# Super 11

Guneet Khosla - gkhosla3
Karthik Nama Anil - kanil3
Pranit Kaul - pkaul9
Prithvi Alva Suresh - psuresh38

## 1. Introduction

Cricket is a sport that is played between two teams of eleven players comprising of batsmen, bowlers, and a wicketkeeper. A fantasy sport is a virtual game played online, where participants assemble their virtual teams of players of a professional sport. The number of fantasy sports users has grown steadily over the years and is still growing.

Selecting a team of players that will perform well is not that complex, is it? (One can just pick any players that have the best performance to date! ) Well, the problem becomes complex when one has to pick a team within a finite budget. A balanced team in Cricket has a good combination of batsmen, bowlers, fielders, and all-rounders adding to the complexity of selection of a team.

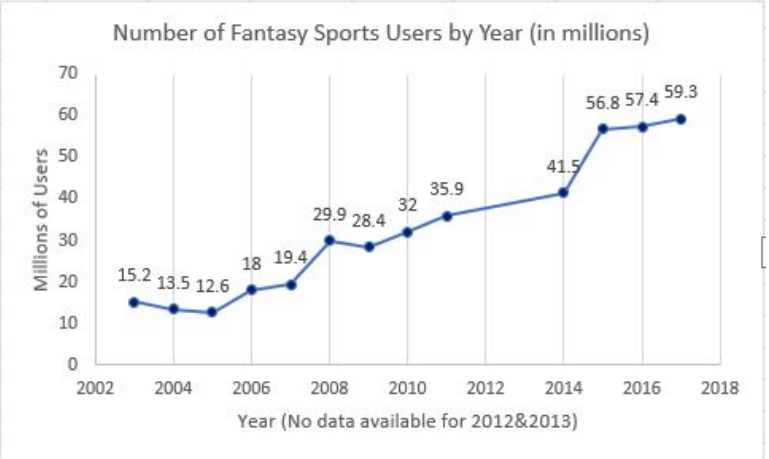


**Figure 1:** Fantasy Sports users (2002-2018) [13]

The main motivation behind this project is to help novice players in drafting a good team that potentially maximize their profits. This will be in the form of smart suggestions for picking players or a complete team suggestion altogether. Furthermore, to increase scope in usability, our method can also be applied to for player selection in dynamic auction environments.

## 2. Problem Statement

Currently, there exists limited research in the area of Fantasy League Cricket team selection. There is a clear amount of work [1, 11] performed on aggregating Cricket data and visualizing performances for players.

There has been only one paper [2] that has explored the integer programming model for selecting an 11 person Cricket team. But, no prior research in this area can be practically applied for the Fantasy league environment for a 20 over match.

We aim to select the best 11 players for a Fantasy Cricket league from the two playing teams based on past player performance under the specified budget constraints. While forming such a team, we will take into consideration various constraints imposed by platforms hosting fantasy cricket contests. For basing our decisions, we will be using IPL Cricket matches dataset from Kaggle (https://www.kaggle.com/manasgarg/ipl) containing ball by ball statistics for matches between 2008-2016.

## 3. Survey

Despite the huge betting market that exists for Cricket, there has been limited research work that has been carried out for prediction in Cricket. Brooks et al., analyze the prediction of test Cricket outcomes using an ordered response model [3]. The results, based on data over the period 1994 to 1999, suggest that the ordered categorized production outcome of test Cricket (win, draw, loss) can be explained by simple measures of the batting and bowling labor inputs. Batting and bowling averages of each player on the team, toss outcome and home ground advantages are considered to determine the relative strength of the team [7]. Scarf and Shi's work deals with forecasting match outcome probabilities by performing ball by ball analysis of the recent test matches [8] between the two competing teams. We will incorporate the techniques proposed in the papers for team performance prediction.

Pranavan et al., provide the optimal set of attributes for evaluation of player performance during the game[9]. Paul et al.propose a more granular approach (for each ball play) to analyze the performance of the Cricket players using isotonic regression to determine the cumulative impact of every player through the course of the game[10].

Financial modeling techniques(such as stocks expectation maximization, and budget restriction optimization) can be modified to suit Cricket fantasy leagues. We utilize the techniques proposed by Scarf, P., & Shi, X. to determine the player's score at the end of each game[8].

## 4. Proposed Methodology

We propose to build an algorithm that predicts the best team for each match by selecting a combination of players from the two competing teams. Our algorithm takes previous seasons data, budget constraints, game constraints, and a list of players playing in the upcoming match as it’s input, and outputs an optimal virtual team satisfying all the constraints.

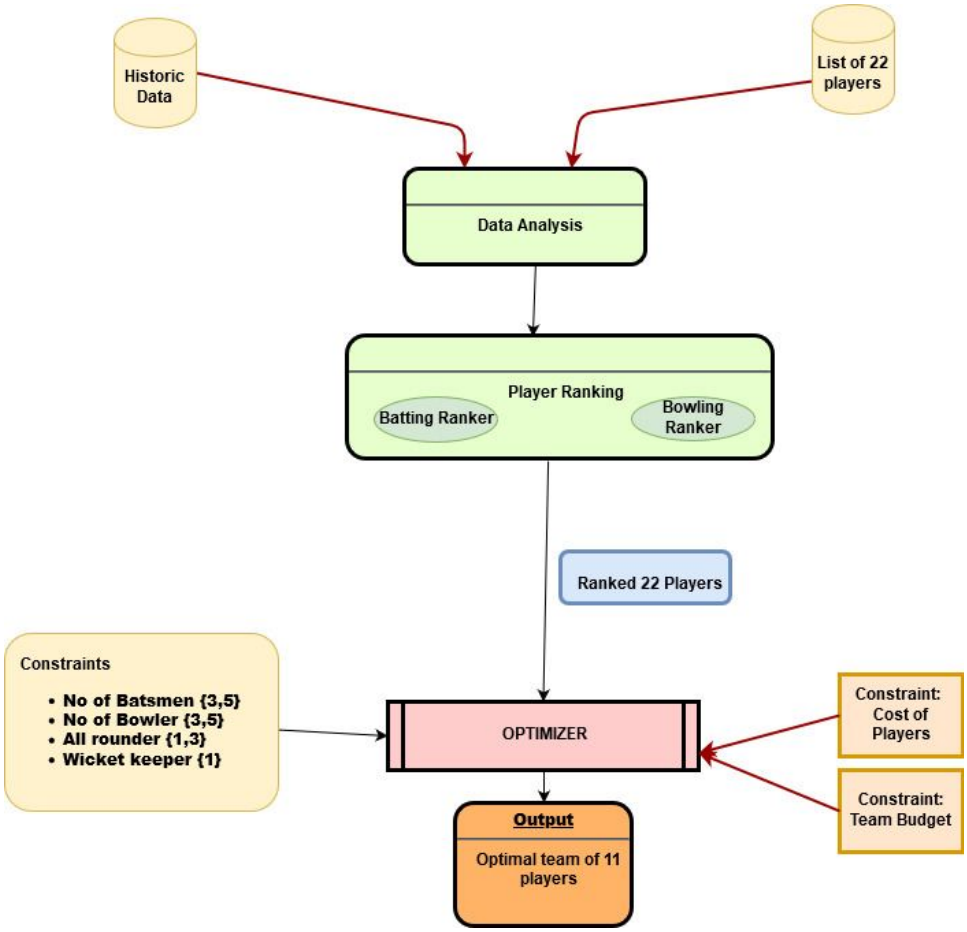


**Figure 2:** Proposed Methodology

### 4.1 Data Collection

For this project, we have used a Cricket dataset (https://www.kaggle.com/manasgarg/ipl) which consists of the ball by ball data of all the IPL Cricket matches from 2008 to 2016. The dataset contains 2 files: deliveries.csv and matches.csv.

- matches.csv contains details related to the match such as location, contesting teams, umpires, results, etc.
- deliveries.csv is the ball-by-ball data of all the IPL matches including data of the batting team, batsman, bowler, non-striker, runs scored, etc.

### 4.2 Data Pre-processing

We utilized Google's Open Refine for data cleaning and standardization. The dataset 'deliveries.csv' contains columns 'is_super_over' and 'penalty_runs' which were irrelevant for our data analysis and hence removed. The inconsistencies in the dataset were removed using the Key Collision Clustering Algorithm. Further, we had to fill in missing data and remove some discrepancies. Then, we performed data analysis.

### 4.3 Data Analysis

We analyzed ball by ball data of all the IPL cricket matches from 2008-2015 to generate player statistics. We built Python scripts that utilized Numpy and Pandas library to efficiently calculate various statistics for each player by aggregating his ball by ball performance imported from 'deliveries.csv'.

Statistics computed for Batsmen:

- Total Runs
- Total Matches
- Strike rate (Pace at which runs are scored)
- Total number of half-centuries
- Total number of centuries
- Highest Score in a match

Statistics computed for Bowlers:

- Total Wickets
- Total Matches
- Economy rate (Average runs given per over)
- Total Maiden Overs
- Most wickets in a match

The wicketkeeper was treated as a batsman and all-rounders were considered both as a batsman and a bowler.

### 4.4 Player Ranker

For each player, the above statistics were weighted as per importance and mapped to a real value called 'Player Performance'. Players with great records were assigned high 'Player Performance'.

Using the Player Performance metric, we calculate the relative ranks of players in each subcategory namely Batsmen, Bowler and All-Rounders, where rank 1 denotes the best performer.

### 4.5 Optimizer

After we have our set of players with their assigned ranks, we will forward this as input for our Optimizer. Also, we have a set of constraints that need to be applied for forming a team (keeping in mind the fantasy sports rules). The constraints for picking our team are -

- Number of Batsmen: 3-5
- Number of Bowlers: 3-5
- Number of Wicket Keepers: 1
- Number of All-Rounders: 1-3
- Total players in the team: 11

Also, we are allocated a budget to pick our team -

- Budget: 100 points

The cost for each player ranges from 7 - 11 points. A budget of 100 points leaves us with an average of 9 points per player. Hence, the cost of buying expensive players (cost > 9 points) have to be offset by the cost of relatively cheaper players (cost < 9 points). We have collected these costs for players from the cricket fantasy app 'Dream 11' [14].

Our problem statement reduces to perform optimization in multivariate scalar functions. So, we plan on exploring the following techniques.

### 4.5.1 Sequential Least SQuares Programming (SLSQP)

The SLSQP method deals with constrained minimization problems, where each constraint is specified using the constraint type and a constraint function. It also allows us to specify lower bounds and upper bounds on input variables using a list of tuples. It is used to optimize set of equations of the form:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to:} \quad & c_j(x) = 0, \quad j \in \mathcal{E} \\ & c_j(x) \geq 0, \quad j \in \mathcal{I} \\ & \text{lb}_i \leq x_i \leq \text{ub}_i, \, i = 1, \ldots, N. \end{aligned}$$

However, the caveat with using SLSQP for our problem is that it tends to find local minima as opposed to the global minima. Therefore, our results will vary and can change with every random seed. Therefore, we are experimenting with other minimization techniques as well.

### 4.5.2 Global Optimization(sub-section under construction)

Global Optimization method finds the global minimum of a function within given bounds, in the presence of potentially many local minima. We aim to explore the following Global Optimization algorithms(section under construction):

- Simplicial homology global optimization
- Dual annealing
- Differential evolution
- Basin-hopping

### 4.6 Innovations

We are introducing profit maximization algorithms (used in finance/trading) to solve a complex problem in the field of cricket. We believe our approach can also be extended to fit many other fantasy sports leagues.

Also, we tweaked our minimization methods to account for the imbalance in batting, bowling, wicketkeeping strength to build the optimizer.

## 5. Experiments and Evaluation(section under construction)

Dream 11 hosts multiple fantasy sports league in India including Cricket(specifically IPL). Dream 11 assigns points for each action like bowling, boundaries, catches, wickets, etc. We applied the same scoring system to our dataset that has a ball by ball data and determines the points each player has earned at the end of the match from each delivery bowled.

| **Regular Points** | |
|---|---|
| **Type of Points** | **IPL T20** |
| Being a part of the starting XI | 4 |
| Every run scored | 1 |
| Every wicket taken (excluding run out) | 25 |
| Catch taken | 8 |
| Stumping/ Run Out (direct) | 12 |
| Run Out (Thrower/Catcher) | 8/4 |
| Dismissal for a Duck (only for batsmen, wicket-keepers and all-rounders) | -2 |

**Table 1:** Dream 11 Fantasy Cricket Points System - Regular Points[14]

| **Bonus Points** | |
|---|---|
| **Type of Points** | **IPL T20** |
| Every boundary hit | 1 |
| Every six-hit | 2 |
| Half-Century (50 runs scored by a batsman in a single inning) | 8 |
| Century (100 runs scored by a batsman in a single inning) | 16 |
| Maiden Over | 8 |
| 4 wickets | 8 |
| 5 wickets | 16 |

**Table 2:** Dream 11 Fantasy Cricket Points System - Bonus Points [14]

We pass 22 players from the two teams playing the match along with the scores they earned from the simulation into an optimiser similar to the one discussed previously. The previous optimiser selected the best team(referred to as "Super 11" in further sections) by minimising the sum of ranks of the players for the given constraints. The optimiser used in the evaluation phase will pick the best team(referred to as "Simulation 11" in further sections) by maximising the total of scores of players earned during the simulation for the given constraints.

We are expecting that in a few scenarios, there will be a high overlap between the players in Super 11 and Simulation 11 aligning with our prediction model. However, this is not the case always, few players who were previously performing well may have a bad day and perform poorly, and on the other end of the spectrum, poorly performing players may shine in a particular match. Currently, we are working on developing the prediction model to ensure that Super 11 and Simulation 11 will have high overlap in players in all scenarios.

## 6. Responsibilities

We have equally divided the total workload among the team members and collaboratively worked on the project.

| **Name** | **Responsibilities** |
|---|---|
| Pranit Kaul | Literature Survey; Data Collection; Model Design; Model Training; Data Analysis; Report Documentation |
| Karthik Nama Anil | Literature Survey; Data Cleaning; Model Training; Model Fine Tuning; Data Visualisation; Report Documentation |
| Guneet Khosla | Literature Survey; Feature Engineering & Selection; Model Design; Model Fine Tuning; UI Development; Report Documentation |
| Prithvi Alva Suresh | Literature Survey; Data Cleaning; Evaluation of Results; Data Analysis; Data Visualisation; UI Development; Report Documentation |

**Table 3:** Responsibilities of Each Team Member

## 7. Timeline & Cost

The project will take approximately 200 man-hours(50 hours for each team member). The project does not involve any monetary costs other than the 200 man-hours. We intend to use our laptops for code development and testing.

Design & implement team performance prediction model is our midterm goal. The final milestone is to develop a minimum viable product with acceptable accuracy for the prediction model team selection.

| Due Date | Original Plan | Updated Plan |
|---|---|---|
| 7th Oct 2019 | Literature Survey & Project Proposal | Literature Survey & Project Proposal |
| 21st Oct 2019 | Design Finalization & Data Pre-processing | Design Finalization, Data Collection & Data Pre-processing |
| 4th Nov 2019 | Evaluate Multiple Data Analysis Models & Interpreting Results | Selection of Prediction Models from Initial Analysis |
| 8th Nov 2019 | Design & Implementation of Training Model for Team Performance Midterm Milestone Presentation & Report | Design & Implement Team Performance Prediction Model & Midterm Milestone Report |
| 18th Nov 2019 | Implementation of Prediction Model, UI Development & Data Visualisations | Design & Implement Team Selection Model, UI Development & Data Visualisations |
| 3rd Dec 2019 | Final Project Poster Presentation, Evaluation of Prediction Model & Fine Tuning Prediction Model | Final Project Poster Presentation, Evaluation of Prediction Models, Fine Tuning Prediction Models & UI Development |
| 9th Dec 2019 | Final Project Report | Final Project Report |

**Table 4:** Timeline Summary

## 8. Conclusion & Discussions

We have presented a novel framework to select the best 11 players for a fantasy cricket league from the two playing teams based on past player performance under the specified budget constraints. We have performed extensive experimentation and evaluation to ensure a good overlap between the selected team and the best team possible for the given constraints.

Our framework can be extended to live auction scenarios as well. The inclusion of certain parameters such as playing field conditions and the weather may further increase the accuracy of prediction.

## References

[1] Hemanta Saikia, Dibyojyoti Bhattacharjee, and Her-manus H Lemmer. Predicting the performance of bowlers in IPL: An application of artificial network. International Journal of Performance Analysis in Sport,12(1):7589, 2012

[2] Debarghya Das, Moneyballer: An Integer Optimization Framework for Fantasy Cricket League Selection and Substitution (2010)

[3] Robert D. Brooks, Robert W. Faff & David Sokulsky (2002) An ordered response model of test cricket performance, Applied Economics, 34:18, 2353-2365, DOI: 10.1080/00036840210148085

[4] G D Sharp, W J Brettenny, J W Gonsalves, M Lourens & R A Stretch (2011) Integer optimisation for the selection of a Twenty20 cricket team, Journal of the Operational Research Society, 62:9, 1688-1694, DOI: 10.1057/jors.2010.122

[5] Budak, G., Kara, İ., Iç, Y.T., Kasımbeyli, R., New mathematical models for team formation of sports clubs before the match, Central European Journal of Operations Research 2019

[6] Faez Ahmed, Kalyanmoy Deb, Abhilash Jindal, Multi-objective optimization and decision making approaches to cricket team selection, Elsevier B.V., DOI: 10.1016/j.asoc.2012.07.031

[7] Allsopp, P. E., & Clarke, S. R. (2004). Rating Teams and Analysing Outcomes in OneDay and Test Cricket. Journal of the Royal Statistical Society. Series A (Statistics in Society), 167(4), 657-667.

[8] Scarf, P., & Shi, X. (2005). Modelling match outcomes and decision support for setting a final innings target in test cricket. IMA Journal of Management Mathematics, 16(2), 161-178. doi: 10.1093/imaman/dpi010

[9] Somaskandhan, Pranavan et al. Identifying the optimal set of attributes that impose high impact on the end results of a cricket match using machine learning. IEEE International Conference on Industrial and… 2017 DOI:10.1109/iciinfs.2017.8300399

[10] Paul J. Bracewell, Ankit K. Patela,b, Jason D. Wellsa (2017) Real time measurement of individual influence in T20 cricket. MathSport International 2017 Conference Proceedings. ISBN 978-88-6938-058-7

[11] Sricharan Shah, Partha Jyoti Hazarika and Jiten Hazarika (2017) A Study on Performance of Cricket Players using Factor Analysis Approach. International Journal of Advanced Research in Computer Science. ISSN No. 0976-5697

[12] Lemmer, H.H.: Team selection after a short cricket series, European Journal of Sport Science, DOI: 10.1080/17461391.2011.587895 (2013).

[13] Fantasy sports growth. Retrieved from https://en.wikipedia.org/wiki/Fantasy_sport

[14] Dream 11 Fantasy Cricket Points System. Retrieved from. https://get.dream11.help/hc/en-us/articles/360021110932-Fantasy-Cricket-Points-System