Fantasy Premier League

A Bayesian Network to help Fantasy Premier League managers predict the number of points a selected player will score in their upcoming fixture.

Dominic Ngoetjana  
 Department of Computer Science  
 University of Cape Town  
[ngwkga001@myuct.ac.za](mailto:ngwkga001@myuct.ac.za)

Mohamed Davidson  
 Department of Computer Science University of Cape Town  
[dvdmoh003@myuct.ac.za](mailto:dvdmoh003@myuct.ac.za)

Muhammad Hassen  
 Department of Computer Science  
 University of Cape Town  
[hssmuh010@myuct.ac.za](mailto:hssmuh010@myuct.ac.za)

Nkosi Gumede  
 Department of Computer Science  
University of Cape Town  
[gmdnko003@myuct.ac.za](mailto:gmdnko003@myuct.ac.za)

ABSTRACT

This paper presents a solution to the problem of deciding whether to have a specified player in a Fantasy Premier League squad. We use a public and freely available dataset to create a Bayesian Network given the variables we assume correlate best with a player’s expected performance. The Bayesian Network will predict the number of points that a player will obtain in their upcoming fixture. It will be evaluated continuously after every set of fixtures. In a world of 6.8 million (and counting) Fantasy Premier League managers with the same objective, the ability to accurately predict which players maximize one's chances of success is becoming increasingly valuable. This paper justifies the design of the most suitable Bayesian Network given the uncertainty of the problem at hand.

CCS CONCEPTS

• Artificial Intelligence • Bayesian Networks

KEYWORDS

Bayesian network, Fantasy Premier League, points prediction game

1. INTRODUCTION

Background and objective

Fantasy Premier League (FPL) is a web-based points prediction game available at [fantasy.premierleague.com](http://fantasy.premierleague.com). The original game is said to have been created by an Englishman named Bernie Donnelly back in 1971 [1]. The plot of the game is to assume the role of an English Premier League (EPL) manager by picking a squad of fifteen players (of varying prices) given a budget of €100 million. There are several rules to adhere to when playing FPL [2]. The main objective is to pick a squad of fifteen players who you predict will score you the highest total number of points over the course of the EPL season. Within the constraints of these rules, each manager battles against millions of other players worldwide. The manager who scores the highest total number of points wins.

The objective of this project is to build a model using a Bayesian Network that can predict the number of points that a player will score in the upcoming game week. This model will then be used by players to determine which players they want to select in their squad for the upcoming game week with the goal of maximizing their points earned from all the players in their squad.

Potential user community

Since 1971, Fantasy Premier League has grown tremendously in popularly. This is primarily due to the emergence of the internet and the increasing popularity of the EPL worldwide. Every week, millions of football fans congregate across the world to watch their favourite players play for their respective top-tier English clubs. Fantasy Premier League draws inspiration from fans’ fascination with the EPL. Statistics are used as a foundation for determining everything in FPL. Besides the global league in which managers compete by default, there exist other public and private leagues which are free to create and join. Private leagues encourage FPL managers to invite their friends to join FPL and compete. This is a form of word of mouth marketing. The FPL also markets itself via a television show called the FPL Show. FPL awards the top weekly, monthly and overall manager(s) with a hamper of prizes [2]. A combination of these factors led to the current total of 6.8 million FPL managers worldwide. All FPL managers play the game with the same objective and set of constraints. What separates the top-performing managers from the rest is their ability to analyze all players efficiently and predict which players to include in their squads with a high degree of accuracy. The potential user community of our Bayesian Network is the subset of managers who prefer using statistical techniques to justify their selection of players.

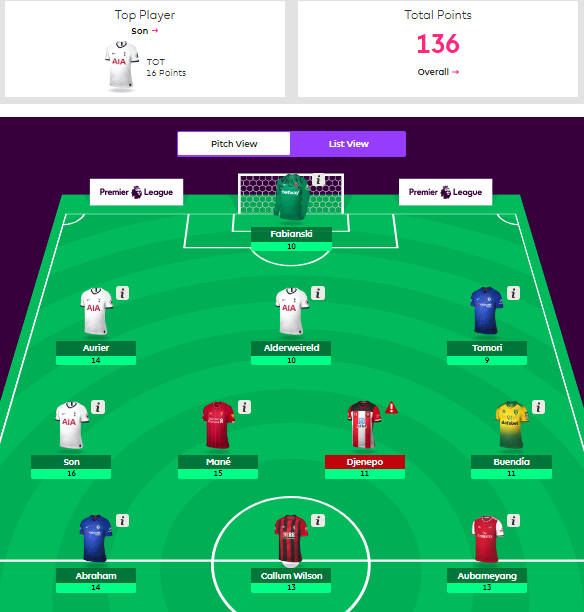


Figure 1: A screenshot of the highest scoring players in a game week.

1. PROBLEM ANALYSIS

The goal with our network is to assist an FPL manager with determining whether to retain or sell a player in the squad based on how the player is expected to perform in the next fixture, i.e., the points that the player is expected to earn. In this section, we’ll consider some of the potential factors that could have an influence on our model design and the decisions around that.

Problem statement

Our targeted problem considers the usefulness of the network, which is the main factor to consider. As previously mentioned, there are over 6.8 million FPL managers worldwide, all vying for the same prizes [2]. This healthy reserve of active and competitive target users makes our network viable and relevant to many. Our system can help managers maintain a squad of players with the highest expectancy of point acquisition and further maximize their performance in the FPL against other players.

Source data

Our data is sourced from [3], “An FPL library that gets all the basic stats for each player, gameweek-specific data for each player and season history of each player”. This means that we’re able to generate up-to-date player stats for each player at any time and for several seasons, across several years. There are thousands of entries to choose from, so there are no concerns for density. The data is filtered through and balanced out to ensure it’s not skewed towards a target (data is unbiased). We’ve also ensured that all the data used in our network is entirely relevant and applicable to the model used.

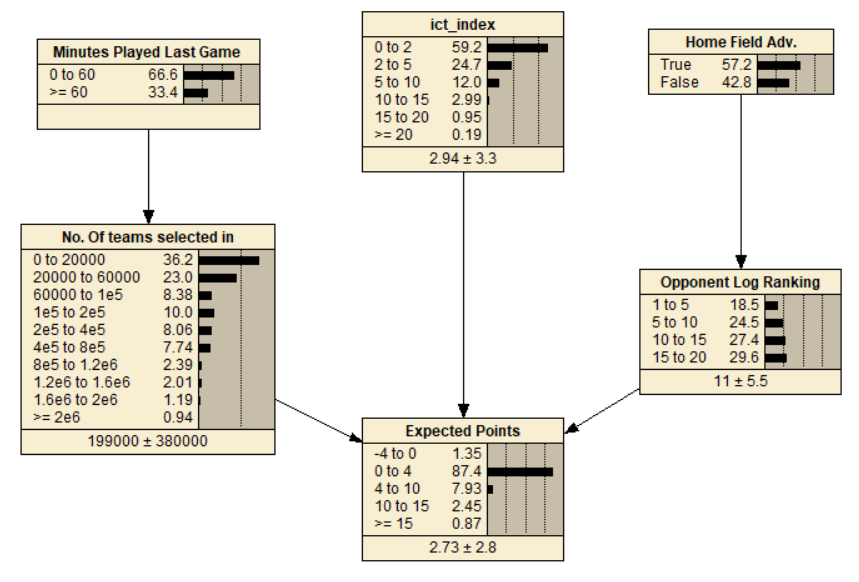
Model design

The model went through various iterations before ending up with an optimal one. The first iteration was merely adding all applicable states (variables) that could affect the expected score of a player to the model. The next iteration involved removing unnecessary variables (after enough deliberation and justification) and then drawing relative links between the variable nodes. The overall model design follows a predictive reasoning approach [5]. Once designed on Netica [4], the model was evaluated with Netica to determine variable sensitivity. We made sure to keep the model simple enough to comprehend, but also complex enough to reliably output a proper expected score, as well as ensure there aren’t too many/few states and links. The model is tested with recent past data from handpicked fixture results.

Usability

Our network is designed to be re-usable, in that it can be used by the same manager across various players in the squad and independently from the others. This then feeds into the manager’s discretion with comparatively evaluating players based on their respective expected points. The network is consistent with its predictions and can reproduce its result for the evidence provided.

1. DECISION NETWORK MODEL



**Figure 2:** *The final design of our Bayesian Network to predict expected points*

The variables were chosen to be used in the model to predict the target variable (Expected Points):

* ICT Index
* Minutes played in the last game
* Home field advantage
* The ranking of the opponent on the log
* The number of FPL teams the player has been chosen for

These variables were chosen based on expert knowledge from FPL experts as well as by identifying variables in the data that have a correlation to earning points.

ICT index is a feature in FPL that offers insight to help guide a player’s strategies and selections [6]. ICT index was determined to be the most important variable for predicting a player’s performance and possible points earned because FPL experts say that it was developed specifically to measure a player as an FPL asset as it offers a view on player performance factors that are known to produce FPL points [6]. The ICT index is a metric calculated by FPL using the player’s current level of influence, creativity and threat in a game. Therefore, ICT index directly affects the target variable.

Minutes played in the last game is important because if a player doesn’t play long enough, they won’t earn points and the longer a player plays, the more time they will have to earn points. Some players tend to not play full games due to fitness levels or because of a decision by the manager. Thus, having less time to earn points.

The number of FPL teams the player has been chosen for was found to correlate with the number of points earned. Through analyzing the data, it was found that players that have been selected for more teams tend to earn more points. This is because FPL managers only pick players who are expected to score high points. Thus, if a player is chosen by many FPL managers it means that the player is expected to score high points by many managers. Since minutes played affects the amount of time a player has to earn points, players who are more likely to play a full game will be chosen and thus have a higher number of FPL managers choosing them.

Teams that play in their hometown are said to have a home-field advantage due to having majority of the stadium being their supporters. This increases the players’ morale and confidence, resulting in better performance. The home team also has the advantage of knowing the field and environment better than the opponent as they train there.

The ranking of the opponent was chosen as it will influence how difficult the game will be which would affect the difficulty of earning points. If the opponent is a high-ranking team it is expected that it would be more difficult for a player to earn points against them than against a lower ranking team. Although this is a measurement of a team, it will still influence an individual player’s performance.

1. MODEL TESTING AND EVALUATION

The model was trained on the 4 game weeks of the current, 2019 Premier League data. The weeks chosen were from the 2nd week up to and including the 5th week. The model was then tested on the 6th, latest game week of the premier league. The results of the test on the 6th week is reflected in the confusion matrix below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Predicted** |  |  |  |
| **-4 to 0** | **0 to 4** | **4 to 10** | **10 to 15** | **15+** | **Actual** |
| 0 | 3 | 0 | 0 | 0 | -4 to 0 |
| 0 | 446 | 19 | 2 | 1 | 0 to 4 |
| 0 | 44 | 7 | 4 | 1 | 4 to 10 |
| 0 | 4 | 4 | 1 | 1 | 10 to 15 |
| 1 | 1 | 0 | 0 | 0 | 15+ |

**Table 1:** Table showing the results produced by our Bayesian Network.

The results from the matrix appear to be good. The “0 to 4” class contains the most classes, the Bayesian Network was able to accurately predict 446 of the 468 instances of this class in the 6th week. Although this appears to be a good result, there is a clear class imbalance that is reflected in both the training game weeks and the testing game week. Most players score between 0 and 4 points in a game week. Since the model was trained on this imbalanced data, the model is biased to predict that a player will score within this range. This is further reflected in the models’ predictions of other categories; in the 4-10 range, the model incorrectly predicts 44 players would earn between 0 to 4 points, whereas it only correctly predicted 7 of the 56 players in week 6 who scored between 4-10 points.

The model is effective in predicting what players would score, but this is only true because most players score between 0 to 4 points. The objective of this model was to predict the points a player would score. It appears that this is not the difficult part of the prediction. Predicting players who lie outside of this category correctly is a difficult task. Furthermore, considering this outcome, a new possible objective should be to predict players in each game week who would score points from 4 upwards.

1. CONCLUSION

Having evaluated the model, we can now explain how the model is useful to fantasy premier league users. Before the upcoming game week, users will be able to input the features mentioned in the previous section. The model will then return a prediction of the points a player might score in the upcoming game week. If the model predicts that the player will score between –4 to 0, the user will not want to select the player to be in their fantasy premier league team. If the model predicts that the player will score above 0 points, the user may be enticed to select that player for their team.

1. REFERENCES
2. Wikipedia. *Fantasy Football (association)*. <https://en.wikipedia.org/wiki/Fantasy_football_(association)>
3. Fantasy Premier League. 2019. *Fantasy*. Available from: [https://fantasy.premierleague.com](https://fantasy.premierleague.com/)
4. Fantasy Premier League Library. 2019. Available from: <https://github.com/vaastav/Fantasy-Premier-League>
5. Netica Application. 2019. Available at: <https://www.norsys.com/netica.html>
6. Kevin B. Korb and Ann E. Nicholson. 2010. Bayesian Artificial Intelligence, Second Edition (2nd ed.). CRC Press, Inc., Boca Raton, FL, USA
7. Fantasy Premier League. 2018. *How the ICT Index in Fantasy works.* Available from: <https://www.premierleague.com/news/65567>