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. Also, 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.

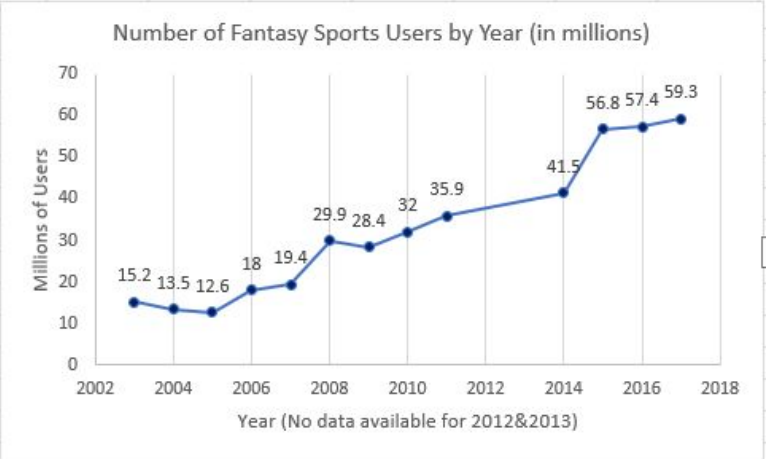


Figure 1: Fantasy Sports users (2002-2018) [14]

2. Problem Statement

The motivation behind this project is to help players all around the world (competing in fantasy league apps) by drafting a good team that can maximize their profits and earn them rewards on the platform. This will be in the form of smart suggestions for picking players or a complete team suggestion altogether.

Our goal is to dynamically select 11 players to form a team on a Fantasy Cricket app for a given match. The cricket fantasy league platform that we have selected is called “Dream 11”. We will generate this team from the two playing teams squads looking back on past player performance. There is a cost associated while picking any player in Dream 11. So, we need to consider these costs as well as , and keep it under the budget based on specified constraints. While forming this

team, we will take into consideration various constraints imposed by platforms hosting fantasy cricket contests.

3. Survey

Limited research is carried out currently in the area of Fantasy League Cricket team selection. Most of the current works concentrate on aggregating data and visualising player performances[1, 11]. Few research works suggest using integer programming for selecting the 11 person Cricket team but they have been applied to longer formats of the game[3] and not T20 Cricket.

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 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.

4.1 Data Collection

For this project, we have used a Cricket dataset[13] which consists of the ball by ball data of all the IPL Cricket matches from 2008 to 2017. The dataset contains 2 files: deliveries.csv and matches.csv which contains 150K rows.

- 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.

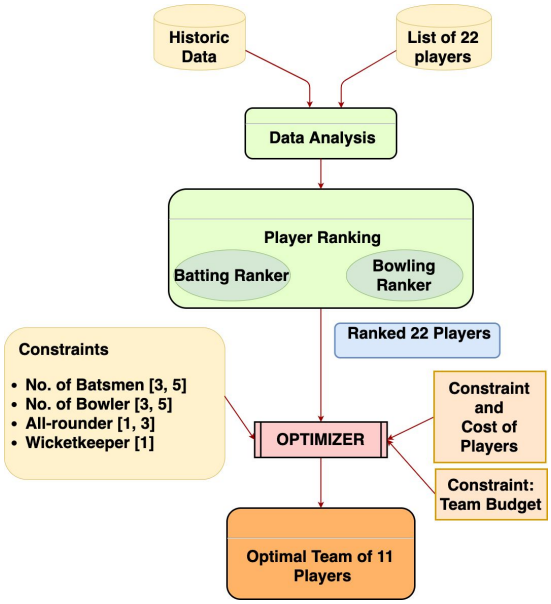


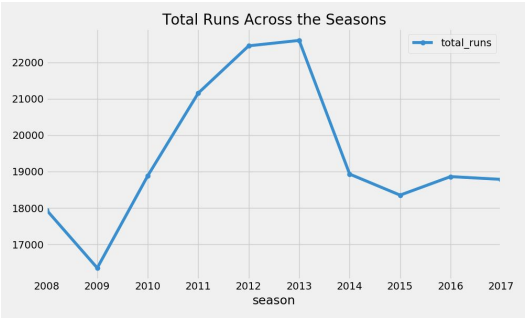
Figure 2: Proposed Methodology

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 Exploratory Data Analysis

We initially performed exploratory data analysis to understand the dataset better and to get an intuition about the distribution of various statistics of players. We constructed a number of plots to calculate player utility values; we analyzed player metrics like runs scored, wickets taken, strike rate, bowling economy and many more. The analysis helped us determine the weights of each metric that is used in the following sections to compute the player ranks.



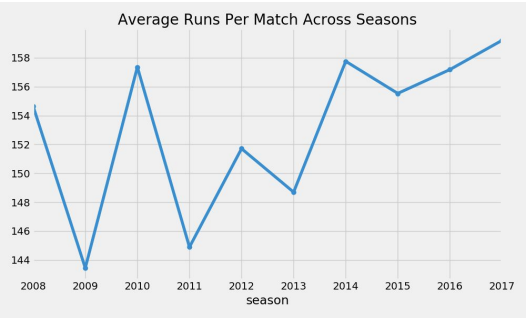


Figure 3: a) Total runs across the seasons b) Average runs per match across seasons

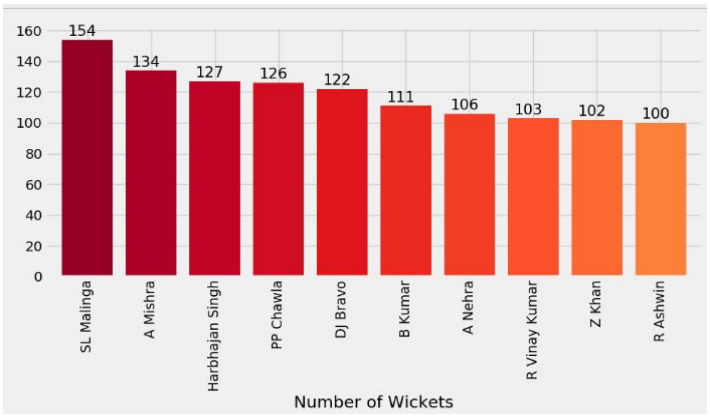


Figure 5: Top 10 wicket takers across all seasons

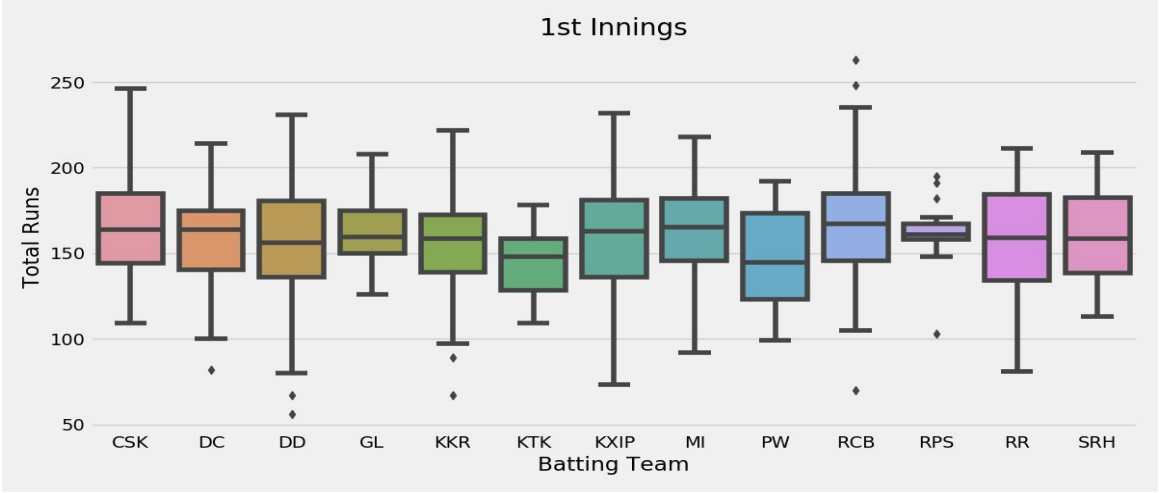


Figure 7: Boxplot of runs scored for an innings by each team across all seasons

4.3 Data Analysis

We analyzed ball by ball data of all the IPL cricket matches from 2008-2016 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'.

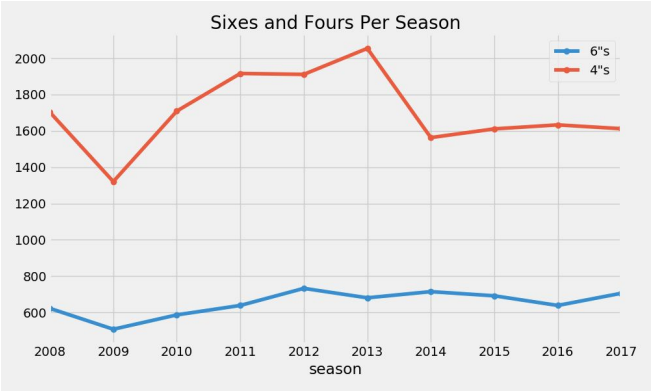


Figure 4: Sixes and Fours scored per season

Statistics computed for Batsmen:

- Total runs scored
- Total matches played
- Strike rate (Pace at which runs are scored)
- Total number of half-centuries and centuries
- Average score
- Number of 4's and 6's each match

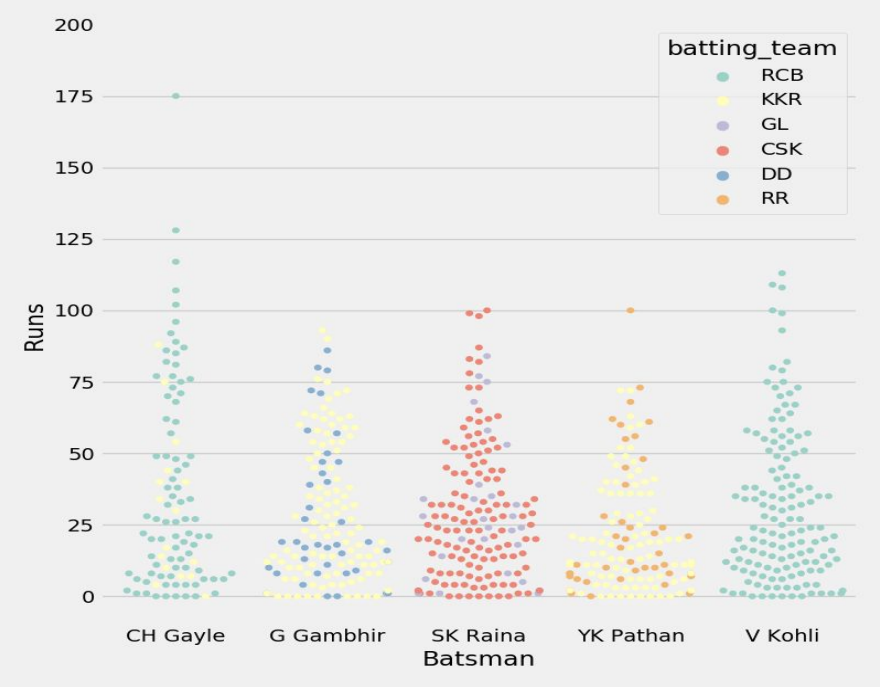


Figure 6: Runs scored in each match with highest strike rate across all seasons

Statistics computed for Bowlers:

- Total wickets taken
- Total matches played
- Economy rate (Average runs given per over)
- Total maiden overs(Overs with zero runs conceded)
- Number of wickets per 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 an integer values called 'Player Utilities'. Players with great records were assigned high 'Player Utilities'.

Using the Player Performance metric, we calculate the relative ranks of players in each subcategory namely Batsmen, Bowler and All-Rounders, where a high utility denotes the better performer.

4.5 Optimizer

After we have our set of players with their assigned utilities, we forward this as input to 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 (as per Dream 11 platform[15]) 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' [15].

Our problem statement reduces to perform optimization in multivariate scalar functions. So, we explored 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

Global Optimization method finds the global minimum of a function within given bounds, in the presence of potentially many local minima. We explored the following Global Optimization algorithms:

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

4.5.3 Integer Programming

We explored the techniques discussed previously, but none of them were suitable for our problem. We identified that Integer Programming is the most accurate method to develop our model and generate results. This owes to the fact that our problem involves solving multiple constraints, coupled with a convex optimization problem (to maximize rewards). So, we have implemented the GLPK package using python. GLPK is a free software and licensed under the GNU General Public License 3. It is used for solving large-scale linear programming (LP), mixed integer programming (MIP), and other related problems.

5. Innovations

Key innovations of our project are -

- We have developed a novel approach to predict a team that can perform well and win maximum rewards for a given match.
- We have developed a model by combining Integer Linear Programming with convex optimization to predict a team with the constraints specified before.
- Modified our optimizer by maximization methods to accommodate a balance in batting, bowling, wicket keeping strength while implementing the optimizer.

6. Experiments and Evaluation

We generated our model by looking at matches played during 2008-2016 (**training set**) and we evaluated their performance on matched played during 2017 (**test set**) . Dream 11 hosts multiple fantasy sports league in India including Cricket(specifically IPL). Dream 11 assigns points to each player for every 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 the outcome of each delivery bowled.

Regular Points	
Type of Points	**IPL T20**
Being a part of the starting XI	2
Every run scored	0.5
Every wicket taken (excluding run out)	10
Catch taken	4

Stumping/ Run Out (direct)	4
Run Out (Thrower/Catcher)	4
Dismissal for a Duck (only for batsmen, wicket-keepers and all-rounders)	-2

Table 1: Fantasy Cricket Points System - Regular Points[15]

Bonus Points	
Type of Points	**IPL T20**
Every boundary hit	0.5
Every six-hit	1
Half-Century (50 runs scored by a batsman in a single inning)	4
Century (100 runs scored by a batsman in a single inning)	8
Maiden Over	2
4 wickets	4
5 wickets	6
Economy rate < 4	8
Economy rate [4, 5)	6
Economy rate [5, 6)	4
Economy rate [6, 7)	0
Economy rate [7, 8)	-1
Economy rate [8, 9)	-2
Economy rate [9, 10)	-4
Economy rate >= 10	-6

Table 2: Fantasy Cricket Points System - Bonus Points [15]

We picked up an optimal team (by looking at the future match) and compared it against our predicted team. This optimal team is basically selected to gauge our predictions , because the optimal team accumulates the maximum reward possible in that match. The results are available in the table below -

Predicted 11 vs Optimal 11			
Match Id	Overlap between teams (/11)	Total points (Predicted 11)	Total points (Optimal 11)
1 - SRH vs RCB	7	235.5	244.8
7 - MI vs KKR	7	240.5	272.7
10 - SRH vs MI	6	197.5	249.3

12 - RCB vs MI	6	165.0	267.75
14 - KKR vs SRH	7	235.5	248.4
27 - KKR vs RCB	6	175.0	283.5
36 - SRH vs KKR	7	235.5	248.4
37 - RCB vs MI	6	189.0	257.85
45 - RCB vs KKR	5	155.0	258.75
47 - MI vs SRH	6	192.0	235.35
53 - MI vs KKR	4	155.0	246.6
57 - SRH vs KKR	7	162.0	175.95

Table 3: Predicted 11 vs Optimal 11 match simulation results

We have displayed findings from our simulated matched above. Overall ,we observed that our predicted team achieved a score that was ~90% as compared to the maximum points achievable. Also, we observed that there was an overlap in players in many simulations.

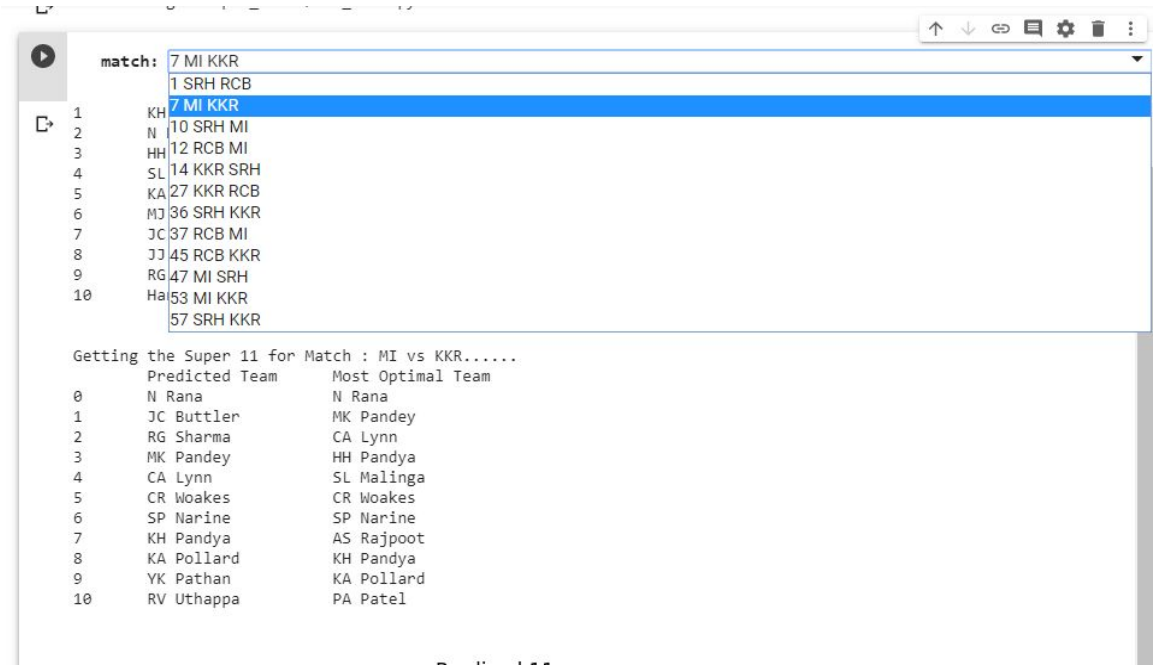


Figure 7: User Interface to select a match and run the simulation

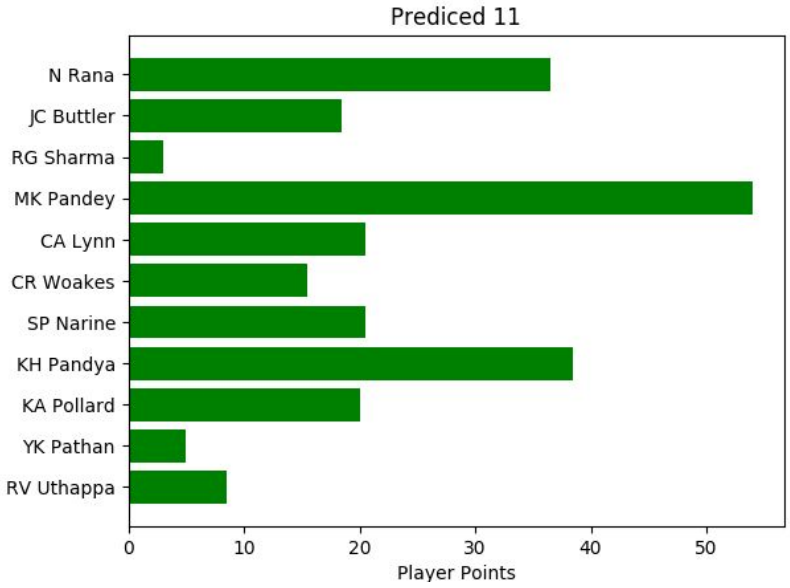


Figure 8: Rewards accumulated by players (Predicted Team) after Simulation

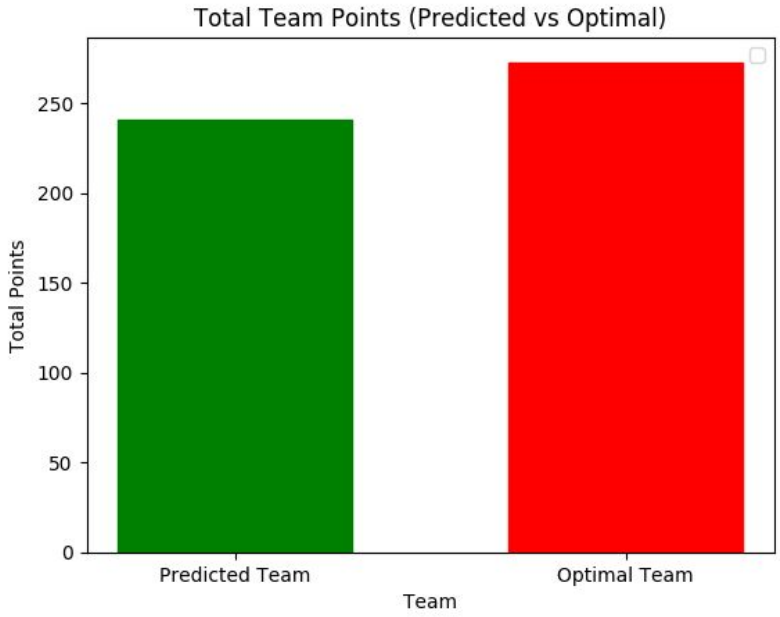


Figure 9: Predicted Team vs Optimal Team

7. 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 4: Responsibilities of each team member

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 significant overlap between the Predicted 11 and the Optimal 11 possible for the given constraints. Furthermore, we observed the Predicted 11 and Optimal 11 had comparable points at the end of the ball by ball simulation for each match.

Our framework can be extended to support player selection in 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 and can be an area of further research.

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] Indian Premier League (Cricket) Ball-By-Ball Cricket Data. Retrieved from https://www.kaggle.com/manasgarg/ipl

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

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