

# HISTORIC ANALYSIS OF PERFORMANCE ACCURACY OF MATCH PREDICTORS

By Michael Dubem Igbomezie



[Link to Upwork Profile](#)  
michaeligbomezie@gmail.com  
November 28th, 2023

## 1. Introduction

This report summarizes the insights and information uncovered from the analysis of three years (three seasons) of historic data on football match scores from approximately twenty-four (24) different leagues. The analysis focused on determining the reliability and accuracy of a system of predictors, based on how much accurate results these predictors have produced in the past three football seasons.

The dataset used for this analysis was extracted from the [worldfootball](#) website. The entire analysis can be group into four main sections;

- Accuracy of predictors
- Most accurate prediction sources
- Teams occurring the most in accurate predictions
- Accuracy trend along season timeline

## 2. Result and Discussion

- A. Accuracy of Predictors: This subsection focuses on the overall accuracy of the predictors independent of season and leagues, as well as when it is dependent on season, league and both.

The result from the analysis show that over the last three football seasons, the predictor produced an overall accuracy of 29.6%. We can also notice the same level of accuracy when the results are analysed by seasons, with the highest accuracy measuring at about 10.6% and occurring in the 2022/2021 football season.

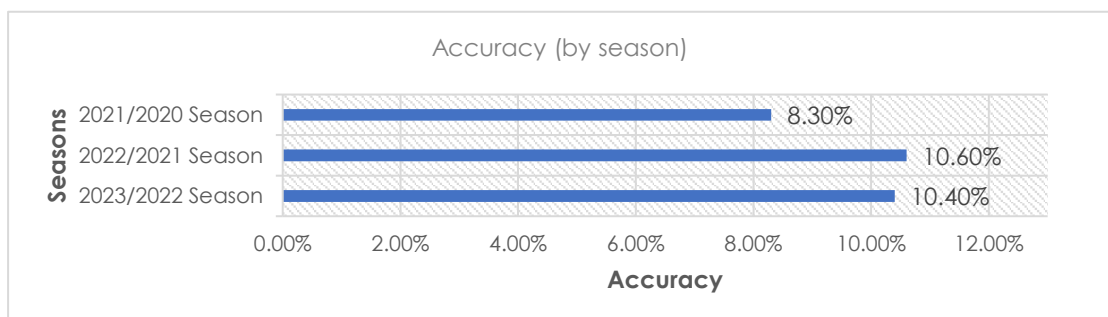


Figure 1. Predictor Accuracy by season

In a similar manner the accuracy was measured by league across all three seasons and the top 5 leagues with the most accurate predictions were extracted.

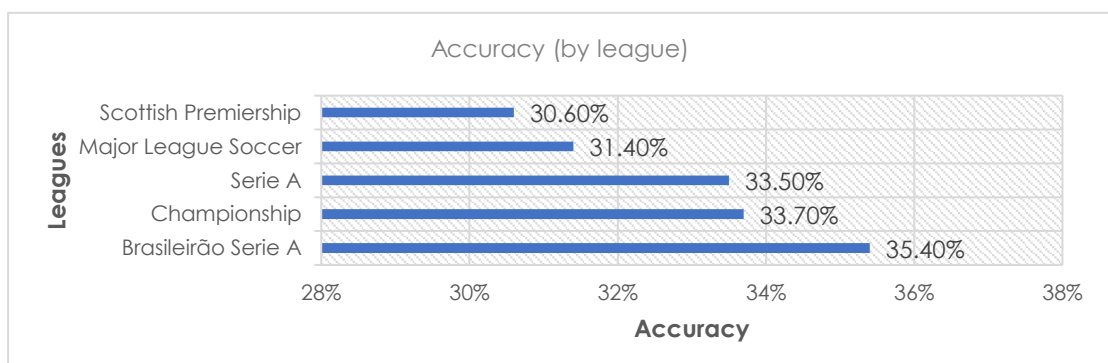


Figure 2. Predictor Accuracy by league

The results above show that the predictors have the highest level of efficiency in the Brazilian league, at 35.4% accuracy out of all prediction made in this league.

Finally the accuracy of the predictors were checked across league for different seasons. This analysis showed that 2023/2022 season had the league with the highest accuracy at 38.9%, followed by 2022/2021 season with the highest league accuracy at 36.8%, and finally the 2021/2020 season with the highest league accuracy at 34.2%. However, what was interesting is that the predictors seem to have a higher accuracy with the more recent football seasons. Also, when the top 5 most accurate leagues were extracted from each season and the leagues were arranged by most occurring leagues in the top 5 most accurate leagues of all three season, the results again show the Brazilian league as having the most occurrence.

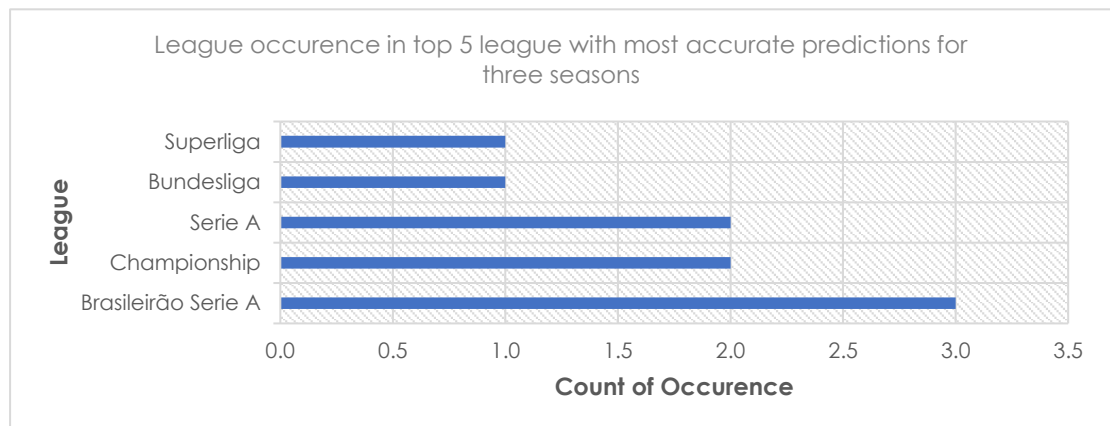


Figure 3. Leagues that occurred the most when top five leagues with most accurate predictions were selected by season.

The analysis conducted on the accuracy of the predictors show a relatively poor performance in accurately predicting match outcomes. Even when the analysis was conducted by season, and then by league, the overall performance of the predictors were still at best sitting around 35-38% accuracy. It was observed that the more recent season showed higher performance for the predictors but more season will have to be observed to confirm if this is a positive trend for the accuracy of the predictors. The Brazilian, Serie A and Championship leagues also kept reoccurring amongst the leagues with the highest predictor performance.

- B. Most Accurate Prediction Sources: This subsection focuses on the sources and types of predictors that gave rise to the most accurate predictions. The three main data sources were the home team historic matches, the away team historic matches, and the home and away team head-to-head match. While the types of predictors included repeating match scores, or consecutively increasing match scores, amongst others. These predictor types were combined with two main filters, which were, skipping certain match entries and filtering by leagues. These combinations were executed for all the different sources of data to make up what is referred to as the prediction sources.

The prediction sources that occurred the highest number of times in all accurate predictions was check across the entire dataset, and from the results, most of the accurate predictions were because of patterns in the home team and away team history. Also four out of the top-five prediction sources were without any kind of filter (neither skipping rows, nor filtering by league) as seen in figure 4.

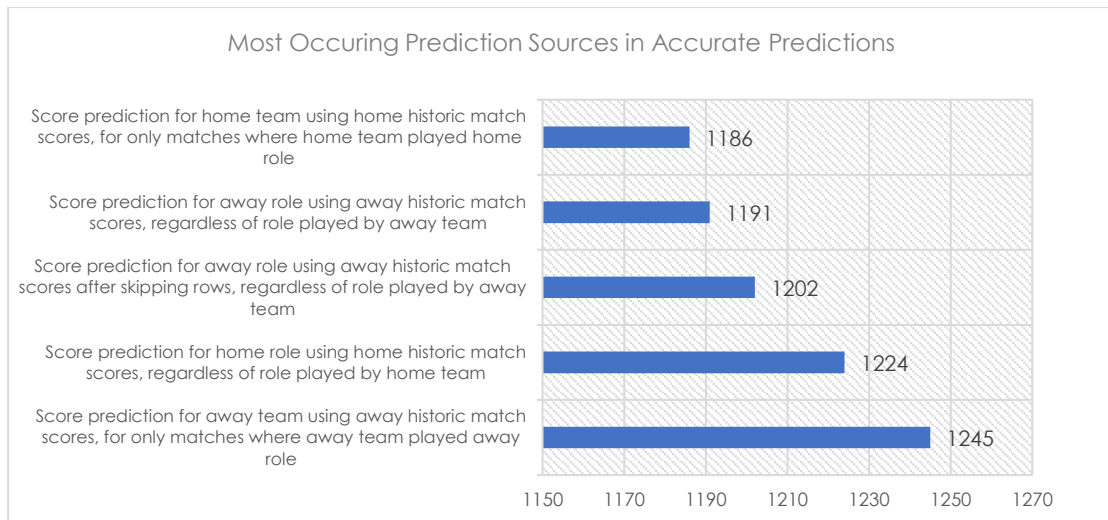


Figure 4. Prediction sources/types that most led to accurate predictions

Similarly the prediction sources were checked with respect to the season and the top five prediction sources were extracted for each season. The most occurring prediction sources in these sets of top five prediction sources were extracted. The results below (figure 5) show a similar pattern, where all the prediction sources that occurred the most in the top five spot for each league were all from the home team and away team history.

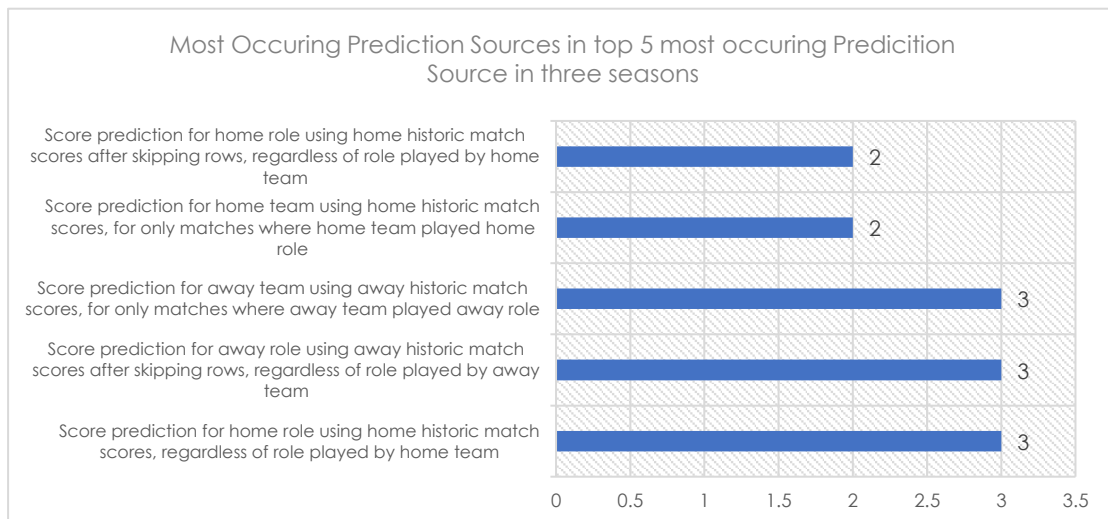


Figure 5. Prediction sources/types that occurred the most when top five prediction sources that most led to accurate predictions were selected by season.

Another similar pattern can also be observed in figure 6, when these prediction source occurrences are checked across each league. The results show that the most occurring prediction sources in the the top five most accurate prediction sources for each league are from the home team and away team histories. Again we also see that four out of the top five prediction sources had no filter applied before being analysed; however, we can also notice that in both instance where there was a prediction source having a filter implemented, it was always the “skipping rows” filter.

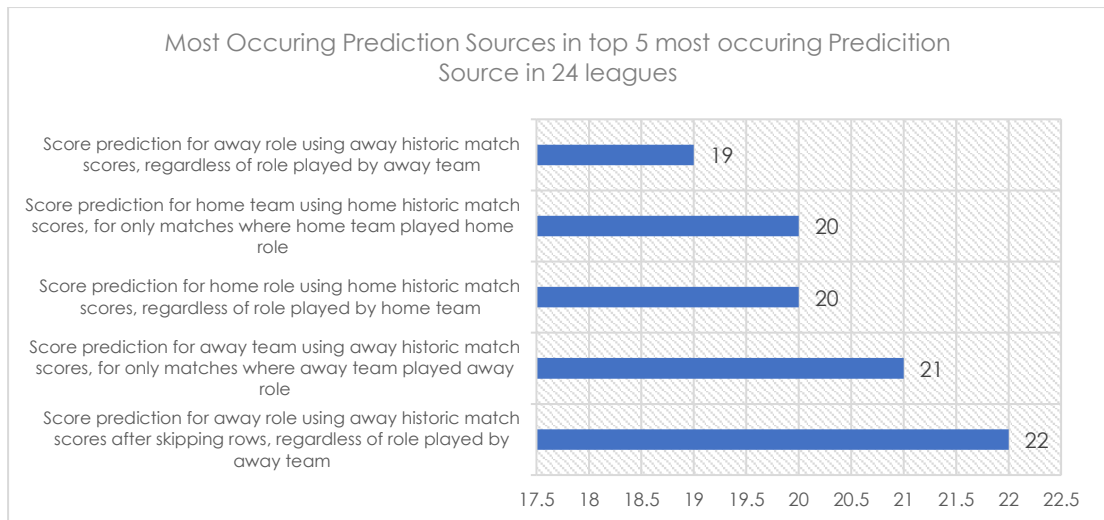


Figure 6. Prediction sources/types that occurred the most when top five prediction sources that most led to accurate predictions were selected by league.

- C. Teams occurring the most in accurate predictions: This subsection is quite similar to the previous section where we checked for the prediction sources that occurred the most in accurately predicted matches; however, the difference is that in this subsection, the teams that occurred the most in accurately predicted matches were checked instead.

After checking the entire dataset for the most occurring team in accurately predicted matches, Queen's Park Rangers turned out to be at the top of this list. Four out of the top five team in this list are European teams with just one team from the Brazilian league, which is Corinthians SP.

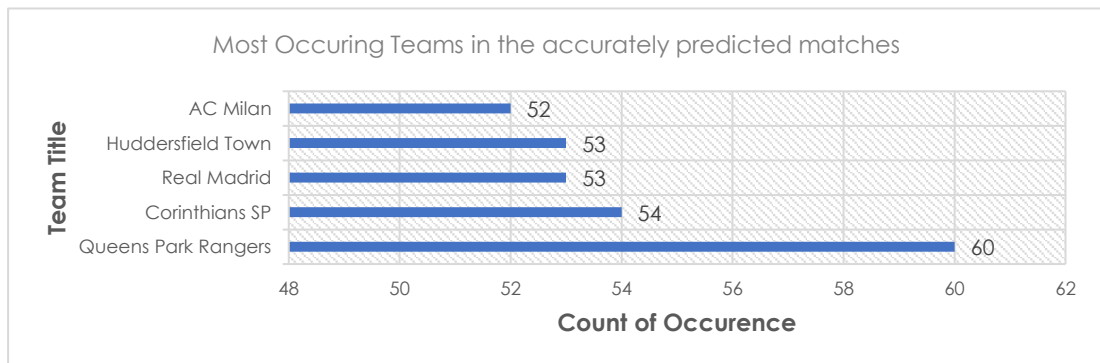


Figure 7. Teams that occurred the most in accurately predicted matches

In addition to how frequently these team appeared in the accurately predicted matches, their number of accurate occurrences were checked against their overall occurrence to see the accuracy associated with each team.

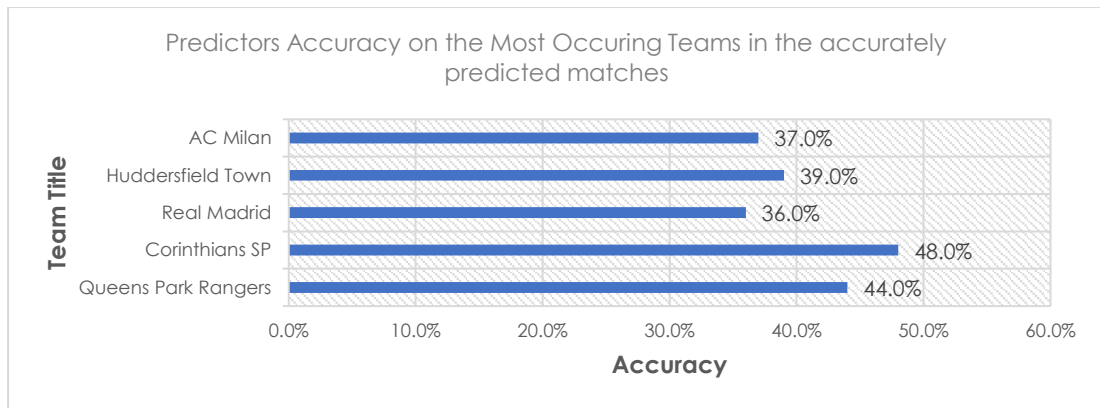


Figure 8. Accuracy of the predictors for all the occurrences of the teams listed in the previous figure.

The results above show that Corinthians SP had the second highest occurrence in the accurately predicted matches and then the highest predictors' accuracy in all of its occurrence in the historic data. This ties into the Brazilian league having the highest accuracy among the other leagues as well. We also see a similar effect for Queen's Park Rangers which also reflects the fact that Championship league had the second highest accuracy. A similar pattern was also observed when running the same analysis by the different seasons, which showed Queen's Park Rangers, Corinthian SP and Huddersfield Town in the top five teams with most accurate predictions.

- D. Accuracy trend along season timeline: This final section was inspired by the observed increase in accuracy of predictors in the more recent seasons. In this subsection, the seasons were split by weeks and the accuracy per week was plotted to check for any pattern in accuracy trend across each season, to have an idea of when the system could be more reliable and possible investigate as to why (if any).

The charts below (figure 9, 10, 11) show the weekly accuracy trend in all three seasons for the top 3 leagues with the highest positive slope. The highest positive slope indicates an increase in accuracy as the weeks of the season pass by.

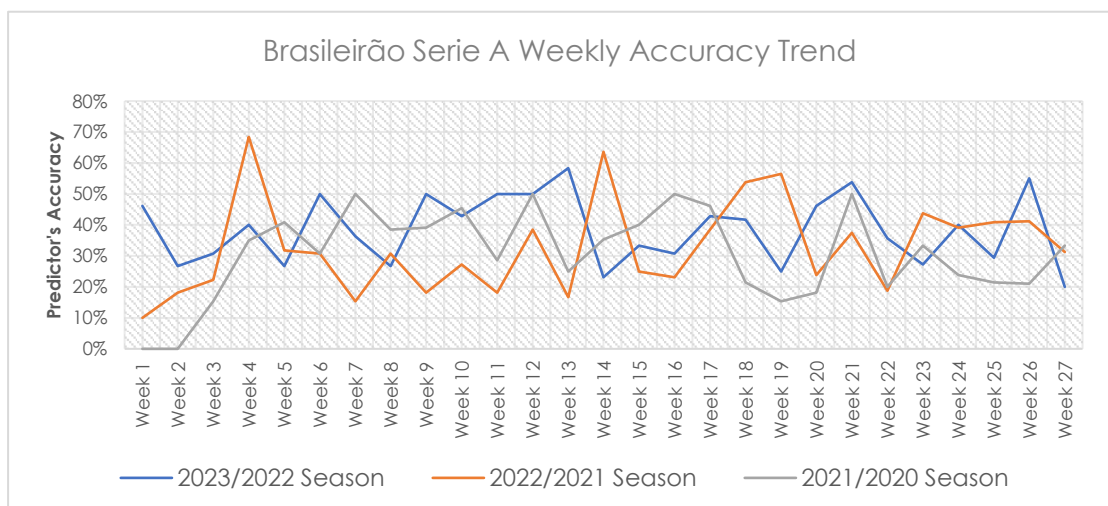


Figure 9. Weekly accuracy trend for Brasileirao Serie A league, for all three seasons.

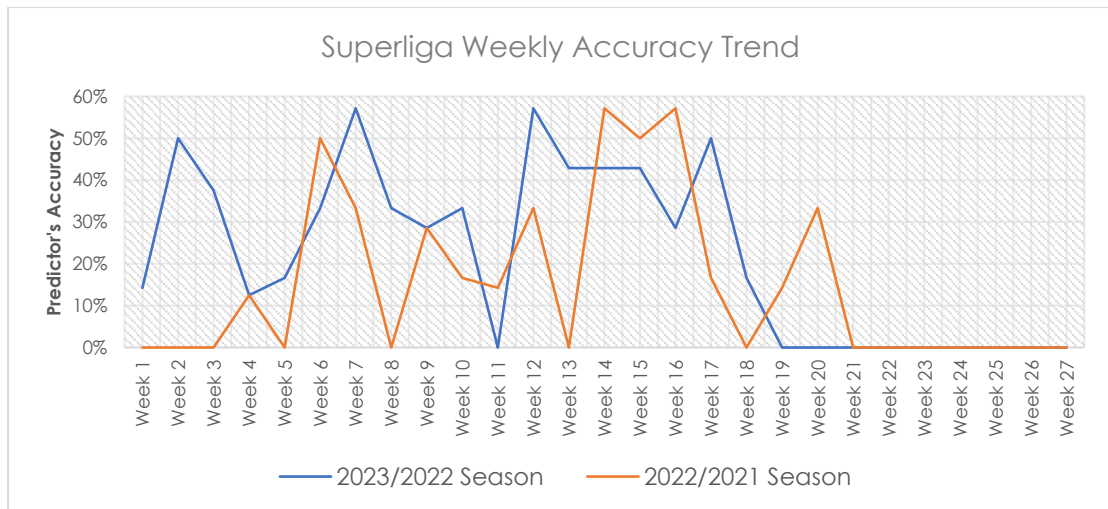


Figure 10. Weekly accuracy trend for Superliga league, for all two seasons

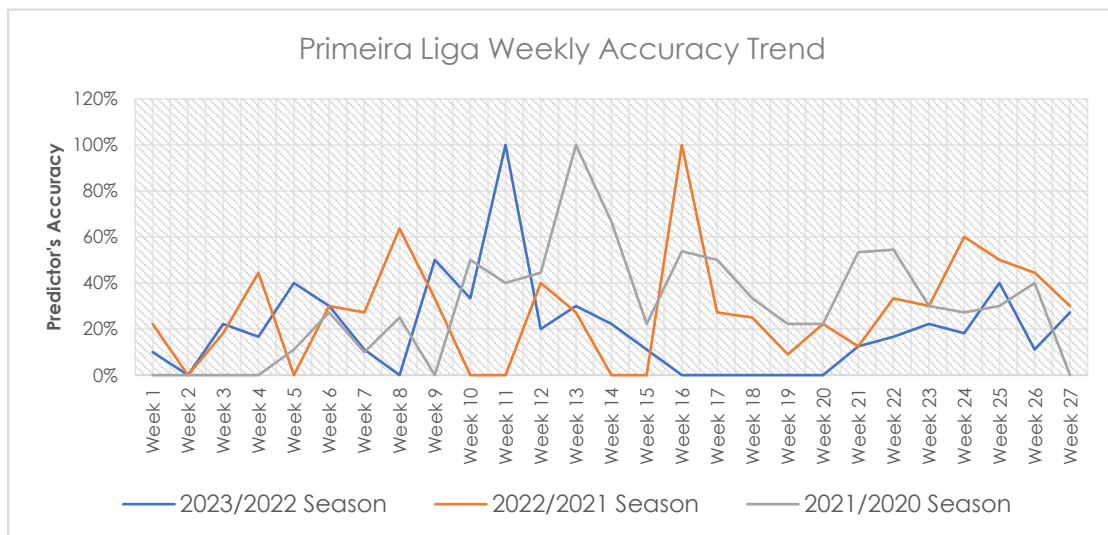


Figure 11. Weekly accuracy trend for Primeira Liga league, for all three seasons

Judging from visual assessment, a significant difference can be spotted in the weekly accuracy trend of the three seasons for all three leagues, with diverging highs and low which indicate no repeated pattern or unreliable trend. However, in addition to visual assessment, the similarity of the weekly trend for each of the season were calculated (using the Euclidean Similarity) to check if there are repeated patterns in the highs and lows of the weekly accuracy by season. The best average similarity amongst the three seasons was recorded for Primera Liga at a score of 0.36, followed by Superliga at a score of 0.46, and then finally Brasileirão Serie A at an average score of 0.49. Considering that these similarities are closer to 0.0 than they are to 1.0, it implies that the weekly accuracy trend for the three season in these top three leagues (with the highest positive slope) are more similar than they are dissimilar which could be looked into in future analysis.

## 2 Conclusion and Recommendation

Given the overall low performance of the predictors across the three season from the leagues' history, it will not be safe to rely solely on the developed system of predictors for match outcomes. However having said this, it is also important to note some trends that emerged from all the analysis conducted on the extracted data. There was an increase in the predictors' accuracy in the more recent football

seasons, although a lot more seasons will need to be analysed to know if this increase in accuracy is a trend and if it is reliable. The Brazilian league and the Championship league were the two leagues in which most of the accurate predictions were made by the predictors. This same pattern also reflected in the team that appeared the most in accurate prediction, in which the top two team were also teams that played in the top two leagues mentioned above. Furthermore, the analysis also showed that the home team and away team history were the major sources of accurate prediction by the system.

My professional opinion would be further monitoring of these systems of predictors before depending on them for any important decisions, as future modification and fine-tuning might be necessary to increase efficient. I would also recommend integrating some league-based results from the analysis into the user interface to further aid with decision-making.