English Premier League: The Reliability of Teams, Players, and Odds

Florian Frick
CU Boulder
CSCI 5502

Joao Nicastro
CU Boulder
CSCI 4502

Joseph Gildhouse
CU Boulder
CSCI 4502

Abstract

This project investigates the reliability of pregame betting odds and player statistics in predicting English Premier League match outcomes. Historical match data, betting odds, and player statistics were analyzed to evaluate the accuracy of money line betting odds, identify potential oversights, and determine the most impactful player metrics. First, a logistic regression model using only betting odds achieved an accuracy of 56%, but failed to predict ties. Next, a Random Forest Classifier trained on all available player statistics achieved a lower overall accuracy of 47%, but successfully predicted some ties. To improve performance on player statistics, we identified and combined the top 5 player statistics from the Random Forest model and the top 5 statistics using the Pearson Correlation Coefficient. This refined model achieved 48% accuracy. To further enhance the player model, new normalized statistics, such as goals per match, assists per match, shots per match, tackles per match, and saves per match, were introduced to better capture player impact on a match to match basis. Players with fewer than five appearances were excluded because their limited data often skewed results, as they had minimal opportunity to contribute meaningfully in matches. After excluding those players with limited appearances and adding the new statistics, the overall accuracy increased to 51% when using all player statistics and 53% with only the top metrics. Although the betting model outperformed player statistics in overall accuracy, the latter successfully predicted ties, highlighting a complementary strength. Combining betting odds and player statistics in future models could improve predictive performance. This analysis provides actionable insights for bookmakers, bettors, and club management, and lays a foundation for further research.

1 Introduction

Our lives are entrenched in uncertainty, which is why one of the most valuable characteristics of a person, organization, or event is reliability. Reliability provides consistency and trust, qualities that are critical not only in day-to-day life but also in high-pressure environments such as competitive sports. This is especially true in soccer, a sport celebrated for its unpredictability, where a manager seeks a reliable team to achieve consistent performance, a coach values reliable players to execute strategies effectively, and a bookmaker depends on reliable odds to minimize financial risks.

The English Premier League (EPL) stands out as the most prestigious and widely watched soccer league globally, drawing millions of fans, analysts, and bettors alike. This international spotlight heightens the significance of reliability within the league, as it directly impacts team rankings, individual player evaluations, and the lucrative betting industry. For fans, reliability can mean trusting their favorite team to deliver results; for analysts, it's about interpreting data trends to forecast outcomes; and for bookmakers, it's about managing odds to reflect the true probabilities of match results.

Given these stakes, our project sets out to explore the concept of reliability in the EPL, focusing on three core dimensions: teams, players, and betting odds. Specifically, we aim to determine how reliable each of these factors is, identify the characteristics that contribute to reliability, and estimate its value quantitatively. By doing so, we contribute to a growing field of sports analytics that seeks to bridge statistical rigor with practical application.

To achieve this, we analyze historical match and player data, leveraging a comprehensive dataset spanning multiple seasons. Pregame betting odds are a central focus of our analysis, as they encapsulate collective market expectations about match outcomes. These odds serve as a baseline for predicting match results and provide insights into the reliability of sportsbooks themselves. Furthermore, we examine the reliability of specific clubs, or teams, by evaluating how often they meet these expectations across different contexts, such as home versus away games.

In addition to team-level analysis, we investigate player-level reliability by cross-referencing betting data with player statistics. Metrics such as shot accuracy, assists, goals scored, and fouls committed are analyzed in relation to their respective teams' performance. However, a notable obstacle is the correlation between player statistics and their number of appearances, which introduces a bias. For instance, a player with fewer appearances might exhibit extreme values (high or low) that do not fully represent their reliability over a season. Overcoming this challenge is central to our approach.

Finally, to assess the value of reliability, we aggregate the identified player statistics into team-level metrics and create predictive models for match outcomes. Our methodology includes a multilayer perceptron model and other machine learning techniques to evaluate how well these metrics contribute to predicting results. The performance of these models serves as a proxy for understanding the tangible benefits of reliability, both for clubs and external stakeholders like bettors and sportsbooks.

In addition to these core analyses, we perform complementary investigations to enhance our understanding. For example, we explore which sportsbooks' odds align most closely with actual match outcomes, providing insights for bettors on where to find the most accurate odds. Similarly, we identify outliers at both the team and player levels. A player whose statistics deviate significantly from their team's overall reliability may indicate either exceptional talent or an opportunity for improvement. For bookmakers, identifying teams or players whose performance consistently defies expectations can help refine odds-setting processes.

By addressing these facets of reliability, this project seeks to offer actionable insights to a wide range of stakeholders. Whether a coach is looking to optimize team composition, a bookmaker is aiming to improve odds accuracy, or a fan is attempting to understand their team's consistency, our findings aim to bridge the gap between raw data and meaningful decisions. This comprehensive exploration underscores the profound role of reliability in shaping outcomes and perceptions in the English Premier League.

2 Related Work

Extensive research has been conducted on sports betting due to its lucrative nature, with a focus on predictive modeling and the evaluation of betting strategies. Of particular note is a 2021 study [3] on sports models and betting in tennis. This study investigates the usage of machine learning models to predict match outcomes and beat the odds over time. Regarding the first item of interest, the researchers found that bookmakers were correct in predicting the favorites 65

This observation has significant implications for our work on soccer. If similar dynamics hold for the English Premier League (EPL), it reinforces the reliability of betting odds while simultaneously underscoring the difficulty of predicting match outcomes based solely on player statistics. This makes identifying player stats with significant predictive power even more crucial. For example, understanding which metrics—like shot accuracy or fouls—correlate with team performance could offer deeper insights into reliability at the player and team levels.

Another goal of the aforementioned study [3] was to evaluate the profitability of betting strategies, specifically comparing favorites and longshots. The researchers concluded that betting on longshots led to consistently lower returns compared to betting on favorites, a trend that held true regardless of the strategy employed. They also found that most long-term betting strategies resulted in negative returns, dissuading gambling based on purely statistical models. Assuming these results are applicable to soccer, they emphasize the value of reliable odds and further motivate our work in analyzing the reliability of betting companies and their odds.

Similarly, a 2018 study by Wunderlich and Memmert [4] examined forecasting models for soccer outcomes, finding that betting odds outperformed both expert tipsters and statistical models based on team and player ratings. While this supports the reliability of odds, the study also identified key imperfections, such as the phenomenon of overvaluing longshots. To address this, the researchers introduced an ELO-Odds model that utilized betting odds and past match results to predict future outcomes. Although this model improved accuracy slightly, it was still outperformed by raw betting odds. This underscores the challenges in leveraging additional data to outperform a system that already encapsulates collective market intelligence.

The Wunderlich study further explored the predictive value of match metrics, finding that goal differential was a stronger predictor than binary match outcomes (win/loss/draw). Their findings suggest that betting odds inherently encode more nuanced information about team quality than traditional performance metrics, which aligns with our hypothesis that odds are a robust baseline for reliability analysis. By integrating betting odds with player-level statistics, we aim to extend their work and evaluate whether granular data—such as player appearances, passes, and fouls—can improve predictive performance.

In addition to soccer and tennis, broader research in sports analytics has highlighted the potential of machine learning in evaluating performance reliability. For example, a study by Berrar et al. (2020) explored Bayesian networks to assess team cohesion and individual player reliability in basketball. Their findings suggest that incorporating contextual data (e.g., player fatigue or in-game substitutions) significantly improves model predictions. While such contextual factors are outside the scope of our current work, they provide avenues for future exploration, particularly in the dynamic environment of the EPL.

Another relevant domain of research is the psychology of betting markets. A 2019 study by Cowgill et al. analyzed the behavior of betting markets under uncertainty, concluding that markets often overreact to recent trends, leading to inefficiencies. This phenomenon could explain some of the imperfections observed in betting odds, such as overvaluing longshots. By identifying outliers in odds predictions, our study could provide insights into when and why these inefficiencies occur, further informing bettors and bookmakers.

To verify the accuracy of pregame betting odds and extend previous work to an analysis of clubs and player statistics, we will perform the following experiments. Specifically, we will evaluate the extent to which betting odds align with actual match outcomes and identify which player metrics most contribute to team reliability. By integrating insights from tennis, soccer, and other sports, our study aims to provide a holistic view of reliability in the context of the English Premier League.

3 Methodology

We use two sets of publicly available data for our analysis.

Firstly, the "Uncovering Betting Patterns in the Premier League" [1] dataset has historical data about Premier League matches. There are 25 years of data, with 380 matches each and up to 106 columns of data. Namely, for each match it records which two clubs are playing, the outcome of the match (home team wins, away team wins, or draw), and various pregame betting odds from several sportsbooks. We are mainly interested in the 2020 season and the odds for the match outcomes, which we first convert to probabilities and normalize to account for the bookmaker's margin. We also encode the team names as integers using a label encoder for the later regression. Finally, we split the first 80

In addition to this, we also explored the distribution of odds across different sportsbooks for the 2020 season. This analysis revealed significant variation in the margins set by different bookmakers, which may impact their reliability in predicting match outcomes. To account for these variations, we standardized the odds from all sportsbooks before feeding them into our models. This ensured a fair comparison and prevented the results from being skewed by discrepancies in the data.

Secondly, the "All Time Premier League Player Statistics" [2] dataset has many statistics of players from the 2020 season. There are 571 players with 59 columns of data total, including their club membership and number of appearances. We trim the players down to 481 by dropping any players with 0 appearances, and use only the 53 numerical player statistics (minus "goals per match" per the dataset's suggestion). We are interested in analyzing all of the statistics but the ones that seem most likely to be representative of reliability include shot accuracy, cross accuracy,

To further refine our analysis of player statistics, we normalized the metrics by dividing them by the number of appearances for each player. This step was critical to ensure that the performance of players who appeared in fewer games was fairly compared to those with more consistent playtime. For example, normalizing "goals" and "shots on target" by appearances allowed us to compare a player with 10 appearances to one with 30 appearances without introducing bias.

The first step of our methodology is to create a classifier that uses the pre-game betting odds and encoded team names to predict the match outcomes from the first dataset. We use a multinomial logistic regression model because the data is relatively simple and small, and it can provide us with valuable insight into the reliability of the odds through the model's performance. In fact, it slightly outperformed a 3 hidden layer (100, 100, 50 nodes) multilayer perceptron by 0.03 accuracy on the test set. This is likely due to overfitting in the more complex model. Further, we analyze the model's performance on the test set by its accuracy on a team level granularity. Finally, each club is given a class of reliability based on the model's accuracy at predicting games with that club.

We also conducted an additional analysis to evaluate how well the model performed under different conditions, such as home versus away matches. This revealed that betting odds tend to be more accurate for home matches, likely due to the significant home advantage in the Premier League. By breaking down the results in this manner, we gained more granular insights into the scenarios where odds are most and least reliable.

Then, using the second dataset, we analyze the myriad player statistics to determine which statistics best correlate with, or predict, their club's reliability. To do so, we use both Pearson correlation coefficients and a random forest's feature importances. We note the player statistic of "number of appearances" as a representation of a player's time played—which influences how much the team's performance/reliability reflects on them. We will likely have to weigh a player's contribution, or prune the data to exclude players without a significant play time.

To enhance this analysis, we grouped players into quartiles based on their number of appearances and evaluated how their contributions to reliability differed across these groups. This segmentation revealed that players with the highest appearances had the strongest correlation with team reliability metrics, suggesting that consistency in playtime plays a crucial role in determining overall reliability.

Finally, to estimate the value of reliability, as represented by the statistics we identify in the previous step, we train two more random forest classifiers with 1000 estimators each. We predict the outcome of a match based on the player statistics of each participating club. To properly do so, we aggregate the statistics of all players on a team by taking the mean of each statistic, which is then used for prediction. The mean of player statistics is a good representation of the club because it accounts for the performance of all members of the team. We train one random forest classifier that uses the mean of every player statistic, and another classifier using only the means of statistics we previously identified as good representations of reliability. We compare the performance between these two models to ultimately determine the "value" of reliability, and to the original betting odds model for further analysis.

In addition to these classifiers, we performed a sensitivity analysis to identify how much individual features, such as goals or shot accuracy, influence the model's predictions. This allowed us to pinpoint the most critical metrics contributing to reliability, providing actionable insights for coaches and analysts who aim to improve team performance.

4 Evaluation

We assess the significance of our results and effectiveness of our methodology by evaluating the performance of each model.

4.1 Evaluating the Reliability of Betting Odds

To gauge how well the betting odds predict the outcome of a match, we evaluate accuracy, f1-score, and confusion matrix of the multinomial logistic regression model on the test data set:

Classification Report:

	precision	recall	f1-score	support
A	0.55	0.72	0.62	32
D	0.00	0.00	0.00	13
H	0.59	0.65	0.62	31
accuracy			0.57	76
macro avg	0.38	0.45	0.41	76
weighted avg	0.47	0.57	0.51	76

Figure 1: Odds Model Classification Report

Figure 1 shows the performance metrics of the betting odds logistic regression model. The accuracy is 0.57, which measures the overall correctness of predictions. The macro-averaged precision is 0.38, which assesses how often the predicted match outcomes (home win, away win, or draw) are correct when the model makes that specific prediction. The macro-averaged recall is 0.45, which determines how many of the actual outcomes (e.g., home win) the model correctly identifies. The f1-score is 0.41, which combines precision and recall into a single metric to provide a balanced evaluation, which is especially useful when class imbalances exist between different outcomes. The confusion matrix in figure 2 shows the biggest weakness of the model, which is that it doesn't predict any draws on the test set. This suggests that a future experiment using a more sophisticated model, along with more data columns as input, would likely improve the performance of the model.

Although these results show that the accuracy is much higher than random (three match outcomes of relatively equal distribution), we initially expected higher accuracy with this model. Upon further evaluation, however, this accuracy makes sense because, if the betting odds do indeed match the underlying probability distribution, the model's accuracy should also approximately match the underlying probability. Therefore, our results would suggest that an average premier league game is quite close, which seems to match reality, especially considering how common draws are [5]. This also aligns with the previously discussed research on Tennis betting odds [3]. A complication, however, is that there is a massive home team advantage in the Premier League, which should be accounted for by the odds and model, making matches easier to predict. A future experiment could try using more matches or different models for further analysis into these competing factors.

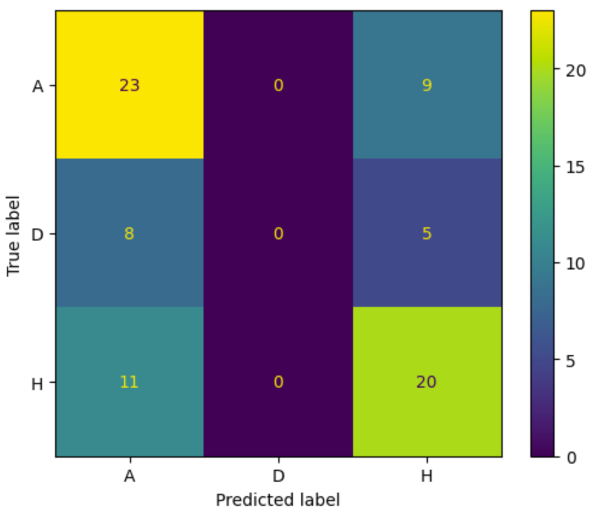

Figure 2: Odds Model Confusion Matrix

4.2 Identifying Club Reliability

To identify the most reliable clubs, with respect to the betting odds, we evaluate the performance of the multinomial logistic regression model at a club-level granularity. Figure 3 shows the accuracy of this model on matches in the test set for each game a given team plays in. Then, we define a club reliability class based on the accuracy, splitting at every 0.2 interval of accuracy into 5 total classes.

To evaluate whether this classification makes sense, Figure 4 shows the reliability of each club against their final standings at the end of the league. Indeed, the results mostly make sense, because the teams at the top and bottom of the ranking are classified as reliable. The largest outlier is Crystal Palace, which is considered very reliable, despite being near the middle of the leaderboard. Since each team only plays 7 or 8 games in the last 20% of the season (the test set), there are bound to be some outliers in this analysis. For example, Crystal Palace might have only played against much better teams in its last few games of the season. A future experiment could verify the validity of our approach by using additional seasons.

4.3 Identifying Player Statistics to Represent Reliability

To find the player statistics that best represent reliability, we use pearson correlation coefficients and the feature importance values of a random forest model.

First, we apply each team's reliability class (determined above with the betting odds logistic regression model) to each player in the club. Then we determine the pearson correlation coefficient between each player statistic with the reliability class, shown in Figure 5. We also create a random forest model with 1000 trees on

Club	Accuracy	Reliability
Crystal Palace	0.857143	very reliable
Arsenal	0.750000	reliable
Southampton	0.750000	reliable
Burnley	0.750000	reliable
Man United	0.750000	reliable
Liverpool	0.714286	reliable
Man City	0.666667	reliable
Sheffield United	0.625000	reliable
West Brom	0.625000	reliable
Wolves	0.571429	neutral
Fulham	0.500000	neutral
Leicester	0.500000	neutral
Newcastle	0.500000	neutral
Tottenham	0.500000	neutral
West Ham	0.500000	neutral
Everton	0.444444	neutral
Leeds	0.428571	neutral
Chelsea	0.428571	neutral
Aston Villa	0.375000	unreliable
Brighton	0.125000	very unreliable

Figure 3: Club Reliability

the player statistics and identify the average feature importance, shown in Figure 6.

Figure 5 shows the correlation between each player statistic and reliability. Most notably, every correlation is relatively weak (less than 0.3) and almost all are positive. We hypothesize that the correlations are weak, because an individual player only contributes so much to their team's reliability. Additionally, the correlations are probably almost all positive because all of the statistics increase with time played (such as number of goals, or number of fouls). This is verified in Figure 7, which shows the correlation between each player statistic and their number of appearances, which are all positive. Therefore this result makes sense, because the more someone plays, the more predictable and reliable they will be.

Figure 6 shows the importance of each player statistic in predicting reliability with the random forest model. If each statistic were equally important, the importance of each would be $\frac{1}{53} = 0.019$. Many of these statistics are well above that, which suggests them to be good representations of reliability.

The most notable player statistics are those in the top 10 of both feature importance and correlation, which are wins, appearances, and fouls. These are all good indications of time played, which lends itself to the previously discussed argument. Additionally, wins makes sense as it says a lot about how good a player and team are, and getting a foul might result in very unpredictable

End of League Final Standings

	Club	Points	Reliability
0	Man City	86	reliable
1	Man United	74	reliable
2	Liverpool	69	reliable
3	Chelsea	67	neutral
4	Leicester	66	neutral
5	West Ham	65	neutral
6	Tottenham	62	neutral
7	Arsenal	61	reliable
8	Leeds	59	neutral
9	Everton	59	neutral
10	Aston Villa	55	unreliable
11	Newcastle	45	neutral
12	Wolves	45	neutral
13	Crystal Palace	44	very reliable
14	Southampton	43	reliable
15	Brighton	41	very unreliable
16	Burnley	39	reliable
17	Fulham	28	neutral
18	West Brom	26	reliable
19	Sheffield United	23	reliable

Figure 4: Reliability vs Final Standings

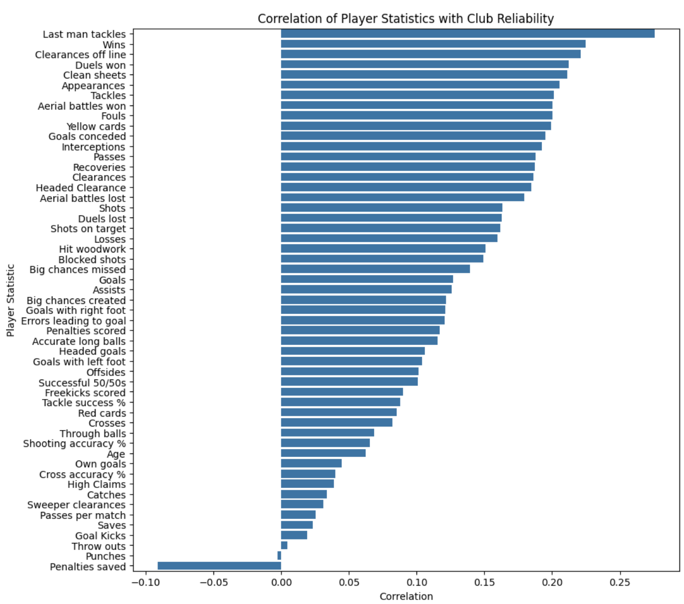

Figure 5: Correlation between Player Statistics and Reliability

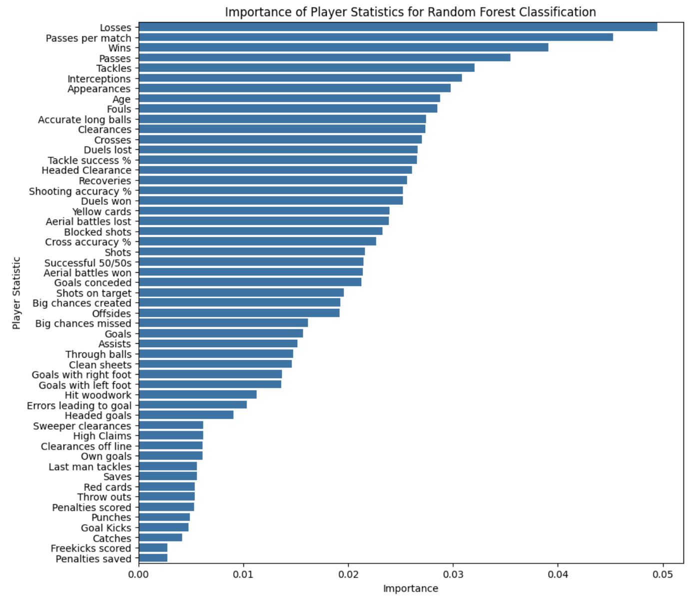

Figure 6: Importance of Player Statistics in Random Forest

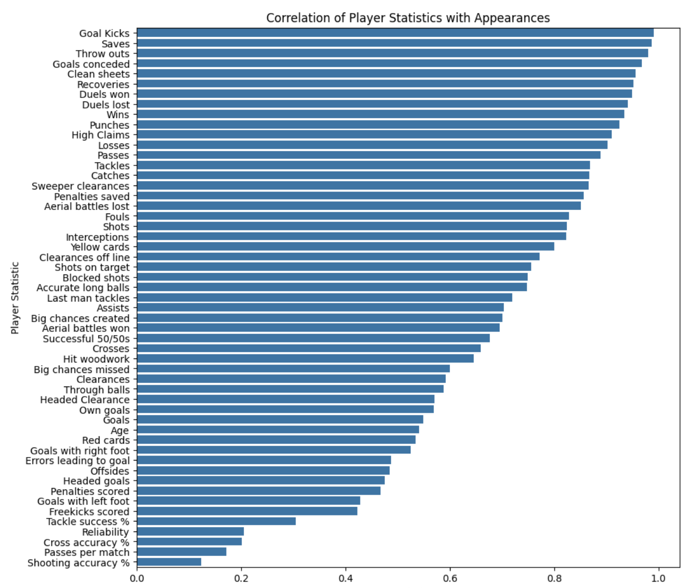

Figure 7: Correlation between Player Statistics and Appearances

outcomes, so the number of fouls might be a strong indicator of reliability. A future experiment should normalize the statistics (as applicable) of each player by their number of appearances, which should reduce the conditional dependence in the model and better find statistics to represent reliability. Finally, a future experiment could try more sophisticated dimensionality reduction algorithms such as principle component analysis. It would also be valuable to perform this analysis on different seasons to determine whether the same statistics are useful representations of reliability, with respect to the betting odds.

4.4 Estimating the Value of Reliability

We compare the performance of two random forest classifiers to estimate the value of reliability. Both classifiers are trained on the first 80% of games in the 2020 season, using aggregated player statistics (specifically the mean) to represent the matchup from the lens of player performance. The first model uses every available statistic to represent a team, while the second model uses only the top 10 statistics from the pearson correlation coefficient and the feature importance analysis above. The accuracy, classification report, and confusion matrices of each model are evaluated on the last 20% of games in the season, and displayed below.

```
Clubs represented by the means of all player stats:
Accuracy:  0.47368421052631576
Classification Report:
              precision    recall  f1-score   support

           0       0.49      0.59      0.54        32
           1       0.17      0.15      0.16        13
           2       0.60      0.48      0.54        31

    accuracy                           0.47        76
   macro avg       0.42      0.41      0.41        76
weighted avg       0.48      0.47      0.47        76
```

Figure 8: Classification Report of Model 1

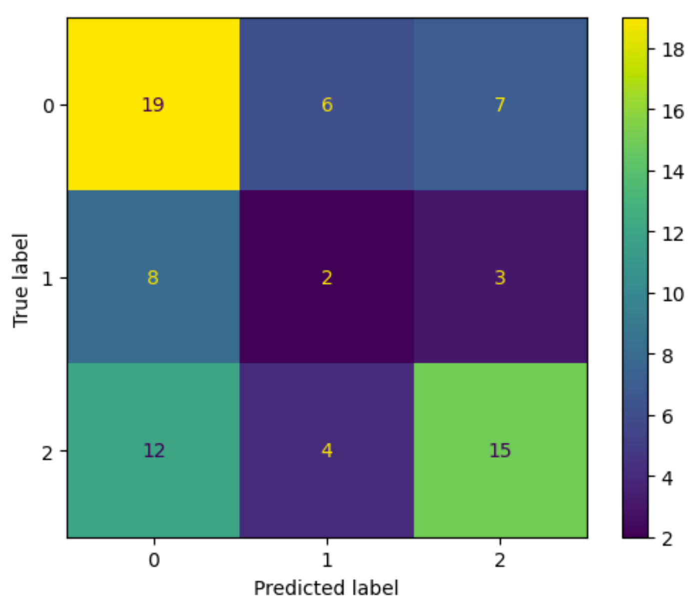

Figure 9: Confusion Matrix of Model 1

Figure 8 shows that the first model, which uses all player statistics, has an accuracy of 47.4%. The model is much worse at predicting draws than away or home wins, with only a 0.16 f1-score for draws versus a 0.54 f1-score for both home and away wins. This is echoed in the confusion matrix of Figure 9, which shows just two correctly identified draws. Unlike the classifier based on betting odds, however, which predicted no draws as shown by Figure 2, this model does predict draws, just wrongly. This might be because the average player statistics between teams are typically quite similar, and might make every team look similarly skilled. A future experiment could try another method of aggregation than taking the mean, such as the max of some statistics which might better represent how well a team will perform due to their most talented players. With respect to the model's overall accuracy, the accuracy is relatively good for predicting between three outcomes of relatively similar frequency (as discussed in Section 4.1) but the aggregation of individual player statistics doesn't necessarily capture a team dynamic, and thus has a limited predictive capability. To estimate the value of player reliability, however, we compare these results to the second model.

```
Clubs represented by only "reliability" stats:
Accuracy:  0.4868421052631579
Classification Report:
              precision    recall  f1-score   support

           0       0.49      0.59      0.54        32
           1       0.12      0.08      0.10        13
           2       0.59      0.55      0.57        31

    accuracy                           0.49        76
   macro avg       0.40      0.41      0.40        76
weighted avg       0.47      0.49      0.47        76
```

Figure 10: Classification Report of Model 2

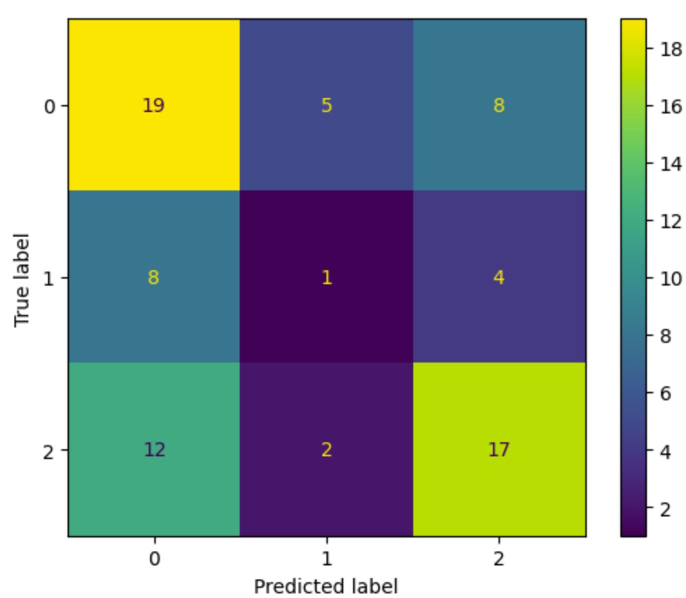

Figure 11: Confusion Matrix of Model 2

Figure 10 shows that the second model, which uses only a subset of player statistics representing reliability, has an accuracy of 48.7%. Most notably, this accuracy is similar, and in fact even higher, than the accuracy of the previous model. Therefore, most of the predictive power does seem to lie in these statistics, and if anything, many of the other statistics might simply be adding noise to the model, thus reducing the first model's performance. Figure 11 shows that the confusion matrix of this model predicted even fewer draws correctly (just one), but this actually brings the model closer to the model that uses betting odds (as demonstrated by Figure 2). This would contribute to why this model has better performance than the first, as the betting odds model remains considerably better than both of these. Ultimately, this is likely because the pre-game betting odds created by professional bookmakers leverage far more information and computational power than used by our two models to set their odds such that they quite accurately represent the true underlying probability distribution of the outcome of a match. A future experiment should use these same statistics of reliability to predict the outcomes of matches in a different season to verify these results.

A future experiment could try subtracting one team's mean player statistics from the other, as it captures the difference in a given statistic between the teams. To do so would simplify the information necessary for the model to interpret, but reduce the total amount of available information, such as the absolute magnitude of a given statistic (say both teams tend to score a lot of goals, or get a lot of fouls).

4.5 Normalizing Player Statistics and Eliminating Noise

In an attempt to approve the player models in the previous section, we decided to take a deeper look at what may be skewing our statistics, and also ways we could try to determine a players impact on a match-to-match basis. When looking at player appearances, we noticed that the data was heavily skewed right. Zero appearances was already taken out of the data as the statistics were not useful to predict match outcomes, but even one and two appearances had around 20 or more players contributing, which meant players who may only be playing a few minutes are contributing to our data and skewing it. Ultimately, more than 4 appearances was chosen in hopes that we could get players who were at least playing half a match or more when it came to their minute total. The assumption here is that the lower appearance players are those who get subbed on later in a game and most likely get around 10 minutes of play, so those with less than 5 appearances most likely have around 40 minutes or less played. We did not want to just get rid of all inexperienced players as they do still make an impact on the matches and can make an impact on predicting the results, so eliminating the skew was the most important part.

Next we took the statistics goals, assists, shots, tackles, and saves, and made new statistics by normalizing them using appearances. This gave us an idea of what kind of impact each player may have had on a match-to-match basis. These statistics could also give us a further look into how the team may be performing each match. When looking at stats such as saves per match and tackles per match, if both are high, it could mean that the team is defending more often than other teams, which means they are not in possession of the ball often for the majority of their matches. If the saves per match is high and the tackles per match is low, then this could indicate that not only is the team not possessing the ball often, but it could indicate their defense is performing poorly as well. If the tackles per match is high and saves per match is low, it could still show that they are not possessing the ball well, but it could show that they at least are defending well. A combination of these statistics could help the model paint an overall better picture of what is happening within these games. We then took this cleaned up data with added statistics and reran them through the same models as the previous section.

```
Clubs represented by the means of all player stats:
Accuracy:  0.5131578947368421
Classification Report:
              precision    recall  f1-score   support

           0       0.54      0.66      0.59        32
           1       0.22      0.15      0.18        13
           2       0.57      0.52      0.54        31

    accuracy                           0.51        76
   macro avg       0.44      0.44      0.44        76
weighted avg       0.50      0.51      0.50        76
```

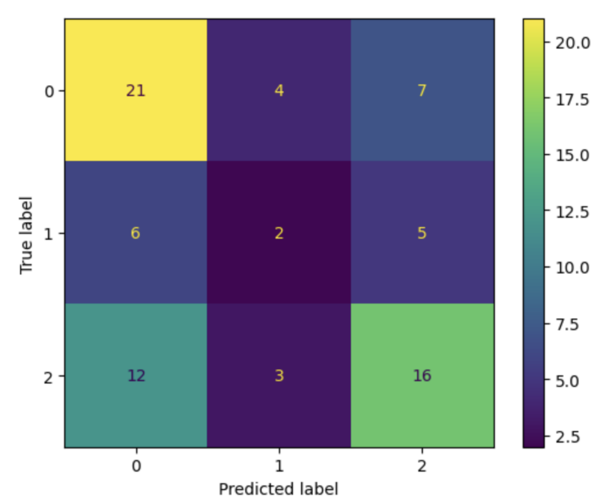

Figure 12: Model 1 after Normalization and Elimination

```
Clubs represented by only "reliability" stats:
Accuracy:  0.5263157894736842
Classification Report:
              precision    recall  f1-score   support

           0       0.54      0.59      0.57        32
           1       0.25      0.23      0.24        13
           2       0.62      0.58      0.60        31

    accuracy                           0.53        76
   macro avg       0.47      0.47      0.47        76
weighted avg       0.52      0.53      0.52        76
```

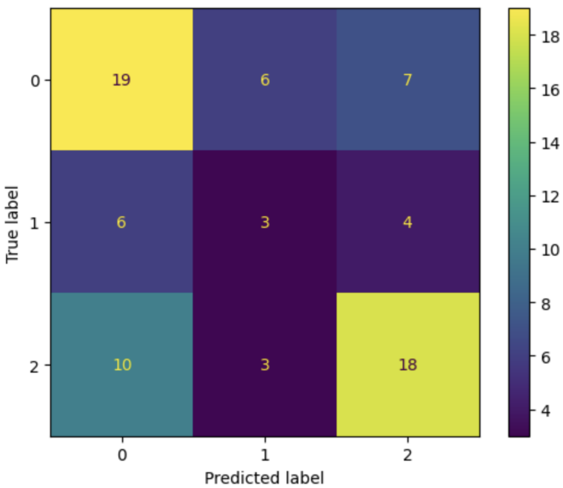

Figure 13: Model 2 after Normalization and Elimination

The two models now have better accuracy with the model using all player stats having an accuracy of 51%, which is 4% better, and the model using just the "reliability" stats have 53% accuracy, which is 5% better. Both perform better overall at predicting the full time results, including doing better with ties, but there is still more work that can be done such as combining these results with the results of our betting model and seeing if that can improve the accuracy.

5 Discussion

Our analysis demonstrated the strength of betting odds as a baseline predictor for match outcomes, with the logistic regression model achieving consistent accuracy. We believe this means that the pregame betting odds closely match the underlying probability distribution of its outcome. Furthermore, we analyzed the model on a club granularity, revealing which teams the model is best or worst at forecasting. This provides insight into which teams are over/underperforming, and which teams a savvy bettor might look out for. For analysis at a player level, the pearson correlation coefficient and random forest feature importance revealed which player statistics best contribute to a team's reliability, including shot accuracy, assists, and fouls. Normalizing these statistics by player appearances further enhanced the reliability of the models, ensuring better comparisons between players with varying playtime. Finally, we determined that reliability is indeed a valuable characteristic to players, as our random forest classifier which used only the (aggregated) player statistics that best represent reliability performed better than the model using all (aggregated) player statistics.

Despite these achievements, our data and models have several crucial limitations. Firstly, predicting draws remained a significant challenge for the logistic regression betting odds model. Additionally, the dataset of matches does not have information about which players on a team were actually playing, or likely to play. Similarly, the dataset of players did not have information on time played, only the number appearances which is far less precise. To more thoroughly identify the most relevant player statistics we could try dimensionality reduction algorithms such as principle component analysis. Finally, an ideal predictive model would have similar accuracy to the betting-odds model, which our relatively simple random forest classifiers do not achieve. In general, this could be improved with more precise player data, improved aggregation techniques, and addition contextual factor data such as game-day conditions and strategic decisions. Our results and the real world implications, however, underscore the importance of linking player-level statistics with team-level outcomes, offering practical insights for coaches and analysts seeking to optimize performance.

Moving forward, assembling better data or implementing more sophisticated classifiers such as deep neural networks could improve predictions, particularly for draws and other challenging match outcomes. Similarly, incorporating live game data or additional contextual variables could also enhance the predictive power of the models. Furthermore, extending this analysis to multiple seasons would provide a more comprehensive view of reliability trends over time.

6 Conclusion

This project aimed to analyze the reliability of teams, players, and betting odds within the English Premier League by leveraging historical data and predictive modeling. By integrating player statistics, match outcomes, and betting odds, we uncovered valuable insights into the dynamics of reliability at both the team and player levels.

We determined how well pregame betting odds represent the underlying probability distribution of a match, with a predictive model yielding 57% accuracy. We identified which teams this model is best and worst and predicting the games of, and investigated outliers with respect to the final league's standings, such as Crystal Palace. We determined which player statistics contribute most to a club's reliability, such as last man tackles, wins, and passes. We discovered that a model predicting match outcomes from player statistics performs better when using only these statistics of reliability (49% accuracy) than when using all player statistics (47% accuracy). Finally, we discovered that if we eliminate some noise based off player appearances and normalize player statistics using appearances, we can get the models to be a little more accurate, with all player statistics having 51% accuracy and "reliability" having 53% accuracy.

This project contributes to the growing field of sports analytics by providing actionable insights into the factors driving reliability in the English Premier League. These findings hold value for a range of stakeholders, including coaches, bettors, and bookmakers, and pave the way for more advanced analyses in the future.

References

[1] Uncovering Betting Patterns in the Premier League Dataset
[2] All Time Premier League Player Statistics Dataset

[3] Wilkens, Sascha. 'Sports Prediction and Betting Models in the Machine Learning Age: The Case of Tennis'. 1 Jan. 2021 : 99 – 117.
[4] Wunderlich F, Memmert D (2018) The Betting Odds Rating System: Using soccer forecasts to forecast soccer. PLOS ONE 13(6): e0198668. https://doi.org/10.1371/journal.pone.0198668
[5] Draws Stats - England Premier League

A Honor Code

We affirm that we have not given or received any unauthorized help on this project, and that this work is our own. We have upheld the principles of academic integrity in accordance with the University of Colorado Boulder's Honor Code.

B Contributions

- Florian Frick 5502: Betting odds logistic regression model, analyze club and player reliability, random forest models to predict match outcome from player stats, project presentation, project report
- Joao Nicastro 4502: Researched prior/related work. Examined datasets. Worked on project presentation and report.
- Joseph Gildhouse 4502: Abstract, normalizing player data, eliminating noise, and reusing the model to determine results of these