Optimization in Team Selection process

Shreyansh Shah*, Devam Shah*,
*School of Engineering and Applied Science,Ahmedabad University, India
Email: *{shreyansh.s, devam.s1}@ahduni.edu.in,

I. Introduction

In today's modern era one may see optimization in every aspect of life. Same goes with different leagues(tournaments) in the sports industry. Very famous examples are the football leagues where players from different countries/clubs play under one team. In such cases one may have to choose the best team from the given players and under a certain budget. So the data regarding the player's past performance is given and based on which we may or may not include the player in the team. Other constraints like formation constraints, team constraints, captaincy also play an important role in the selection process. Thus, optimization of the selection process will benefit the sponsors in terms of cost and also increase the chances of winning the league.

One may also need to assume that we have perfect forecast of the players, substitutes are present in the team and no transfer of players is taking place.

II. Problem Description

As a manager of a specific team you have to optimize the team selection process by selecting the best players from the given pool of players from different clubs, In other words maximizing the overall points of the teams and minimizing the entire buying cost of the team. We have 1000 million dollars as out budget and cannot spend more than that.

The statistics regarding 647 player's career is given in the data set. Which includes type of player(defender, goalkeeper, etc.), number of goals scored and assisted, cost of the player, etc.

	first_name	second_name	goals_scored	assists	total_points
0	David	Ospina	0	0	19
1	Petr	Cech	0	0	124
2	Damian Emiliano	Martinez	0	0	0
3	Laurent	Koscielny	2	0	100
4	Per	Mertesacker	1	0	16
...	...	...	...	...	...
642	Jordan	Hugill	0	0	3
643	Patrice	Evra	0	0	2
644	Marcus	Browne	0	0	0
645	Josh	Pask	0	0	0
646	Grady	Diangana	0	0	0

647 rows × 20 columns

Fig. 1: Snipshot of the dataset

minutes	goals_conceded	creativity	influence	threat	bonus	bps
380	3	0.0	75.8	0	1	86
3039	48	0.0	722.4	0	7	627
0	0	0.0	0.0	0	0	0
2225	32	123.8	692.0	137	16	556
364	7	3.0	119.6	38	2	81
...	...	...	...	...	...	...
22	0	0.4	0.0	0	0	4
227	10	6.0	42.8	0	0	32
0	0	0.0	0.0	0	0	0
0	0	0.0	0.0	0	0	0
0	0	0.0	0.0	0	0	0

Fig. 2: Snipshot of the dataset

ict_index	clean_sheets	red_cards	yellow_cards	selected_by_percent	now_cost	element_type	team_code
7.7	2	0	0	0.2	48	1	3
71.9	11	0	1	4.5	53	1	3
0.0	0	0	0	0.6	40	1	3
95.5	9	0	4	1.5	60	2	3
16.1	0	0	0	0.5	47	2	3
...	...	...	...	...	...	...	...
0.0	0	0	0	0.0	49	4	21
5.0	0	0	0	0.2	41	2	21
0.0	0	0	0	0.0	45	3	21
0.0	0	0	0	0.0	40	2	21
0.0	0	0	0	0.0	45	3	21

Fig. 3: Snipshot of the dataset

The constraints here are:

- Formation constraints: The team needs different types of players like goalkeeper, defender, midfielder, forward. Thus, we use 2-5-5-3 criteria respectively for selecting different types. where 1-4-4-2 are the playing 11 and rest 4 players will be extra if anyone gets injured.
- Team constraints: It ensures that at most 3 players are selected in any team from the same club.
- Budget constraint: There is 100 million dollar of cash assigned for buying the players.

Now, the problem can be defined as a linear programming question where we need to use libraries to solve the problem, and get the optimised team.

III. Solution Approach

For solving the linear problem we decided to use the "simplex Algorithm", which is an efficient way to find the optimal solution because it does not calculate the value of objective

function at every iteration, but starts with edges(corner) of the feasible set where main variables are 0. Later, shifting from edge to edge, while improving the objective function's value at every shift. This process continues till the optimal solution is achieved.

So in this fashion we select the best 15 players from the given pool with taking cost and other constraints into consideration. where my objective function is to optimize is the total number of points scored by the final team.

Now, let's start by defining the objective function. The basic idea is the maximize the total points scored by the team and hence assuming p_i is the points scored by player i, and assuming x_i denotes whether the player i is selected or not, then the objective function can be defined as follows:

$$max \sum_{i}^{N} p_i * x_i \tag{1}$$

where, N is the total number of players. and the variable x_i is a binary variable which can take only two values i.e. x_i=0,1. If x_i=0, then that player isn't chosen and if x_i=1 it means that player is chosen.

Now, using the rules defined earlier, here we'll mathematically defined each constraint. First of all starting with the cash constraint, we have total 100 million dollar funds available to complete the team. Assuming c_i is the cost(in million dollar) of the player i, and x_i is the same binary variable which indicates whether the player is selected or not then, the cost constraint can be given as follows:

$$\sum_{i}^{N} c_i * x_i \tag{2}$$

Now, our base team is one goalkeeper, four defender and midfielder and two forward players and additionally we'll select one extra player for each category which will be the substitution players. So the in total we'll have to select two goalkeepers, five defender and midfielders and three forward players. Now, assuming x_i is the same binary variable which indicates whether the player is selected or not then, the cost constraint can be given as follows:
Goalkeeper constraint:

$$\sum_{i}^{A} x_i = 2 \tag{3}$$

Defender constraint:

$$\sum_{i}^{B} x_i = 5 \tag{4}$$

Midfielder constraint:

$$\sum_{i}^{C} x_i = 5 \tag{5}$$

Forward constraint:

$$\sum_{i}^{D} x_i = 2 \tag{6}$$

where, A,B,C and D are the total number of goalkeepers, defenders, midfielders and forward players respectively.

Now, the final remaining constraint is the team constraint which is there can be maximum three players selected in a team from the same club. So assuming T is the total number of clubs and for club i there are total N_i players and the variable x_j indicated that whether that player from that particular club is selected or not then, this constraint can be given as follows:

$$\sum_{i}^{T} \sum_{j}^{N_i} x_j <= 3 * T \tag{7}$$

IV. Outcome and Results

We have got the optimal(maximized) points in the given constraints with the name and other details of the 15 players in the team.

```
Status: Optimal
Optimal Solution to the problem:  2520.0
Individual decision_variables:
x128 = 1.0
x130 = 1.0
x175 = 1.0
x276 = 1.0
x310 = 1.0
x311 = 1.0
x327 = 1.0
x329 = 1.0
x396 = 1.0
x474 = 1.0
x511 = 1.0
x555 = 1.0
x595 = 1.0
x65 = 1.0
x81 = 1.0
```

Fig. 4: Index of the players and status(1.0)

The image above shows the status and the maximum points obtained from the given pool of players.

```
Selected players in the final team:
César Azpilicueta
Marcos Alonso
Luka Milivojevic
Jamie Vardy
Mohamed Salah
Roberto Firmino
Nicolás Otamendi
Raheem Sterling
Ayoze Pérez
Lukasz Fabianski
Ben Davies
Abdoulaye Doucouré
Ahmed El-Sayed Hegazi
Mathew Ryan
Pascal Groß
Total cost= 999
Expected Points= 2520
```

Fig. 5: Final optimal team

This is the final team of 15 with maximum points.

V. Conclusion

Using this code, we can optimally select the team which will be able to score more points while keeping several constraints in-line. Though, here due to the data set limitation, we had to

assume that all players will score same points scored in the previous season. Furthermore, to improve this optimization we can also include the transfer conditions which will make this problem more usable and concrete.