Optimize the Lineup for Fantasy English Premier League

by

Siva Ramakrishnan

submitted in fulfillment of the requirements for the Final Project

for BANA 7020 - Optimization

Professor:

Dr. David Rogers

Abstract

This optimization project is an attempt to find the best lineup for the official Fantasy English Premier League football. The best lineup will be based on two factors – the points scored by the player in the past and the current form of the player. The project attempts to select a team of 15 players using a Multi-objective model using the Goal Programming approach. Alternate Scenarios where objective was either to maximize the overall score or overall form were also explored.

Table of Contents

[List of Figures iv](#_Toc481058486)

[List of Tables v](#_Toc481058487)

[Introduction 1](#_Toc481058488)

[Rules of the Fantasy Premier League 1](#_Toc481058489)

[Data Collection 3](#_Toc481058490)

[Knapsack Models 4](#_Toc481058491)

[Model Formulation for Maximizing Score 4](#_Toc481058492)

[Solving the Multi-Dimensional Knapsack model 5](#_Toc481058493)

[OpenSolver 5](#_Toc481058494)

[GAM Model 6](#_Toc481058495)

[Goal Programming Model 7](#_Toc481058496)

[Model Formulation for Goal Programming 7](#_Toc481058497)

[Solving the Goal Programming model 8](#_Toc481058498)

[Conclusion 9](#_Toc481058499)

[Key Learnings 9](#_Toc481058500)

[Bibliography 9](#_Toc481058501)

[Appendix A – Players selected from the Goal Programming Model 10](#_Toc481058502)

List of Figures

[Figure 1 Webpage Screenshot 3](#_Toc481057104)

[Figure 2 Excel data screenshot 3](#_Toc481057105)

[Figure 3 Open Solver model 6](#_Toc481057106)

[Figure 4 OpenSolver Output 6](#_Toc481057107)

[Figure 5 Maximize Score GAMs output 7](#_Toc481057108)

[Figure 6 Maximize form GAMs output 7](#_Toc481057109)

[Figure 7 Goal Programming - OpenSolver 8](#_Toc481057110)

List of Tables

[Table 1: Point Allocation 2](#_Toc481057051)

[Table 2 Selected players from the model 10](#_Toc481057052)

# Introduction

The official premier league website hosts a Fantasy sport contest where participants can build a virtual team from a pool of players who are playing in the current season of the league with some constraints. These virtual teams accumulate score based on the performances of the selected players. Teams with the best overall score win the fantasy league.

The chances of a player scoring more points in the future games can be assumed to be proportional to both his existing points and the current form i.e. a player would have a better chance of getting more points in the upcoming games if his past track record (that is his existing points) is good, a player’s chance of getting more points will also be high if his current form is high.

Thus, this problem can be modelled as a linear programming where there are multiple objectives of maximizing both the cumulative score of the team and the cumulative form of the team.

The approach that was followed was Goal programming with goals of a high score and a high form close to their individual maximum values. Base multi-dimensional knapsack models were also developed to find the maximum possible overall score and the maximum possible overall form.

The project was completed in four phases

**Phase 1:** Study of Knapsack, Goal programming models and Model Formulation for the current project

**Phase 2:** Web Scraping and Data Cleanup in excel

**Phase 3:** Solving the two Knapsack models for maximizing the overall score and overall form of the team in excel using Open Solver and GAMS

**Phase 4:** Building the Final goal programming model using Open Solver

# Rules of the Fantasy Premier League

The premier league Consist of 20 member clubs. Each of these 20 clubs are a shareholder premier league.

There is an official Fantasy league on the Premier league website and each participant must follow certain rules while creating a virtual team. These rules are:

* The Team should consist of 15 players with:
  + 2 Goalkeepers
  + 5 Defenders
  + 5 Midfielders
  + 3 Forwards
* The total budget of the Squad must not exceed £100 million
* Total number of players selected from a single team should not exceed 3

The Premier league website also maintains statistics of all the players in the league including:

* Player’s Cost
* Player’s Current form (on a scale of 0 to 10)
* Points the player has accumulated so far.

Points are allocated to the players based on their performance in the league. Listed below are some of the rules taken from the website (Fantasy Premier League, n.d.):

Table 1 Point Allocation

|  |  |
| --- | --- |
| Action | Points |
| For playing up to 60 minutes | 1 |
| For playing 60 minutes or more (excluding injury 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 goal assist | 3 |
| For a clean sheet by a goalkeeper or defender | 4 |
| For a clean sheet by a midfielder | 1 |
| For every 3 shot saves by a goalkeeper | 1 |
| For each penalty save | 5 |
| For each penalty miss | -2 |
| Bonus points for the best players in a match | 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 |

The Premier league website maintains the overall score for each player based on the above rules.

# Data Collection

The Data was collected from the Statistics page of the premier league website (<https://fantasy.premierleague.com/a/statistics/total_points>). Each webpage was scrapped using a C# code. Indicator variables for each team and Player position were also created using the code and the entire dataset was than exported to an excel file.

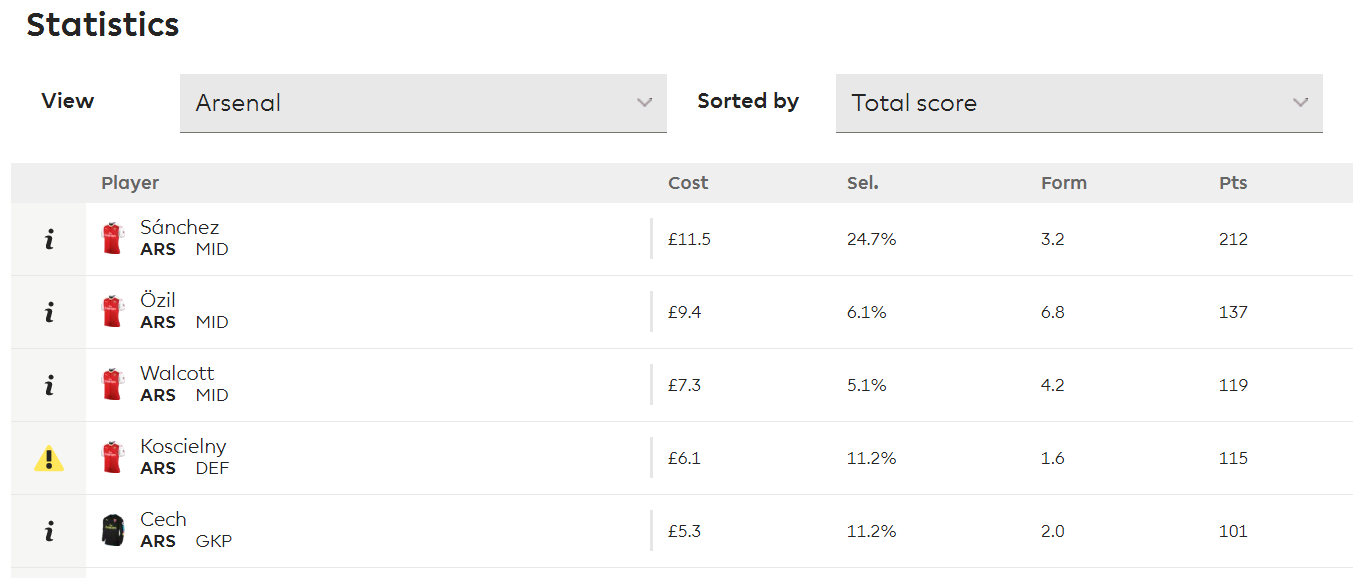


Figure 1 Webpage Screenshot

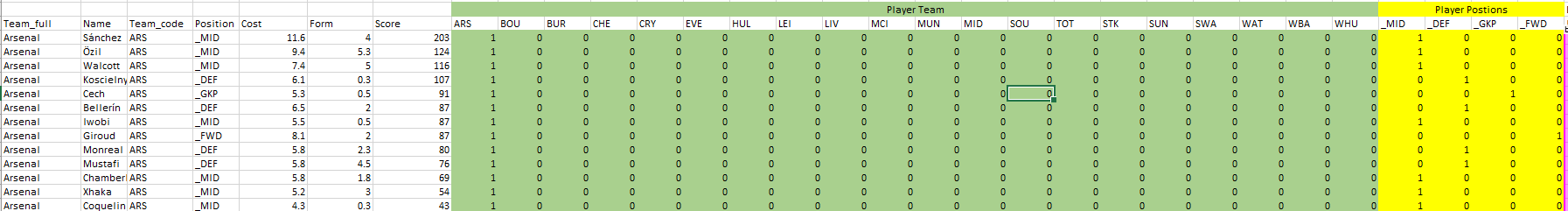


Figure 2 Excel data screenshot

Overall the complete data on 550 players from 20 different teams was collected.

# Knapsack Models

Knapsack models are one of the widely-studied optimization problems with a broad range of applications. The most common version of the knapsack problem is the 0-1 knapsack problem. The basic idea behind this problem is maximize the overall value obtained from selecting items from a set of n items subject to a single constraint that the weight of the selected items doesn’t exceed a certain limit.

This form of knapsack model can be extended to the case of multiple constraints. Such models are called Multi-Dimensional Knapsack Problems.

The aim of maximizing the overall score or overall form of the selected team can exactly be modelled using the Multi-Dimensional Knapsack problem.

## Model Formulation for Maximizing Score

The Objective function would be of the form:

Ma

Here be the binary variable for selecting the player i . It is 1 if player i is selected and 0 otherwise. i goes from 1 to N (N is the total number of players – 550 in this case)

– Points accumulated by player i.

The Constraints of the model are:

Total Players From a single team must be less than or equal to 3 (20 such constraints occur as there are 20 teams)

=> ARS is an indicator variable. It is 1 if belongs to team Arsenal and 0 otherwise

.

.

.

=> MTD is an indicator variable. It is 1 if belongs to team Manchester United and 0 otherwise

Player Position Constraints:

=> is an indicator variable. It is 1 if plays as a Forward and 0 otherwise

=> is an indicator variable. It is 1 if plays as a Midfielder and 0 otherwise

=> is an indicator variable. It is 1 if plays as a Defender and 0 otherwise

=> is an indicator variable. It is 1 if plays as a Goalkeeper and 0 otherwise

Budget Constraint (Total Budget is 100 million)

where => Cost of Player i (in millions)

The Same multidimensional knapsack formulation can be used for maximizing the form of the team. The only change would be that would represent the current form of the player instead of the Points accumulated.

## Solving the Multi-Dimensional Knapsack model

### OpenSolver

Maximizing the overall Score

The Standard Excel Solver plugin has a limit of 200 decision variables. Since there are more than 550 decision variables (one each for each player), excel solver could not be used. One of the alternatives of excel solver is the open-source OpenSolver which provides the same functionality but without any limits.

The Knapsack model was built for maximizing the overall points of the selected players in the open solver

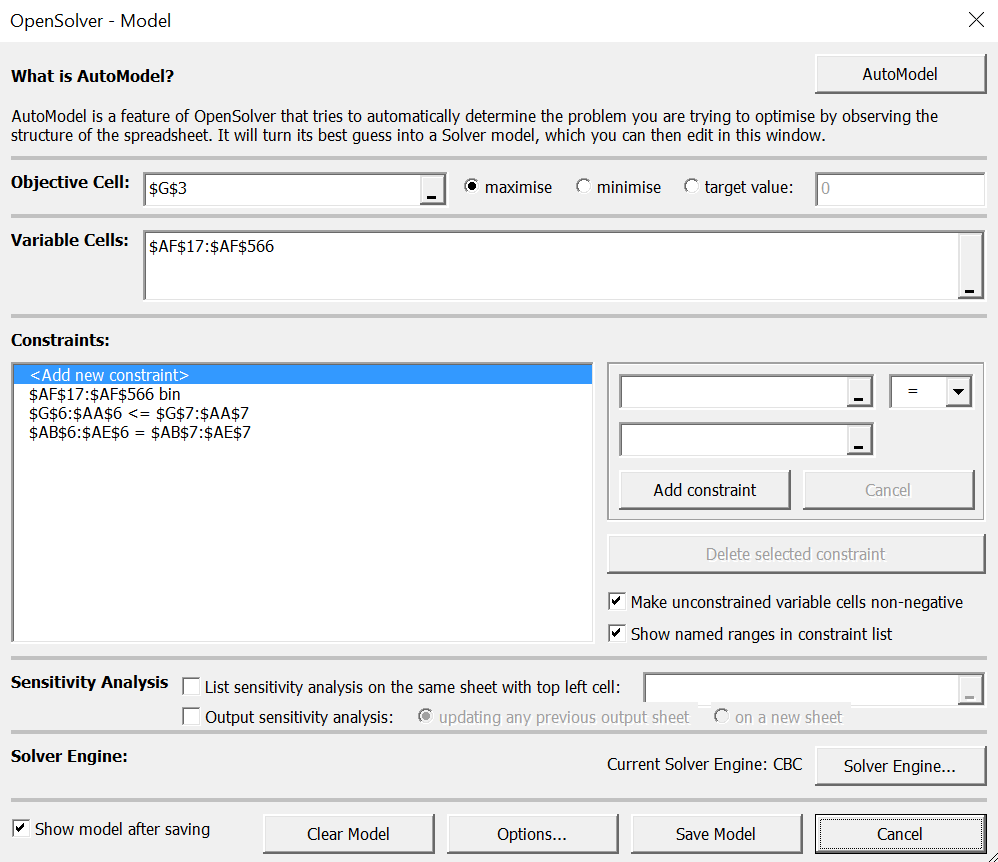


Figure 3 Open Solver model

The maximized score obtained was 2118.

A maximum of 3 players from the team Chelsea were selected.



Figure 4 OpenSolver Output

Maximizing the overall Form

Similar to the knapsack model solved above, a model to maximize the overall form in the team was built and solved.

The maximized value of the cumulative form was 106.7

### GAM Model

The Same models were also solved using GAM. Shown below are the outputs from the model to maximize the Score and the model to maximize the form.

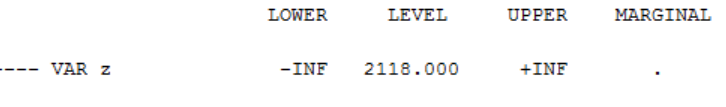


Figure 5 Maximize Score GAMs output

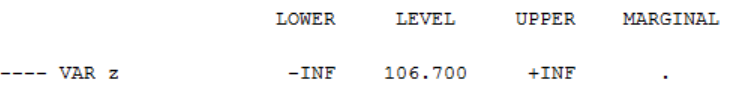


Figure 6 Maximize form GAMs output

Overall the outputs from both the OpenSolver and GAM match.

# Goal Programming Model

Goal programming is an approach where we consider multiple criteria when making the final decision.

A traditional linear programming model consists of a single objective function that is either maximized or minimized, a goal programming model consists of a set goals which are given priorities. The Final feasible solution may not meet all the goals. The objective of the model is of satisfy the true constraints present in the model and come close to meeting all the goals.

## Model Formulation for Goal Programming

The maximized values obtained for overall score and the overall form from their individual knapsack models were 2118 and 106.7 respectively.

In order to try and maximize both simultaneously, the final goals for score and form where set to the desired values of 2000 and 100 respectively (close to their maximum values). Both the goals were given equal weightage.

Goal constraints that get added to the model in addition to the constraints mentioned in the knapsack model are:

All values are

Here and represent the score and form of the player

Since the objective is to maximize both score and form (each having the same weight) the final objective becomes

*Minimize*

Here and represent the maximum possible values of overall score and form obtained from the individual single objective models (they are 2118 and 106). Each slack variable is divided by these terms so that the scales or units of measurement of score/form do not influence the actual weightage and deviation variable are in the same range (Pukkala, 2013).

## Solving the Goal Programming model

The Goal programming model was solved using OpenSolver.

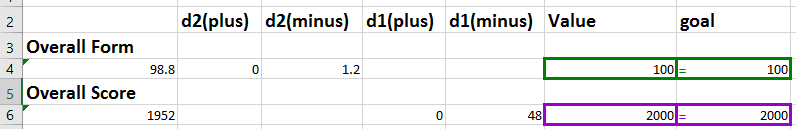


Figure 7 Goal Programming - OpenSolver

The optimized values obtained for the overall score and overall form are 1952 and 98.8 respectively.

# Conclusion

Using the Goal programming approach, we were able to come up with an optimized lineup of 15 players with a good balance between score and form which come close to the maximum values of single objective models.

# Key Learnings

The key learnings from this project were:

* Solving real-life problems that involve huge datasets and a large number of constraints and variables
* Learned to use OpenSolver to solve problems involving many variables
* Better understanding of Knapsack problems and their practical applications
* Better understanding of model solving using GAMS
* Application of Goal programming for solving multi-objective problems.

# Bibliography

(n.d.). Retrieved from Fantasy Premier League: https://fantasy.premierleague.com/help/

Pukkala, T. (2013). *Multi-objective Forest Planning.* Springer Science & Business Media.

# Appendix A – Players selected from the Goal Programming Model

Below is the list of 15 players that were selected from the 550 players using the Goal programming model:

Table 2 Selected players from the model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team | Name | Cost | Form | Score |
| Bournemouth | King | 6 | 6.8 | 143 |
| Bournemouth | Boruc | 4.5 | 4.8 | 100 |
| Bournemouth | Afobe | 5.5 | 7 | 78 |
| Burnley | Heaton | 5.2 | 6.3 | 134 |
| Burnley | Mee | 4.7 | 5.8 | 95 |
| Chelsea | Alonso | 6.8 | 4.3 | 143 |
| Crystal Palace | Zaha | 5.9 | 6.8 | 131 |
| Everton | Lukaku | 10.4 | 8.3 | 199 |
| Everton | Jagielka | 4.7 | 7.8 | 71 |
| Man Utd | Ibrahimovic | 11.4 | 4.8 | 163 |
| Middlesbrough | Gibson | 5 | 5 | 110 |
| Spurs | Alli | 9.2 | 8.3 | 188 |
| Spurs | Eriksen | 8.8 | 7.8 | 175 |
| Spurs | Son | 7.1 | 8.5 | 134 |
| West Brom | Dawson | 4.8 | 6.5 | 88 |