FPL Capstone Report

Farhan Kassam | April 11, 2023

Problem Statement

In this project, I attempted to create a model to predict the points a player would receive in a Fantasy Premier League team based on their real-life performance. This project would be beneficial for heavy FPL players who want a data-driven team to rise the ranks in both their mini-leagues and the world at large. Players who can score in the top ranks generally receive prizes for their efforts.

Soccer is becoming increasingly dependent on individual player statistics which are then used to calculate the points earned by a player in FPL. The Premier League themselves have a model to predict player points, but if our model is successful, we may have a better chance at winning the game and being awarded those prizes.

Data Collection

The data was collected from an open-source Fantasy Premier League API. A cleaner version of this data is available in a public GitHub repository hosted by Anand, Vaastav. This user uses automated python scripts to pull from the FPL API to keep various player statistics stored in the GitHub containing both raw and cleaner versions of the data.

The dataset that I used in this project was from the 22/23 Season for matches played between game-weeks 1-24. I used the merged-gw comma separated file which contained information on all the game-weeks 1-24 for all players in the Premier League. The original data had 44 columns and 15,866 rows.

Data Cleaning and Preprocessing

Firstly, I checked the data to see if there were any missing values and imputed them appropriately to avoid any data loss for the modelling phase. I removed any players who did not get play time, to include players who were injured, benched for rotation, or substitute players for a given match. This removed approximately 55% of the rows where the points earned by a player would be 0 since they did not play in the match.

I selected features to use both in the following modelling and team selection notebooks. The data then included player statistics earned in the match such as how long they played (minutes), number of goals and assists, number of saves, if the player was booked (given a yellow or red card), as well as FPL special statistics of influence, creativity, and threat. Lastly, I also included the points a player was expected to score, a player's value in the Fantasy game and the true points scored.

For the exploratory analysis, I decided to discover which positions contribute the most to the features that earn points and forfeit points. This will help us decide which positions may be worth investing more of our budget in for our FPL team.

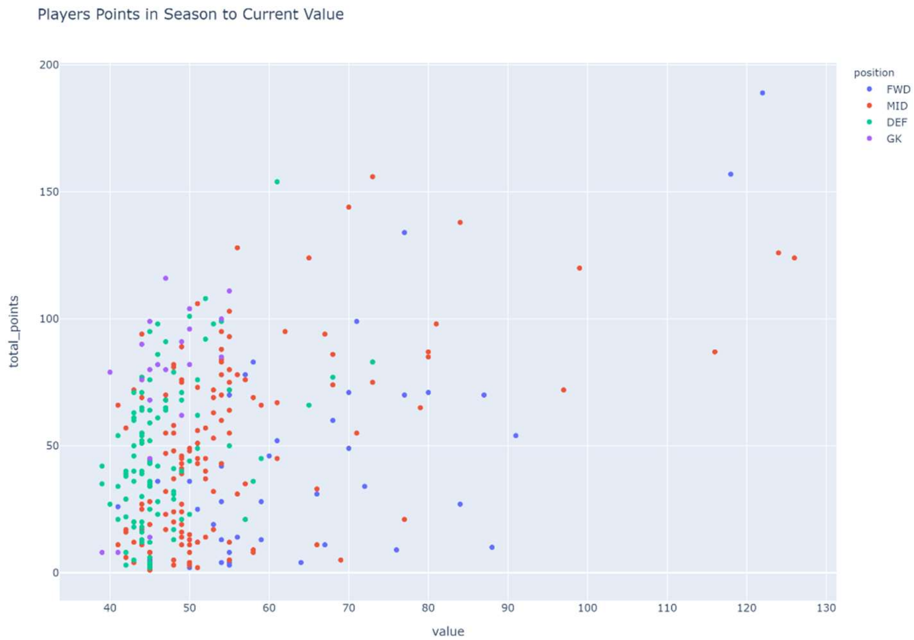

Figure 1: Distribution of total points earned in season to date to their current value by position.

The above chart shows the value of a player by the points they have earned in the season so far, denoting the possibly undervalued player picks where the total points they have earned are high and their value is low.

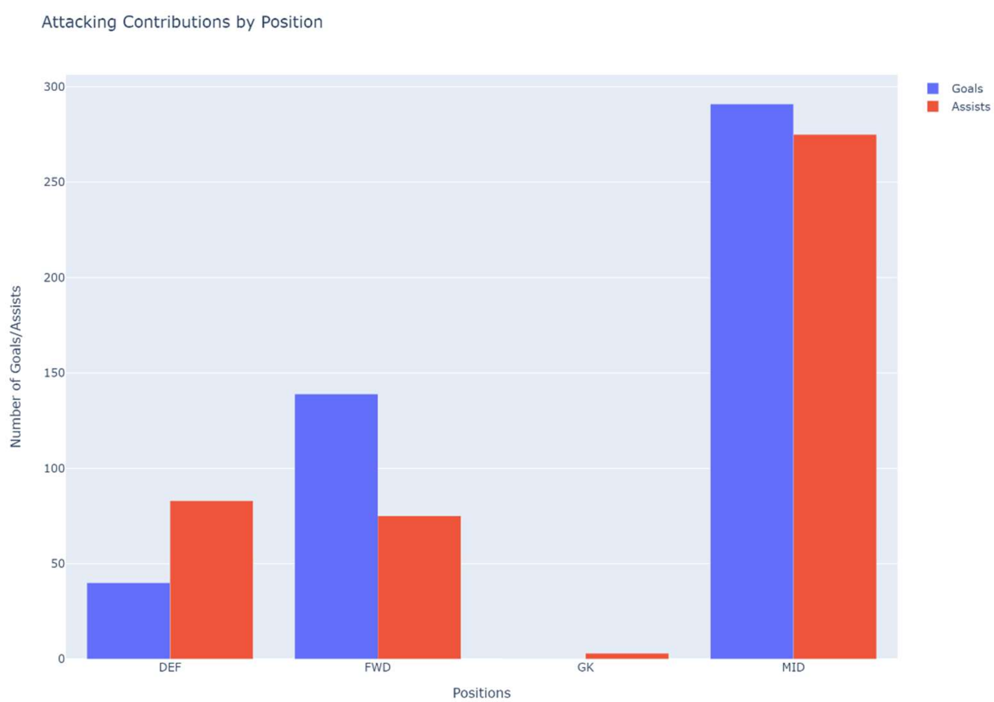

Figure 2: Attacking contributions by position.

In figure 2, we discover that the most valuable attacking players are our midfielders as they have the most goals and assists compared to any other position followed by our forwards. This provides evidence on why we should invest more in midfielders for our FPL teams.

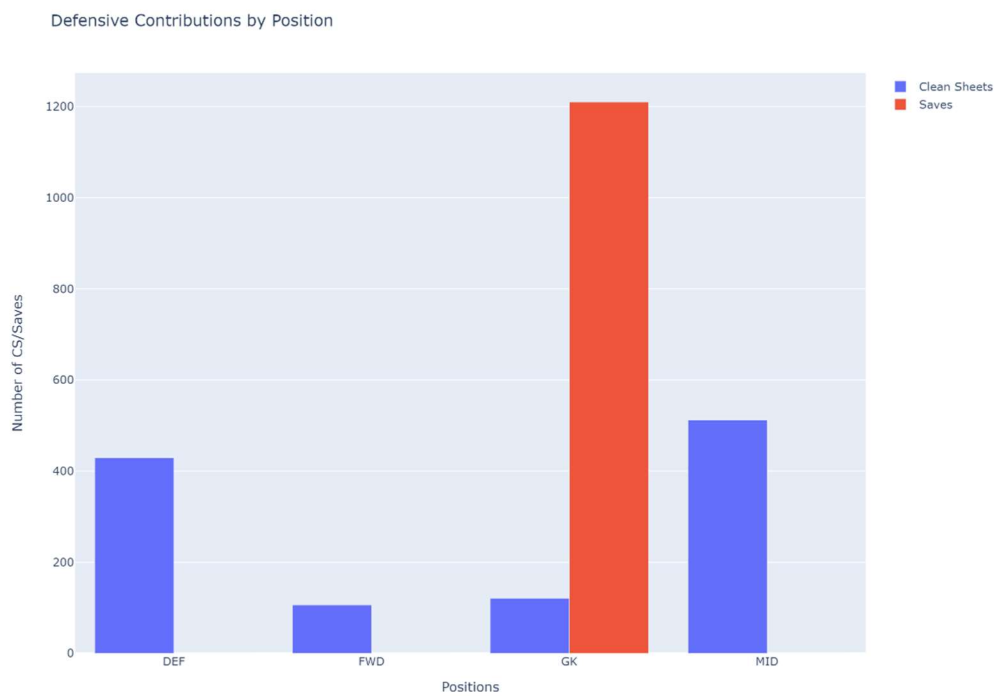

Figure 3: Defensive contributions by position.

In figure 3, we discovered that the most defensive contributions came from our defensive positions defenders and goalkeepers. The amount of clean sheets and saves are far lower than the attacking contributions. Additionally, the FPL scoring scheme doesn't value defensive efforts as much as attacking efforts. With both of these reasons, it is safe to conclude that we should invest more in attacking positions such as midfielders and forwards.

Modelling and Insights

I selected the features as previously mentioned ran a preliminary linear regression model. After viewing the most important features which ended up being goals scored, assists and clean sheets, I decided to run a more sophisticated model with hyper parameter optimization. I used a lasso regression since I wanted to minimize the effect of non-important features. The resulting important features for lasso regression were identical to the linear regression.

Lastly, I decided to setup a random forest regressor which would have multiple decision trees working together with the goal of increasing the variance explained in the points earned in each match. The important features here were also identical.

The model with the best accuracy was the random forest regressor whose results are shown in figure 4.

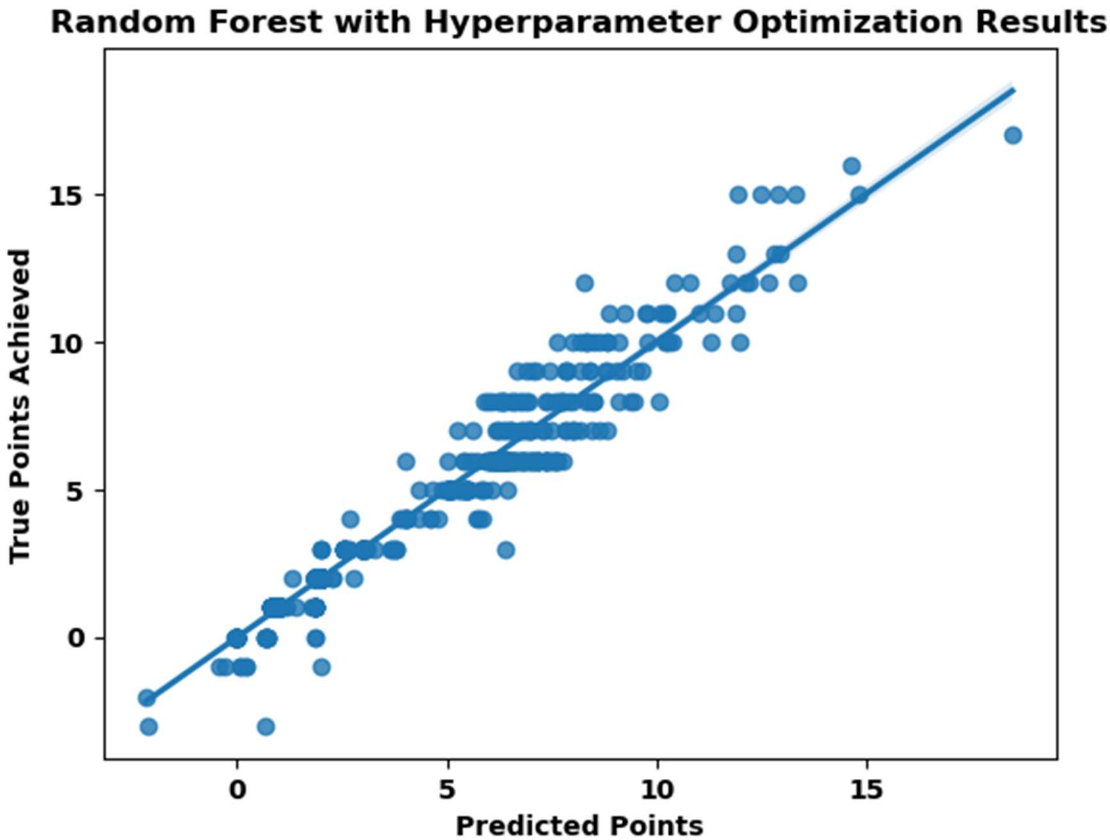

Figure 4: Random Forest Regressor model results.

The model was able to correctly estimate the variance in the points earned (r-squared value) based on the features provided with an accuracy of 97% on the train set and 96% on the test set.

Findings and Conclusion

I exported the model's predictions to compare the team selection of FPL's predicted points, my model, and the true results for game-week 24. I found that my model was an 80% match for the team selected whereas FPL's prediction was a 46% match. The overlap between my predictions and FPL's was 40%.

This brings me to question the validity of the results I acquired from the modelling section, particularly the validity of my feature selection. The features I used to predict points earned were from the results of the match and therefore have a direct relationship to the points earned. Additionally, these features would not be available until the match has ended so they cannot be used to predict beforehand which defeats the purpose of our objective to predict player points. This explains why the model's r-squared value was so high since my model essentially became a replica of the scoring scheme/rules of FPL.

In the future it would be beneficial to include expected player statistics such as expected goals and expected assists as these may be more appropriate predictors of a following match.

References

1. Anand, Vaastav. (2023). *FPL Historical Dataset*. https://github.com/vaastav/Fantasy-Premier-League/.