmental factors certainly have an effect of the player's performance in each game. One of our focuses in adding extra features was to consider environmental factors, especially the level of travel fatigue for the away team and the support from the audience. The graphs 9 and 10, unsurprisingly, show that the higher level of fatigue increases the chances of a home win, as does a higher attendance on the game as most of the audience tend to be the supporters of the home team.

Environmental factors generally tend to bode better for the Home team. However, another factor, which is to do with the schedule of the games, not necessarily does so. The number of games that the team has played before the given match might increase the fatigue, wherefore putting the team with the higher number of matches played in the previous 14 days at disadvantage, as shown in Figure 11

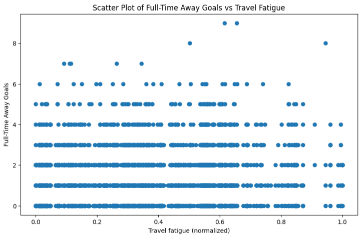


Figure 9: Effect of travel fatigue on performance

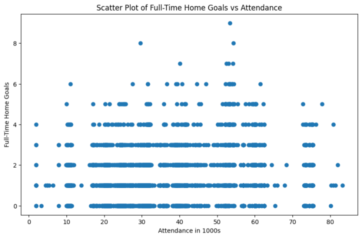


Figure 10: Effect of a higher attendance of the game

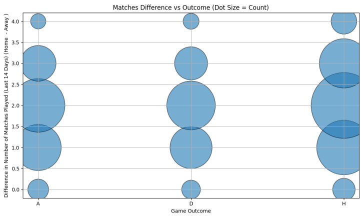


Figure 11: Effect of the number of previos matches

In the history of the English Premier League, it has been accepted as commonplace that different teams have different Monetary Values. Although the Value can be attributed to different factors, including advertisement policies and regional support, it is reasonable to assume that the teams that are valued higher should be generally more successful. This rationale has largely been proven true, as can be seen in Figures 13 and 14.

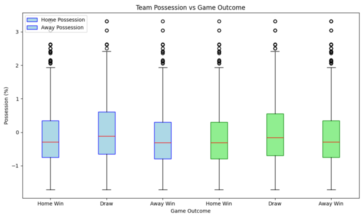


Figure 12: Possession effect on the outcome

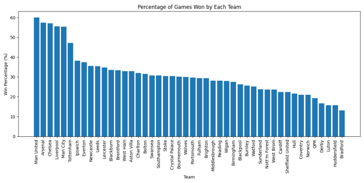


Figure 13: Number of Games won per team

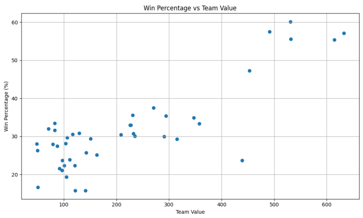


Figure 14: Effect of the team's monetary value

2.3 Final Features

Thus, the list of features that we have incorporated is as follows: Match Density, Referee Strictness, Season, Round, Team points, Team strength, Team defence strength, team form, difference in team form, team win streak, team valuation, average team possession, team set piece efficiency, team penalty efficiency.

The full explanations of all the features can be found in the Appendix.

3 Methodology Overview

The methodology was grounded on the literature review conducted at the beginning that reviewed a variety of models used to approach a problem like this, followed by detailed feature engineering exploring additional data. Training and evaluation techniques were also explored and all decisions detailed below were made after testing different mechanisms and seeing which one performed the best, giving an optimum result.

3.1 Literature Review

There is a number of credible research articles on the subject of predicting the results of football matches using Machine Learning. Although not all of them are tailored towards the English Premier League[5], they share a lot of useful similarities. An article by Fatima Rodrigues and Angelo Pinto [6] shows explores the performance of different Machine Learning approaches and their accuracy rates. The approaches that generated the highest levels of accuracy for the research team were SVM (63.95%), Xgboost(63.95%) and Random Forest(63.95%)[6], which is remarkably closed to the results achieved by a different team led by R.Baboota[8]. However, Deep Neural Network (DNN) considered in article by Ashiqur Rahman [7] but omitted in the Rodrigues/Pinto article, has achieved a relatively close accuracy of 63.3% on the FIFA matches in 2018[7]. In the article by Sergei Anfilets et. al.[10] we can see a similar approach the building the model with an accuracy of 61.14% on test data. It was also observed in a research article that using a K-Means model as part of a broader machine learning approach was beneficial and improved prediction and reliability in predicting Football World Cup Championship results by 55%[11]. Thus, based on prior research, K-Means, SVM, DNN and XGBoost have appeared to be the most prospective ones for development. This conclusion seems natural all the more, since K-Means and SVM models are naturally suited for Classification tasks[18][19], whereas the DNN[16] and XGBoost are able to work with various types of tasks as long as sufficient data is available which is true in the case of the EPL Results prediction.

3.2 Additional Data and Feature Engineering

Based on what was observed through the course and the readings conducted, it was evident that innovative feature engineering and adding additional data would be vital for any model to perform well after data preprocessing.

Before anything, data preprocessing was done, which involved standardizing naming conventions (e.g., reconciling "Manchester United" with "Man United") , converting numerical values into uniform formats, data cleaning and imputation which is described in more detail in section 2. This process ensured consistency when merging external features with the training dataset.

To enrich the dataset and improve model performance, a series of features were engineered using domain knowledge. An important addition was the 14-day match density, which measured the number of matches each team played in the 14 days preceding a given match. This feature, calculated using rolling windows over historical match schedules, aimed to capture the potential impact of player fatigue and match congestion on team performance. These features reflected not just overall team performance but also recent form and a paper explored and explained the advantage that rolling averages had in temporal pattern capture, revealing team dynamics and scoring patterns[12]. External data sources, such as team market values, possession statistics, and set-piece efficiency, were incorporated through web scraping.

In order to look at more innovative features which have an influence on game prediction, we discussed various suggestions as a team. A key feature identified was referee strictness, quantified as a weighted average of red and yellow cards issued by each referee. Specifically, the formula (3×red cards+yellow cards)/4 was used to capture the higher impact of red cards on match dynamics. On a similar vein, team strength, defensive strength, goal-scoring efficiency, and win streaks were also derived from historical data using both aggregate and rolling averages. These metrics were largely based on finding a proportion between goals scored or conceded over the matches played or defining win rates for teams.

A paper [13] exploring team performance metrics in football match prediction emphasised on using comparative metrics as features, making it important to incorporate differences in statistics as features, which was done by including difference in form points between home and away teams. Where form was based on looking at points gained over a rolling window by the home or away team.

3.3 Overview of Model Training and Evaluation

3.3.1 Training

Training a model involved partitioning the final data set into training, test and validation sets, using the training set to understand complex relationships between the range of features and the match outcome. This particular split allowed the models to learn from a significant subset of the overall data while using the validation set for hyperparameter optimisation and early stopping.

3.3.2 Regularisation Methods

To effectively mitigate overfitting and enhance the generalisation capabilities of the predictive models, a multifaceted approach was employed during the training phase. One of the primary strategies involved the implementation of regularisation techniques, such as L2 regularization and dropout layers, which constrain the complexity of the models by penalising large weights and randomly deactivating neurons during training. This prevents the models from becoming overly tailored to the training data, thereby reducing the risk of capturing noise instead of underlying patterns. Additionally, early stopping was utilized as a crucial mechanism to halt the training process once the model's performance on the validation set ceased to improve. This not only conserves computational resources but also safeguards against the models learning spurious correlations that do not generalize well to unseen data. These techniques are well-supported by Srivastava et al. demonstrating the effectiveness of dropout in preventing overfitting in deep neural networks[14].

3.3.3 Hyperparameter Tuning

In order to fine-tune the models for optimal performance, comprehensive hyperparameter optimisation was systematically conducted, with different parameters being optimised differently in the variety of models we explored. This involved looking at a range of hyperparameter values, such as learning rates, tree depths, and regularization coefficients, to identify configurations that balance bias and variance effectively. Also, cross-validation methods, including Repeated Stratified K-Fold, were employed to validate the models across multiple subsets of the data, thereby providing robust estimates of their performance and ensuring consistency. By meticulously tuning hyperparameters and employing cross-validation, the models had greater resilience against overfitting, aligning with best practices outlined by Bergstra and Bengio in their comprehensive survey on hyperparameter optimisation[15]. This framework meant that the models not only performed well on the training data but also generalised effectively to new, unseen match scenarios.

3.4 Alternative Models

Current state of the field of Machine Learning offers a wide variety of different methodologies and strategies of Machine Learning. In order to narrow it down to some of the potential strategies, we have explored a range of publications, as discussed above.

The goal of predicting future match outcomes using minimal data—limited primarily to the competing teams and the referee—led the team to focus on algorithms that can effectively handle sparse or incomplete information. In this context, we selected **XGBoost**, a gradient-boosting method built on decision trees, as our principal model. XGBoost excels at capturing non-linear relationships, managing missing values, and maintaining robust performance on tabular data, making it particularly well-suited for scenarios where comprehensive historical data might be lacking. However, aiming to get a proof of the quality of our choice, we have also developed smaller models using several alternative strategies: SVM, K-Means clustering and a DNN.

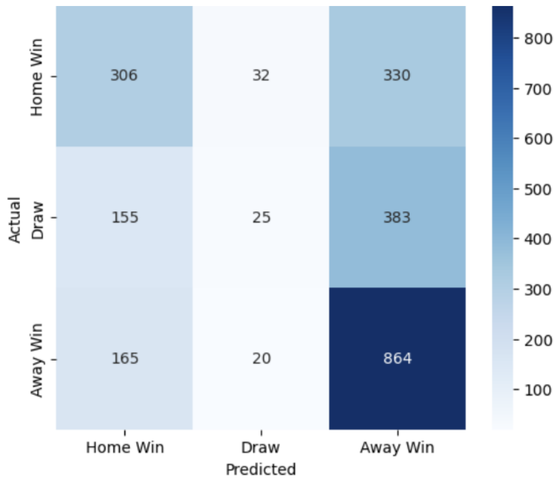


Figure 15: SVM Confusion Matrix

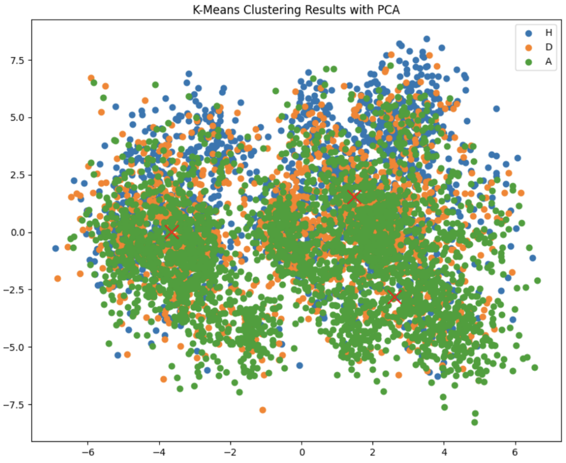


Figure 16: K-Means Clutering

SVM or Support Vector Machines is a class of supervised Machine Learning algorithms that is based on finding the optimal hyperplane to perform classification of points to different classes. Hyperplane, in this case, becomes the decision boundary in the N-dimensional space, where N is the number of features considered.[2][1]
SVMs are innately well-suited for binary classification.[18] The problem of predicting English Premier League match results is multi-class classification problem, working via a K-class Discriminant model[17], the results of this are explored in section 5. As can be seen in the Confusion Matrix in Figure 15, this model deals significantly better with predicting the outcomes of Away or Home win as opposed to the Draw result.[21][20]

K-Means Clustering is an unsupervised Machine Learning algorithm that is very suitable for exploratory data clustering and pattern recognition. This method is considered the simplest of the unsupervised learning algorithms

family, making it a suitable choice to apply unsupervised ML algorithms to this problem. K-means clustering assigns a class to a data point based on the point's distance from the centre of the cluster under consideration.[19] K-means clustering algorithms works cyclically, by first randomly assigning the cluster centres to points and then recalculating the mean's coordinates via categorizing each item to its closest mean.[3][4]
In order to visualise the clusters and centroids, we used Principle Component Analysis, the model's clustering can be seen in Figure 16, which shows that the cluster centres are very close together, making it hard to separate the data, which offers an explanation for the low accuracy score mentioned in section 5. Here, the Hungarian algorithm for optimal cluster-to-label mapping was incorporated to ensure better alignment between predicted clusters and true labels.

DNN : At the same time, we also chose to experiment with a Deep Neural Network (DNN). While XGBoost offers interpretability and strong performance out of the box, a DNN can uncover highly complex patterns when larger or more granular datasets (e.g., detailed player statistics or in-game metrics) become available. By testing a DNN, we aimed to determine whether its capacity for modelling intricate interactions could yield an additional performance boost. Deep Neural Networks are a subset of Neural Networks with several hidden layers. The Neural Network discussed further is a Multilayer Preceptor, which is a Feed-Forward-Network (thus does not involve any loops back to preceding layers) with inputs in the first layer and outputs in the last layer, separated by a set of hidden layers. Since the model is used for Multi-class Classification, it's Evaluation Function can be mathematically described as follows[16] for the output of the k^{th} layer:

$$\tilde{o}_k^{(0)} = \sum_j w_{kj}^{H+1} z_j^{H(i)} + b_k^{H+1}$$

4 Model Training & Validation

4.1 Training

The dataset was partitioned into training, validation, and test sets to facilitate unbiased evaluation and to safeguard against data leakage. An initial split allocated 80% of the data for training and 20% for testing, followed by a further division of the training set into training and validation subsets. This hierarchical splitting enabled the models to learn from a substantial portion of the data while retaining a separate validation set for hyperparameter tuning and early stopping criteria.

4.2 Validation

After training our model, we need to validate it. Validation is a crucial step in machine learning. While a high training accuracy might look promising, it doesn't necessarily indicate that the model performs well on unseen data. Training accuracy only reflects the model's ability to handle data it has already seen, which partially tests its generalization. The real challenge lies in assessing how well the model responds to new, unseen data—this is where validation comes into play. Validation acts as the bridge between training and deployment, helping us understand whether the model is underfitted, overfitted, or affected by data leakage. These insights are essential to ensure the model's reliability in real-world scenarios post-deployment.

Given that several models have been trained, we evaluate their ability to predict the outcomes of EPL football matches by assessing how well they generalize beyond the training data. This involves checking training and validation accuracy on unseen data. The primary metric for evaluating accuracy is the F1-score because it balances precision and recall, which is crucial for our imbalanced dataset. Training and validation losses are also critical indicators; together with accuracy metrics, they reveal whether the model is overfitting, underfitting, or experiencing data leakage. After identifying these issues, we optimize the hyperparameters of each model to maximize predictive performance. Performance after optimization is a key determinant in selecting the best model.

For all models, we used the simplest form of validation: the train-test split. This approach was chosen because we had a moderately sized dataset and this method is computationally efficient while providing a straightforward evaluation of model performance. We split the dataset into 80% for training and 20% for validation.

We first explored hyperparameter optimisation on the DNN. Initially, the DNN achieved an accuracy of 86% on the training set without any fine-tuning. This initial model had a learning rate of 0.001, a dropout rate of 0.3, and no regularization. Compared to the benchmark 53% accuracy of football match prediction models (such as bookmakers' predictions), this result appeared reasonable. However, a closer examination of validation accuracy and loss revealed a different picture. While training accuracy was 86%, validation accuracy was only 47%, indicating a significant disparity. Moreover, the validation loss curve trended upwards, suggesting that the model is overfitting.

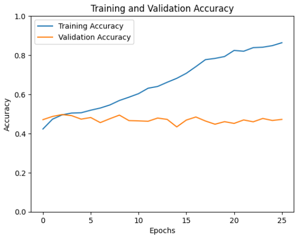


Figure 17: Original DNN accuracy

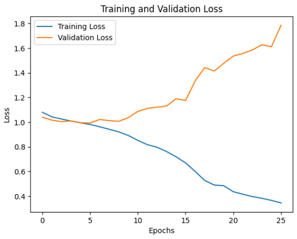


Figure 18: Original DNN losses

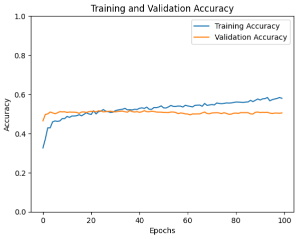


Figure 19: Fine-tuned DNN accuracy

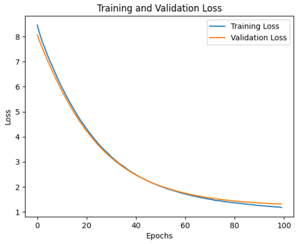


Figure 20: Fine-tuned DNN losses

To address these issues, we adjusted the learning rate to 00005, increased the dropout rate to 0.5, and introduced L2 regularization with a coefficient of 0.01. After fine-tuning, the gap between training and validation accuracy narrowed, within 0.1 of each other. Furthermore, both training and validation loss curves showed a consistent downward trend, reflecting better generalization and a more robust model. These fine-tuning efforts have drastically reduced overfitting seen in the original DNN.

The fine-tuning process of the XGBoost decision tree is largely similar. For this approach, we chose a learning rate of 0.001, a dropout rate of 0.1 as our final hyperparameters. We also used DART (Dropouts meet Multiple Additive Regression Trees) as a booster in XGBoost which is effective for reducing overfitting due to its integration of dropout techniques and also combined early stopping with this in the model.

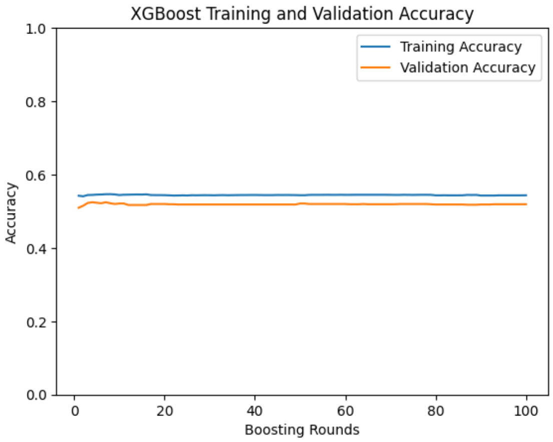


Figure 21: Fine-tuned XGBoost training and validation accuracy

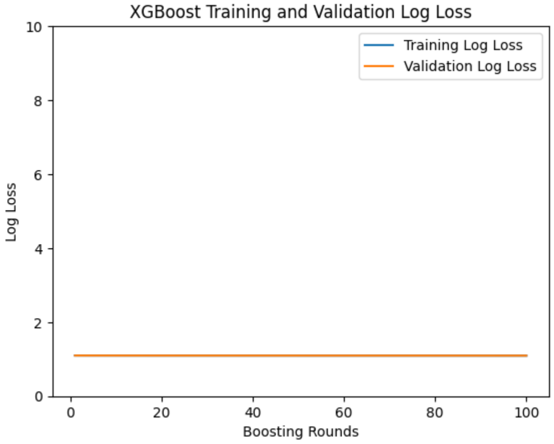


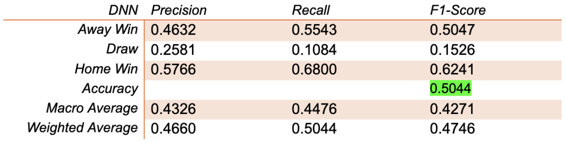
Figure 22: Fine-tuned XGBoost training and validation losses

5 Results

We used classification report to be able to compare and evaluate the models based on a set of varying parameters.

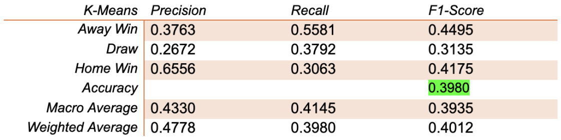
Based on the reports seen above, one can see that the Gradient Descent Boosted Decision Tree (XGBoost) has the highest accuracy of 54%. However, this is at the cost of not being able to predict any draws as seen in table 4. On the other hand, looking at Table 2 for K-Means, a significantly lower accuracy but much better ability to predict draws can be seen. This seems to knead into our hypothesis that draws were harder to predict due to the complexity of matches that end in a draw and indicates further avenues of improvement which is covered in section 7.2 in more detail.

Overall, due to the XGBoost Decision-Tree model not only displaying a higher accuracy but also superior F1-scores for both Home Win and Away Win categories which indicates a balanced performance in precision and recall. Although it can't predict draws, the high recall scores mentioned suggests robust performance in identifying decisive match outcomes which are more common as compared to a draw. Hence, since out of the models evaluated the XGBoost



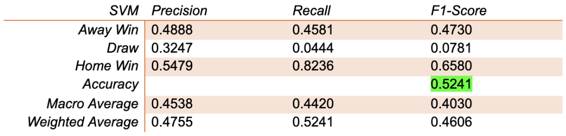
DNN	Precision	Recall	F1-Score
Away Win	0.4632	0.5543	0.5047
Draw	0.2581	0.1084	0.1526
Home Win	0.5766	0.6800	0.6241
Accuracy			0.5044
Macro Average	0.4326	0.4476	0.4271
Weighted Average	0.4660	0.5044	0.4746

Table 1: DNN Report Classification



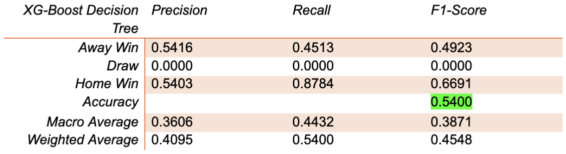
K-Means	Precision	Recall	F1-Score
Away Win	0.3763	0.5581	0.4495
Draw	0.2672	0.3792	0.3135
Home Win	0.6556	0.3063	0.4175
Accuracy			0.3980
Macro Average	0.4330	0.4145	0.3935
Weighted Average	0.4778	0.3980	0.4012

Table 2: K-Means Report Classification



SVM	Precision	Recall	F1-Score
Away Win	0.4888	0.4581	0.4730
Draw	0.3247	0.0444	0.0781
Home Win	0.5479	0.8236	0.6580
Accuracy			0.5241
Macro Average	0.4538	0.4420	0.4030
Weighted Average	0.4755	0.5241	0.4606

Table 3: SVM Report Classification



XG-Boost Decision Tree	Precision	Recall	F1-Score
Away Win	0.5416	0.4513	0.4923
Draw	0.0000	0.0000	0.0000
Home Win	0.5403	0.8784	0.6691
Accuracy			0.5400
Macro Average	0.3606	0.4432	0.3871
Weighted Average	0.4095	0.5400	0.4548

Table 4: XGBoost Report Classification

Decision Tree turned out to be the most reliable model, we have chosen this as the main model using which we have created the final prediction set.

6 Final Predictions on Test Set

To make our predictions for the test set, the 10 test football matches have been added to the training set so that their associated features can be computed. Data from the training set are still required because features such as 14-day match density require data prior to the test matches. Features that are not relevant to the test set but are present in the training set are still added to the test set so that the number of features between the training set and the test set matches up. Unavailable values are substituted with NaN or a boolean false (or zero).

Date	Home Team	Away Team	Prediction
01-Feb-25	AFC Bournemouth	Liverpool	H
01-Feb-25	Arsenal	Man City	H
01-Feb-25	Brentford	Spurs	H
01-Feb-25	Chelsea	West Ham	H
01-Feb-25	Everton	Leicester City	H
01-Feb-25	Ipswich Town	Southampton	H
01-Feb-25	Man Utd	Crystal Palace	H
01-Feb-25	Newcastle	Fulham	H
01-Feb-25	Nottingham Forest	Brighton	A
01-Feb-25	Wolves	Aston Villa	H

7 Conclusion

In the course of the work, 13 new features have been added to the dataset and 4 models trained, the latter being based on K-Means, Support Vector Machine, Deep Neural Network and XGBoost machine learning paradigms. Although all models have performed better than a purely "coin toss" accuracy of 33.33%, XGBoost strategy has proven to be the most successful with the Accuracy of 54% , exceeding the industry standard of 53%. The main drawback of the chosen model is it's inability to predict Draws. Other models considered, although demonstrating a lower overall accuracy, have a higher f1-score in predicting Draws, reaching 31.35% for K-Means.

7.0.1 Future Developments

The model could be further improved by focusing on developments in accuracy of Draw prediction. One potential way to do it could be using Voting/Stacking Ensemble Methods combining the models we tested (DNN, SVM + XGB Decision Tree). This way, the benefits of different models can be combined to improve overall performance.[22]

References

[1] S. Suthaharan, "Support vector machine," in *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, Springer, 2016, pp. 207–235.

[2] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[3] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.

[4] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[5] J. Stübinger, B. Mangold, and J. Knoll, "Machine Learning in Football Betting: Prediction of Match Results Based on Player Characteristics," *Applied Sciences*, vol. 10, Dec. 2019. doi: 10.3390/app10010046.

[6] F. Rodrigues and Â. Pinto, "Prediction of football match results with Machine Learning," *Procedia Computer Science*, vol. 204, pp. 463-470, 2022. doi: 10.1016/j.procs.2022.08.057.

[7] M. A. Rahman, "A deep learning framework for football match prediction," *SN Applied Sciences*, vol. 2, Jan. 2020. doi: 10.1007/s42452-019-1821-5.

[8] R. Baboota and H. Kaur, "Predictive analysis and modelling football results using machine learning approach for English Premier League," *International Journal of Forecasting*, vol. 35, pp. 741-755, Apr. 2019. doi: 10.1016/j.ijforecast.2018.01.003.

[9] A. Joseph, N. E. Fenton, and M. Neil, "Predicting football results using Bayesian nets and other machine learning techniques," *Knowledge-Based Systems*, vol. 19, pp. 544-553, Nov. 2006. doi: 10.1016/j.knosys.2006.04.011.

[10] S. Anfilets, S. Bezobrazov, V. Golovko, A. Sachenko, M. Komar, R. Dolny, V. Kasyanik, P. Bykovyy, E. Mikhno, and O. Osolinskyi, "Deep multilayer neural network for predicting the winner of football matches," *International Journal of Computing*, vol. 19, pp. 70-77, Mar. 2020. doi: 10.31891/1727-6209/2020/19/1-70-77.

[11] Bai, Yanyang, Zhang, Xuesheng, Prediction Model of Football World Cup Championship Based on Machine Learning and Mobile Algorithm, Mobile Information Systems, 2021, 1875060, 11 pages, 2021. https://doi.org/10.1155/2021/1875060

[12] Yeung CCK, Bunker R, Fujii K. A framework of interpretable match results prediction in football with FIFA ratings and team formation. PLoS One. 2023 Apr 13;18(4):e0284318. doi: 10.1371/journal.pone.0284318. PMID: 37053253; PMCID: PMC10101499.

[13] Adnan, N.A., Al Hakim Mohd Asri, L., Mustapha, A., Razali, M.N. (2024). The Football Matches Outcome Prediction for English Premier League (EPL): A Comparative Analysis of Multi-class Models. In: Ghazali, R., Nawi, N.M., Deris, M.M., Abawajy, J.H., Arbaiy, N. (eds) Recent Advances on Soft Computing and Data Mining. SCDM 2024. Lecture Notes in Networks and Systems, vol 1078. Springer, Cham. https:$//doi.org/10.1007/978 - 3 - 031 - 66965 - 1_40$.

[14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov.Dropout: a simple way to prevent neural networks from overfitting.

[15] James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, null (3/1/2012), 281–305.

[16] D. Hosseini, “Machine Learning: Neural Networks,” Moodle, 2024.

[17] D. Hosseini, “Machine Learning: The Non-Linear SVM,” Moodle, 2024.

[18] D. Hosseini, “Machine Learning: Discriminant Classification & the Linear SVM,” Moodle, 2024.

[19] D. Hosseini, “Machine Learning: Clustering,” Moodle, 2024.

[20] M. McGregor, “SVM Machine Learning Tutorial – What is the Support Vector Machine Algorithm, Explained with Code Examples,” freeCodeCamp.org, Jul. 01, 2020. https://www.freecodecamp.org/news/svm-machine-learning-tutorial-what-is-the-support-vector-machine-algorithm-explained-with-code-examples/

[21] A. Sasidharan, “Support Vector Machine Algorithm,” GeeksforGeeks, Jan. 20, 2021. https://www.geeksforgeeks.org/support-vector-machine-algorithm/

[22] O. Sagi and L. Rokach, “Ensemble learning: A survey,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 4, Feb. 2018, $doi : https : //doi.org/10.1002/widm.1249$.

8 Appendix

Feature	Category	Definition	Calculation	Purpose
Referee Strictness	Match Context	Measures referee behavior in issuing penalties (yellow/red cards).	(3 × red cards+yellow cards)/total matches officiated	Accounts for potential referee influence on match outcomes.
14-Day Match Density	Team Context	Number of matches played by a team in the last 14 days across competitions.	Rolling count of matches for the home/away team over the previous 14 days.	Reflects player fatigue and match congestion effects on performance.
HTPS / ATPS	Team Form	Rolling points sum for home/away team up to the previous round.	W=3 points,D=1 point,L=0 points ; rolling sum for last matches.	Captures team performance trends over recent matches.
HTS (Home Team Strength)	Team Strength	Average points earned by the home team across all previous matches.	total points earned by HT/total matches played by HT	Reflects the overall strength and consistency of the home team.
ATS (Away Team Strength)	Team Strength	Average points earned by the away team across all previous matches.	total points earned by AT/total matches played by AT	Reflects the overall strength and consistency of the away team.
HGSR / AGSR	Team Strength	Average goals scored by home/away team across all previous matches.	total goals scored by HT or AT/total matches played by HT or AT	Measures team scoring efficiency over all past matches.
Home_DS / Away_DS	Team Strength	Average goals conceded by home/away team across all previous matches.	total goals conceded by HT or AT/total matches played by HT or AT	Reflects team defensive strength over all past matches.
Home_Form_Points / Away_Form_Points	Team Form	Sum of points from the last 10 matches for home/away team.	Rolling sum for last 10 matches: W=3, D=1, L=0.	Captures recent performance trends for the team.
Home_Goal_Diff_Form / Away_Goal_Diff_Form	Team Form	Net goal difference in the last 10 matches for home/away team.	Rolling sum of Goals Scored−Goals Conceded over the last 10 matches.	Reflects recent offensive and defensive balance for the team.
Home_Win_Streak / Away_Win_Streak	Team Form	Number of consecutive wins for home/away team in the last 10 matches.	Count of consecutive matches with 3 points earned.	Indicates team momentum based on recent consecutive wins.
Home_H2H_Win_Rate / Away_H2H_Win_Rate	Match Context	Percentage of past matches won by the home/away team against their opponent.	Head-to-head wins/total head-to-head matches.	Accounts for historical performance in specific team matchups.
HTV / ATV	Team Context	Team market value normalized over all years.	Normalized team market value (up to 2010).	Reflects team financial resources and player quality.
HTPos_Avg / ATPos_Avg	Team Context	Average possession percentage for home/away team.	Possession data normalized over seasons.	Reflects ball control and dominance in matches.
HSPE / HPE / ASPE / APE	Team Context	Set-piece and penalty efficiency percentages for home/away team.	Normalized values up to the 2009/2010 season.	Accounts for team proficiency in scoring through set pieces and penalties.