ISyE 323 Operations Research - Deterministic Modeling

Creating the Perfect Fantasy Football Team Through Optimization

Group: Anjali Sheth, Manavv Shah

Table of Contents

1. Executive Summary

Fantasy premier league is a league where a person gets to pick a squad of 15 players from the premier league using the allocated budget of 100 million euros. The main aim of any person taking part in a fantasy premier league is to try and maximize their total points by making a decision of which players to include in their squad. The points calculated for each player in the premier league is based on many factors, like the number of goals they score, number of assists they make, number of red cards or yellow cards they get, and the number of minutes they play. For any premier league fan, fantasy premier league is an exciting way to make use of their knowledge and intuitive ability to compete against. Since there are 632 players in the premier league, it is hard to manually check all possible combinations of squads that maximise points and fit within the allocated budget. Hence, we have used optimization using the data from the first 9 gameweeks of this season to help any person competing in the fantasy premier league to select the ideal team for the next few gameweeks that will be played this season.

The optimization model consists of a decision variable for each of the 632 players to decide if they will be in the fantasy premier league squad. The aim is to maximise the sum of points scored by each player in the squad. Fantasy premier league has a few rules that we have added as constraints in our model in order to make sure that the team created complies with these rules. The constraints include staying within the allocated budget, meeting the position requirements according to the selected formation, and selecting a maximum of 3 players from any of the 20 teams that take part in the premier league.

We used the JuMP package on Julia, to run our optimization code which eventually gave us a squad of 11 on field players that includes the team - Sá, Alexander-Arnold, Cancelo, Van Dijk, Rüdiger, Salah, Son, Bernado, Gallager, Vardy, Dennis. This is the team that we suggest our client picks in the future game weeks to gain an edge in points in comparison to their competitors. We have made a few assumptions that were difficult to model and have not considered a few factors that are hard to predict due to being highly uncertain, and this may cause the ideal team to change in the near future. We must keep updating factors after each gameweek to get the most accurate team possible.

2. Problem Statement

We are creating an optimization model that helps our client pick their fantasy premier league team in a way that maximizes their points within the allocated budget. With 632 players in the premier league, each with their individually associated cost and points, it is hard to manually analyze all options while complying with the fantasy premier league rules. This model is intended to help our client pick the team after using data of the first 9 game weeks in the 2021-2022 season.

For each of the 632 players, we will have a decision variable that decides if the player is in the team or not. A total of exactly 11 out of the 632 players should have a decision of yes. Generally, the fantasy premier league gives a budget of 100 million Euros and makes you pick a team of 15 players, where 11 players will be on the field and 4 will be substitutes. In our model, we will be disregarding substitutes as it is extremely uncertain to predict if the substitutes will play or not. On average, each player in the fantasy premier league costs 4.5 million Euros and hence, the 4 substitutes would have cost around 18 million Euros in total. We will deduct this amount from our 100 million Euros budget, and hence use only 82 million Euros while selecting the 11 players that will be on field.

Each of the 632 players has an associated cost, points they have scored up till now, the team they are playing for, and the position they play at. The 11 players should be distributed in a way that there is exactly 1 goalkeeper, 4 defenders, 4 midfielders, and 2 forwards. Another rule to keep in mind is that only a maximum of 3 players can be selected from one team out of the 20 teams that play in the premier league.

A lot of factors cannot be included in our model due to the uncertainty associated with it, and hence previous data will not be helpful in making any decisions. We will assume that all the players have played an equal amount of time in the last 9 game weeks and hence their points are used in a fair manner. All players that are selected in the team will play for the entire 90 minutes assuming that they will not be injured or will not test positive for covid-19 during the remaining season. A few other assumptions that we will make are that the team the player plays, position the player plays, and cost of the player will stay the same throughout the season and it will not change mid-way through the season.

3. Model Description

We have used data from the official fantasy premier league website of this 2021-2022 (Final dataset: https://fantasy.premierleague.com/player-list/) season that has taken place up till now. The data consists of the list of all the players in the 20 teams that have taken part in the premier league this season. Each player has an associated cost, total points scored up till now, the team they are currently playing for, and the position they play at.

We have 1 binary decision variable in our model. The first decision variable (Xi) is for the selection of the 11 players. It equals to 1 if the player is selected in the fantasy team, and equals to 0 if the player is not selected in the fantasy team.

$$Set\ I\ =\ \{Mendy,\ Allison,\ Ederson,\ \dots\ ,\ Ferguson\}$$
$$X_i = \{1,\ if\ player\ i \in I\ is\ chosen\ for\ the\ team;\ 0\ otherwise\}$$

Additionally, we introduced some parameters for cost, points and position to help us with our model:

$$cost_i\ =\ \{6.3, 6.3, 6.5,\ \dots\ , 6.7\}$$
$$points_i\ =\ \{68, 72, 69,\ \dots\ , 0\}$$
$$position_i\ =\ \{GK, GK,\ \dots\ , FWD\}$$

The objective of our optimization model is to maximize the points scored by the selected fantasy team. Hence, this will equal to the first selection decision variable multiplied by the points of each of the players and the second captain variable multiplied by the points of each player. Hence, if the player is selected, Xi will equal to 1 which means that their points will be added to the sum of points. Only the 11 players for which Xi equals to 1, will points be added, and for others as Xi will be 0, their points will not be added to the sum.

Objective Function Formula:

$$max\ \Sigma\ (x_i)\ \times\ (points_i)\ \ \forall\ i \in I$$

The constraints of the optimization model take into consideration the rule that the fantasy premier league imposes on the people that take part in the league. We have a total of 7 constraints to model those rules:

1) Budget: The total budget is 82 million euros. We will multiply the first selection decision variable with the cost of each player and sum up this value. Hence, if the player is selected, Xi equals to 1 which means that their points will be added to the total cost, and for others as Xi will be 0, their points will not be added to the total cost. Xi multiplied by each player's cost will give us the total cost and this should be less than or equal to 82 million euros.

Budget Equation:

$$\Sigma\,(x_i) \times (cost_i) \leq 82 \quad \forall\, i \in I$$

2) Total players: The total number of players that can be selected are 11. For this we will take a sum of the first selection decision variable. Hence, if the player is selected, Xi equals to 1 which means Xi can be equal to 1 a maximum of 11 times.

Total players Equation:

$$\Sigma\,(x_i) = 11 \quad \forall\, i \in I$$

3) Maximum of 3 players selected from 1 team: 20 total teams take part in the premier league, and for each team we will check that not more than 3 players from that team are selected. This means that we will have 20 iterations in this constraint. For each iteration, we will take a sum of the first selection decision variable if the team of that player equals the team in that iteration. If the if statement is true, and the first decision variable equals to 1, it will be added to the sum. This will give us a total count of selected players from that team, which should be less than or equal to 3. We will repeat this iteration for the other 19 teams as well.

$$Set\ uniqueTeams = \{Chelsea,\ Arsenal,\ \dots\ ,\ Liverpool\}$$

Max players from each team Equation:

$$team_i = uniqueTeams_j \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i \leq 3$$

$$\forall\, i \in I, \forall\, j \in uniqueTeams$$

4) Exactly 1 Goalkeeper: We will take a sum of the first selection decision variable if the position of that player equals Goalkeeper. If the if statement is true, and the first decision variable equals to 1, it will be added to the sum. This will give us a total count of the goalkeepers in the team, which should be equal to 1.

Exactly 1 Goalkeeper Equation:

$$position_i = \text{"}GK\text{"} \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 1 \quad \forall\, i \in I$$

5) Exactly 4 Defenders: We will take a sum of the first selection decision variable if the position of that player equals Defender. If the if statement is true, and the first decision variable equals to 1, it will be added to the sum. This will give us a total count of the defenders in the team, which should be equal to 4.

Exactly 4 Defender Equation:

$$position_i = \text{"}DEF\text{"} \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 4 \quad \forall\, i \in I$$

6) Exactly 4 Midfielders: We will take a sum of the first selection decision variable if the position of that player equals Midfielder. If the if statement is true, and the first decision variable equals to 1, it will be added to the sum. This will give us a total count of the midfielders in the team, which should be equal to 4.

Exactly 4 Midfielder Equation:

$$position_i = \text{"}MID\text{"} \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 4 \quad \forall\, i \in I$$

7) Exactly 2 Forwards: We will take a sum of the first selection decision variable if the position of that player equals Forward. If the if statement is true, and the first decision

variable equals to 1, it will be added to the sum. This will give us a total count of the forwards in the team, which should be equal to 2.

Exactly 2 Forwards Equation:

$$position_i = "FWD" \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 2 \quad \forall\, i \in I$$

8) Sign Constraints:

$$x_i \in \{0, 1\} \quad \forall\, i \in I\,, \qquad uniqueTeams_j \geq 0 \ \forall\, j \in uniqueTeams$$

4. Analysis and Results

After running our optimization model on Julia using the JuMP package, we got an objective value of 1022, and a list of 11 players for who the decision variables equal to 1 which represented that those 11 should be added in the fantasy premier league for a person who wants to win the league and get the highest possible points compared to others. The objective value of 1022 is the sum of points of the selected 11 players in the team. The 11 players include 1 goalkeeper, 4 defenders, 4 midfielders, and 2 forwards. These players have proven to be efficient together as they score high points and stay within the budget provided. This team has been formed according to all the player's track record since the last 9 gameweeks that have been played this season.

We suggest that the client chooses this as their base team for the future gameweeks by selecting these 11 players while creating their FPL team. However, there are a lot of uncertain factors in the future, like injury of a selected player, a selected player having to play against tougher teams in the future gameweeks, or just lost form of a selected player that can cause these results to change in the future. On the other hand, a non-selected player having to play against easier teams in the future game weeks or gaining form of a non-selected player can also cause more optimal results than the one suggested currently in the future. Hence, while we can tell our client to choose these 11 players in the team right now, we would ask them to use a little bit of

their knowledge and intuitive ability to make slight changes to this base team as each gameweek passes.

While we have tried to include all possible constraints through our data, there are a few factors that cannot be modeled currently. The cost of each FPL player varies slightly due to demand fluctuations, and hence it can be that the team that we suggest right now does not fall into the required budget in the near future. For this, we might have to run our optimization model on Julia every gameweek to get a team for the client that maximizes the points while staying within the budget. Additionally, there are a lot of uncertain factors that limit the model. It is uncertain if a player from our current team will be injured during that gameweek, and if a substitute player will have to replace him. We can add this as a future addition to our model, by taking into account the average number of times that players get injured and are replaced by substitutes during the game. It also often happens that without an injury an on field player is replaced by a substitute and we should check how often this happens based on previous data. Hence, in order to make our model more efficient and accurate, we must include information about substitutes and further model a team of 15 players with 11 on field players and 4 substitute players. This will also let our model have a budget constraint of the total 100 million euros, and we will be able to disregard our assumption of substitutes costing around 18 million euros on average.

4.1 What-if Analysis

Since the optimization model is based on the points of each player that is in the fantasy team, the number of players in each position can make a difference in the points as well. For example, in our base formation we follow 1-4-4-2 (Goalkeeper, Defenders, Midfielders, Forwards) which gives us the total points for that formation as 1022. However, changing the formation will result in some players being removed and new players being added as they may not satisfy the constraints. To fully encapsulate these various scenarios and check how they affect the points we performed a 'What-If Analysis' on a few modern formations that are commonly used.

I. Scenario 1: Formation 1-4-4-2

In this formation we get the team - Sá, Alexander-Arnold, Cancelo, van Dijk, Rüdiger, Salah, Son, Bernado, Gallager, Vardy, Dennis for a total point value of 1022 and total cost of 71.84 million euros.

II. Scenario 2: Formation 1-4-3-3

In this formation we get the team - Sá, Alexander-Arnold, Cancelo, van Dijk, Rüdiger, Salah, Bernado, Gallager, Vardy, Antonio, Dennis for a total point value of 1016 and total cost of 72.118 million euros.

III. Scenario 3: Formation 1-5-3-2

In this formation we get the team - Sá, Alexander-Arnold, Cancelo, van Dijk, James, Rüdiger, Salah, Bernado, Gallager, Vardy, Dennis for a total point value of 1017 and total cost of 72.377 million euros.

IV. Scenario 4: Formation 1-3-4-3

In this formation we get the team - Sá, Alexander-Arnold, Cancelo, van Dijk, Salah, Son, Bernado, Gallager, Vardy, Antonio, Dennis for a total point value of 1016 and total cost of 71.575 million euros.

V. Scenario 5: Formation 1-3-6-1

In this formation we get the team - Sá, Alexander-Arnold, Cancelo, Rüdiger, Salah, Mané, Son, Bernado, Mount, Gallager, Dennis for a total point value of 1021 and total cost of 71.061 million euros.

4.2 Results

For the what-if analysis in section 3.1, we can dive deeper into picking the best formation for your fantasy football team. For the formation in Scenario 1, we get a total points of 1022 with a cost of 71.84 million euros. This results in savings of 82 - 71.84 = 10.16 million euros. Any

savings in our model means that we get more of our budget to spend behind substitutes, which helps if any of the 11 on field players get injured or do not play for the entire 90 minutes. Since, this is the first scenario we can assume this as the best scenario for now. Next, Scenario 2 gives us the total points of 1016 with a cost 72.118. This results in savings of 82 - 72.118 = 9.882. We can conclude that this scenario is not ideal since we are losing some points and spending more. Therefore, Scenario 1 still remains as the best scenario. Next, Scenario 3 gives us total points of 1017 and a cost of 72.377 million euro, leaving us with 82 - 72.377 = 9.623 million euros in savings. We can yet again conclude that Scenario 1 stays as the best formation however, Scenario 3 proves to be better than Scenario 2 since we have a +1 increase in points for a minimal cost. Next, Scenario 4 gives a total points value of 1016 at a cost of 71.575 million euros which results in a savings of 82 - 71.575 = 10.425 million euros. Again, Scenario 1 stays the top choice; however, Scenario 4 is better than Scenario 2 but does not prove to be better than Scenario 3. This is because points are the most important factor and between Scenario 2, 3 and 4 - Scenario 3 has the highest points for a very small difference in cost and Scenario 2 and 4 provide the same amount of points but Scenario 4 is less expensive than Scenario 2. Finally, Scenario 5 gives us a total point value of 1021 with a cost of 71.061 million euros, leading to savings of 82 - 71.061 = 10.939. Scenario 5 comes close to Scenario 1, however, it still falls one point short of being tied with Scenario 1. Given the points of Scenario 2, it proves to be the second best option. We can conclude the rankings of the different formations as:

(1-4-4-2) > (1-4-3-3) > (1-5-3-2) > (1-3-4-3) > (1-4-3-3)

Additionally, from our analysis we can see that there are some players who appear on all the formations and are considered to be best in the league for now. Those players are Sá, Alexander-Arnold, Cancelo, van Dijk, Salah, Bernardo and Dennis. Therefore, it is advisable to have these players in your team no matter what to maximize the points. Furthermore, at the start of every season the budget is replenished, therefore it is advisable to use up all your budget as the savings are of no use in this model. Although we consider savings irrelevant in this model, a more complex optimization model could be implemented that follows all the rules of FPL which will utilize the savings that are accumulated.

5. Appendix

Team	Points	Cost	Position	Name
Chelsea	68	6.3	GK	Mendy
Liverpool	72	6.4	GK	Alisson
Man City	69	6.5	GK	Ederson
Aston Villa	67	6.6	GK	Martínez
Burnley	47	6.7	GK	Pope
Spurs	63	6.8	GK	Lloris
Man Utd	66	6.9	GK	de Gea
Wolves	77	6.1	GK	Sá
Arsenal	75	6.11	GK	Ramsdale
West Ham	64	6.12	GK	Fabianski
Wolves	0	6.13	GK	Patrício
Everton	42	6.14	GK	Pickford
Leeds	49	6.15	GK	Meslier
Leicester	44	6.16	GK	Schmeichel
Chelsea	6	6.17	GK	Arrizabalaga
Man Utd	0	6.18	GK	Henderson
West Ham	0	6.19	GK	Areola
Arsenal	4	6.2	GK	Leno
Brighton	53	6.21	GK	Sánchez
Crystal Palac	54	6.22	GK	Guaita
Brighton	0	6.23	GK	Ryan

A.1 View of dataset

$$Set\ I\ =\ \{Mendy,\ Allison,\ Ederson,\ \dots\ ,\ Ferguson\}$$

$$Set\ uniqueTeams\ =\ \{Chelsea,\ Arsenal,\ \dots\ ,\ Liverpool\}$$

$$X_i = \{1,\ if\ player\ i \epsilon I\ is\ chosen\ for\ the\ team;\ 0\ otherwise\}$$

$$cost_i\ =\ \{6.3, 6.3, 6.5,\ \dots\ , 6.7\}$$

$$points_i\ =\ \{68, 72, 69,\ \dots\ , 0\}$$

$$position_i\ =\ \{GK, GK,\ \dots\ , FWD\}$$

$$max\ \Sigma\ (x_i)\ \times\ (points_i)\ \forall\ i \epsilon I$$

Subject To: $\Sigma\ (x_i)\ \times\ (cost_i)\ \leq\ 82\ \forall\ i \epsilon I$

$$\Sigma\ (x_i)\ =\ 11\ \forall\ i \epsilon I$$

$$team_i = uniqueTeams_j \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i \leq 3 \quad \forall i \in I, \forall j \in uniqueTeams$$

$$position_i = "GK" \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 1 \quad \forall i \in I$$

$$position_i = "DEF" \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 4 \quad \forall i \in I$$

$$position_i = "MID" \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 4 \quad \forall i \in I$$

$$position_i = "FWD" \rightarrow \delta = 1;\ \delta = 1 \rightarrow \Sigma x_i = 2 \quad \forall i \in I$$

$$x_i \in \{0, 1\} \quad \forall i \in I$$

$$uniqueTeams_j \geq 0 \ \forall j \in uniqueTeams$$

A.2 Complete Program