iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46598

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Data Analytics as a Key to Fantasy Premier League

Arya L Gowda
Department of Electronics and Communication Engineering, Visvesvaraya Technological University, Bangalore, Karnataka, India

***Abstract:** This paper aims to provide the reader with a review of the Fantasy Premier League (FPL) driven by data science, A competition in which participants select a fictional team from among the players in the real-world league and earn points according to the actual performance of the players. Fantasy Premier League (FPL) performance predictions tend to be algorithmically based solely on historical statistical data. The main factor driving fantasy sports are three main factors which are Competition, Bragging rights and money. In this paper we present a study for recommending FPL player selections for upcoming game weeks.*

***Keywords:** Fantasy Premier League (FPL), Analysis, Model, Statistics, Data.*

I. INTRODUCTION

Fantasy Premier League (FPL) is the official fantasy football game for the English Premier League, with more than 9 million players actively competing each season to reach the highest tally of points. Fantasy sports are growing in popularity, with the Fantasy Premier League leading the way in players per season. This growing popularity makes it increasingly difficult to achieve a respectable rank, let alone win in any official or paid league. Through the analysis of statistical data such as previous performances, difficulty rate of upcoming matches, analysis of the betting market, opinion of the public and experts through articles on social media and social media, we can improve our understanding of who is likely to perform well in upcoming matches.

This game requires calculations, and if working with data is not your best skill, your players are very likely to lose. Yes, you have to be a data geek to hit the crossbar accurately. The game and environment change every week, so it's important to start high. The relationship between the fantasy game and its Physical counterparts raises the question of whether those who participate in the former suffer (or benefit from) the same combinations of skill and chance that makes their physical equivalent enjoyable.

II. HISTORY

Fantasy football was developed in March 1962 at the Plaza Hotel in New York City, By Wilfred "Bill" Winkenbach. Wilfred was known for his association with Oakland robbery. He also created Fantasy Golf in the late 1950s. The first official football league was formed in 1963 with Scotty Sterling and George Ross. (Brown) They developed a system of organization and a set of rules. The league was called Greater Oakland Professional Pigskin Prognosticators League (GOPPPL) which was created via a draft. (Newman) The League consisted of his eight teams put together by three creators and friends.

During 1969, the creator ‘Andy Mousalimas’ of the GOPPPL introduced his league to a sports bar Kings X Sports Bar in Oakland but, the fantasy league did not work as plan.

The first national football league was developed in 1989, Pigskin Playoff, created by Lee Marc, Robert Barbiere, and Brad Wendkos. Pigskin Playoff was then called Daily Fantasy Sports (DFS). Fantasy sports was once played by “creating daily line-ups and participate for cash prizes”. (Militello) During, this time players used football statistics as a simple scoring system. The participants created fantasy teams, earned points for plays based off performances, traded players, and competed to gain enough points to win prizes. Proper technology was unavailable at this time.

Twenty years later, over a million people in the United States are playing fantasy football. In 1997, CBS launched their first free version of their online fantasy game. It was the first start of making a multi-billion industry. Other sport sites notice and instantly started their own games. By September 2006, over 18 million people were playing fantasy football competing against their opponents weekly. Many years later, DirectTV introduced their first RedZone channel. (Adam) The RedZone channel is a special game day channel that showed every single touchdown and big play from all NFL games. Fantasy football has improved greatly over the last 40 years. Today, “According to the FSGA, more than 59 million people played fantasy sports in 2017. 80% of them played fantasy football.” (McCormick) Created in 1962, the game Over time it has grown into one of the wealthiest multi-billion-dollar industries and is still growing and improving.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

III. FANTASY PREMIER LEAGUE

FPL [2] is a game that casts you in the role of a fantasy manager. Your task is to pick a squad of players who then score points based on their real-life performances. However, once the first week expires, you are limited to one free game transfer per week. Free transfers can only be accumulated by him twice, although there is always the possibility of carrying over to the next game week.

At the start of every season each individual is given £100m to create a team of 15 players whose price ranges between £4.5m - £13m.There are 63 goalkeepers, 204 defenders, 252 midfielders, 111 forwards; considering everyone is healthy the possible combinations are:

Goalkeepers: 63 players, 2 in squad = 2000 combinations

Defenders: 204 players, 5 in squad = 2.8 billion combinations

Midfielders: 252 players, 5 in squad = 8.1 billion combinations.

Forwards: 111 players, 3 positions = 222,000 combinations

TOTAL (product of all the above): Around 1028 combinations (10 octillion).

A. FPL Squad Builder Interface

The simpler part of the Points System is when a player scores a goal he is rewarded with points; Similarly, if the goalkeeper keeps a cleansheet in real life he will be awarded with points. As the picture (1) shows we cannot choose 15 players of the same position and therefore we have to assemble a complete team. Each position on the field has its own scoring system. Goalkeepers and defenders are most often rewarded for saving penalties {5 points}, saving goals {4 points}, etc. A midfielder sets a goal {3}. Goal striker {4}. However, players in different positions can receive different rewards, so you'll want to make sure how much each action is actually worth. For example, defenders can score in real life too. However, it is rare, and this is reflected in the points a defender receives {6 instead of 4} when a real opponent scores. On the other hand, midfielders also play a defensive role and score points for that as well. However, in this respect they are not as important as the defenders, and points can only be scored if the team keeps the opponent away. Also note that players can lose points for negative actions in real life, such as: For example, sending off {-3 points} or failing a penalty {-2 point}. Your captain gets double points every week. So, if the captain scores a goal at forward, he gets 8 points instead of the usual 4. If the captain does not play for any reason, the vice-captain takes over.

Action	Points
For playing up to 60 minutes	1
For playing 60 minutes or more (excluding stoppage time)	2
For each goal scored by a goalkeeper or defender	6
For each goal scored by a midfielder	5
For each goal scored by a forward	4
For each assist for a goal	3
For a clean sheet by a goalkeeper or defender	4
For a clean sheet by a midfielder	1
For every 3 shots saved by a goalkeeper	1
For each penalty save	5
For each penalty miss	-2
Bonus points for the best players in a match	1-3
For every 2 goals conceded by a goalkeeper or defender	-1
For each yellow card	-1
For each red card	-3
For each own goal	-2

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

B. Points Breakdown

Chips are special features of the game. These are:

1) *Wildcard:* The Wildcard allows you to transform your squad by making an unlimited number of free transfers in that Gameweek. It lasts beyond that week and is your new team.
2) *Free Hit:* This is like the Wildcard but for just one Gameweek, allowing unlimited free transfers for that round of games. Your old team resumes afterwards.
3) *Triple Captain:* For one Gameweek, Captains earn 3x the normal points instead of 2x.
4) *Bench Boost:* If you're having trouble getting 4 players on the bench, you can use this chip to score points from all 15 players. Just for one round, though.

IV. DATA-SET

The vast data set available from the official website for doing our analysis which is considered adequate by many researchers for arriving at a decision are:

•Total points •Round points •Price •Team selected by % •Minutes played •Goals scored •ComnAssists •Clean sheets •Goals conceded •Own goals •Penalties saved •Penalties missed •Yellow cards •Red cards •Leagues Saves •Bonus •Bonus Points System	•Influence •Creativity •Threat •ICT Index •Form •Times in Dream Team •Value (form) •Value (season) •Points per match •Transfers in •Transfers out •Transfers in (round) •Transfers out (round) •Price rise •Price fall •Price rise (round) •Price fall (round)

V. LITERATURE SURVEY

Selection of the initial squad if the most vital part of the game as it lays foundation to starting strong. James Asher *et al* [3] took an economic approach to the game by mainly emphasising on maximising Return on Investment (ROI) within the £100m budget. You can think of this as two potential teams with the same budget constraint of £100 million. This shows that the team with the highest ROI will be the team with the most points when maximizing the budget. There are a few cautions discussed later. The only problem with the ROI model is that FPL allows players to be selected as captains. As such, teaming up with low- or mid-value, high-ROI players may not be optimal since it will not take advantage of the double points. So, a good way to pick a balanced team is pick at least two high-value, high-return players to captain each week.

O'Brien JD, Gleeson JP, O'Sullivan *et al* analysed statistics from multiple seasons of fantasy football using a variety of statistical and graphical tools to highlight importance of decision made by managers [4]. Skill-based selection was evident in all aspects of the game, including efficient use of transfers, strong financial awareness, and exploitation of short- and long-term strategic opportunities such as captain selection and chip mechanics. Behaviour and their decision-making constitute their skill levels. These decisions suggest a combination of skill and luck that exists in fantasy sports, are the same as their physical counterparts.

In the first week of selection, the mistake often made by participants is choosing the players that had good statistics the previous season, this does not result in fruition most times as the form of individual players across seasons is not the same and only few players are consistent.

The factor we can consider is the performance of players in pre-season games, the statistics to look at may not only be goals but the number of chances they create, successful passes (which assesses how well they are fitting in the team i.e., chemistry), fitness and in certain cases defensive blocks and tackles. The suitable approach for the first week can be a Linear Regression model or a Basic Neural Network model by training it with datasets which includes historical data of above-mentioned parameters to assess a player before selection in the team.

Considering a simple Multiple Linear Regression using a full set of variables. Coefficient values can be considered an important measure because linear regression finds only one coefficient value for each input variable. This model outperforms a simple baseline that predicts last week's points as next week's points and all players' average points as individual player's point predictions. The reason for using Linear Regression is that the function is based on a linear representation form. Based on this formula, linear programming tries to find the sweet spot. So, you can get the best results.

Taking a Neural Network with the desirable data fed to it would predict an output where players with higher skill and prowess standout and their probability of scoring a goal or stopping a penalty or providing an assist standout. The Neural Network architecture is destined to contain hidden layers in which the likeliness of high rated players like Mohammad Salah to score a goal, Kevin De Bruyne to make an assist or even Dean Henderson to make a crucial save. This concept can be compared to the same model YouTube uses to suggest food videos to me as I am more likely to watch food videos than videos on automobiles. These are learnt from previous examples of how often the above-mentioned players score, assist or stop goals and then fed into the network to provide results of highly desirable picks for the fantasy team.

Fantasy Sports rely on a lot of historical data and finding the suitable model to formulate the data we have in hand. Robinson, Chelsea *et al* [5] studied player careers who played majority games in the 2019 season in the National Football League (NFL). Based on player requirement and statistics, this model showed important data pertaining to particular positions and their dominance in that position. The model ARIMA, short for 'Auto Regressive Integrated Moving Average', is a forecasting algorithm; It is based on the idea that future values can be predicted using only information from the historical values of a time series. There was nothing in this model that precluded injuries from performance, and if there was anything in this model that involved injuries, this project would have been better, but that's hard to predict.

Past performance is an uncertain indicator of future performance. It is known that performance depends on game difficulty. Sometimes teams play tough games away, sometimes home games are very easy. Player performances also differ accordingly and hence their points. For Example, Bernardo Silva usually played the entirety of the game making several crucial passes and being an active member in the squad, comparing him to Jorginho who was a defensive midfielder having fewer crucial passes and lesser game time but because he was the penalty taker for Chelsea racked up more points than the former. This shows sometimes data we are considering to compare two entities are not enough.

Sertalp B. Cay et *al* proposed an Expected Value (EV) model to predict the points to be scored and a detailed distribution of the statistical data provided by the user [6]. A scale is used here ranging from 1-5 scoring team fixture difficulty based on data available from FPLReview and DraftHound. Wee Rogue uses a similar approach to generate "predicted goals conceded/scored per game against an average Premier League opponent" rather than against any average league opponent. This model is better suited for distinguishing between good and bad fixtures for FPL purposes than using a 1-5 rating.

When making decisions in FPL, we want to consider both the value and the likelihood of the consequences the decision may have. Each decision has a set of possible outcomes. Each result has two properties of interest. A specific probability of occurring and a specific value if it does occur. We need a way to combine the value and likelihood of outcomes to determine what decisions to make. Enter the expected value (EV), which is the cornerstone of game theory. In other words, the EV of a decision is the probability-weighted average of its resulting values.

Akhil Gupta [7] proposed a collaborative process consisting of three phases, Data Pre-processing, Time Series Modelling and Linear Optimisation with the dataset of points from the previous three seasons.

As Han noted [8], today's real-world databases are typically huge in size and likely come from multiple disparate sources, resulting in noisy, missing, and inconsistent data. Poor quality data leads to poor mining results. When one faces such an issue the most common thing the other would suggest is to undergo data pre-processing, i.e., matching the consistency of rows, missing string fields replaced by '-' as a string liberal and necessary blank IDs to be created for matching of datasets.

The aggregated data after the above-mentioned stage were tried by two different models separately their result cast together to form a hybrid version:

1) ARIMA, or Autoregressive Integrated Moving Average.
2) LSTM-RNN, or Long Short-Term Memory Recurrent Neural Networks.

Each one is designed to compensate for the others lacking. Time series data for each player was used for forecasting the next year. ARIMA and LSTM-RNN were run separately and their results mixed in proportion. These 38 gameweek points were added together to determine each player's point total for the following season.

The final stage Linear Optimisation seeks to achieve the best mathematical model whose pre-requisites are represented by linear relationships. The objective function is maximized or minimized with respect to some

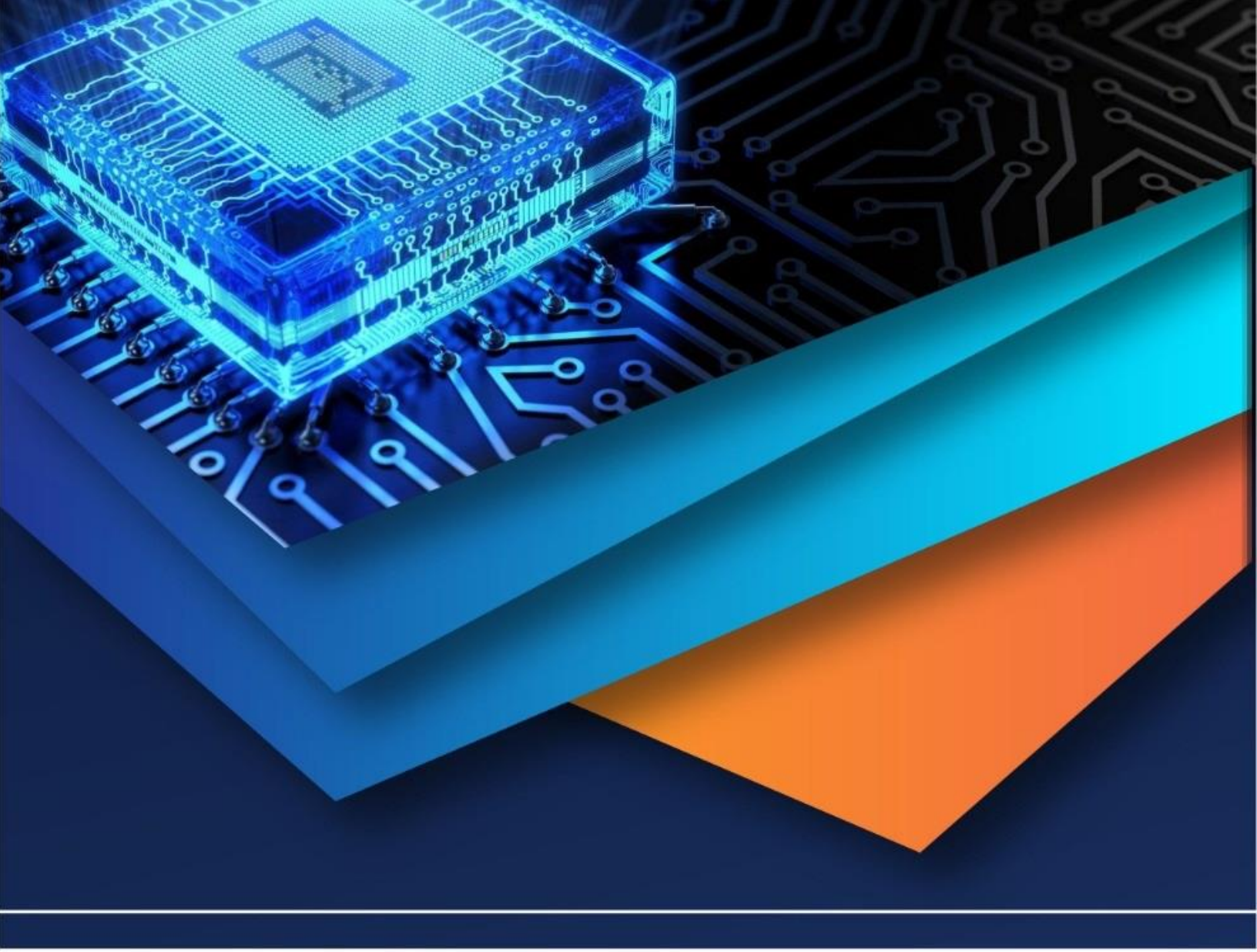
Circumstance-dependent constraints. This optimization approach contributed significantly to the reduction of post-time series forecasting effort. To formulate the objective function and constraints, we need to understand the problem correctly. Once that's done, it's just a matter of inputs and outputs.

VI. CONCLUSION

Fantasy Premier League is a game with increasing popularity, just like any other fantasy sport and is not subject to die down as long as the sport is alive. There has been an average increase of a close to a million users every year with more than 9 million users in the present time. There is a huge scope in terms of finding the best team and maximise the points one scores through the week; The choices in the game being purely data driven or luck, an individual prefers hard convincing to make this choice. This paper provides a brief if not detailed view of approaches by various researchers and research groups on how Data Science and Mathematics can be the convincing factor and play the major role required to your fantasy sport success. Particular attention has been given to models researchers have used and their advantages that helped them achieve their success in the fantasy sports world. The same models can well be used for other fantasy sports in Cricket, Basketball, American Football and so on. This paper can pave way for future developments and blooming ideas in this field. Either as the source or the path leading to the source.

REFERENCES

[1] Zegura, Nathan and Augustyn, Adam. "Fantasy sport". Encyclopedia Britannica, 25 May. 2016, https://www.britannica.com/sports/fantasy-sport. Accessed 30 August 2022.
[2] https://www.premierleague.com/news/1252542
[3] James Asher. “How to Use Data to do better in Fantasy Football” https://medium.com/the-sports-scientist/how-i-used-data-science-to-get-into-the-top-1-on-the-return-to-fantasy-premier-league-98829d4f65e5
[4] O’Brien JD, Gleeson JP, O’Sullivan DJP (2021) Identification of skill in an online game: The case of Fantasy Premier League. PLoS ONE 16(3): e0246698. https://doi.org/10.1371/journal.pone.0246698
[5] Robinson, Chelsea, "“The Prediction of Fantasy Football”" (2020). Mathematics Senior Capstone Papers. 20. https://digitalcommons.latech.edu/mathematics-senior-capstone-papers/20
[6] A Friendly Introduction to FPL Analytics by Sertalp B. Cay https://alpscode.com/blog/intro-to-fpl-analytics/
[7] Gupta, Akhil. "Time series modeling for dream team in fantasy premier league." arXiv preprint arXiv:1909.12938 (2019). https://doi.org/10.48550/arXiv.1909.12938
[8] Han J, Kamber M, Pei J. Data Preprocessing. Data Mining Concepts and Techniques 2011; 83-124: Morgan Kauffman Publications



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY