

Super 11

Guneet Khosla - gkhosla3

Karthik Nama Anil - kanil3

Pranit Kaul - pkaul9

Prithvi Alva Suresh - psuresh38

Abstract

Cricket is a sport that is played between two teams of eleven players comprising of batsman, bowlers and a wicket keeper. For those unfamiliar, it is comparable to baseball regarding it's gameplay. A fantasy sport is a type of game, often played using the Internet, where participants assemble imaginary or virtual teams of real players of a professional sport.

The idea of picking a set of players and running a contest based on their performance stats has been around for quite a long time. After the dot-com boom of the internet, more and more people started investing in selecting their "roster" by anticipating how well the combination of players will perform during a season. This introduced us with the concept of fantasy sports where participants assemble imaginary of virtual teams of players of some professional sports. The number of fantasy sports users has grown steadily over the years and is still growing.

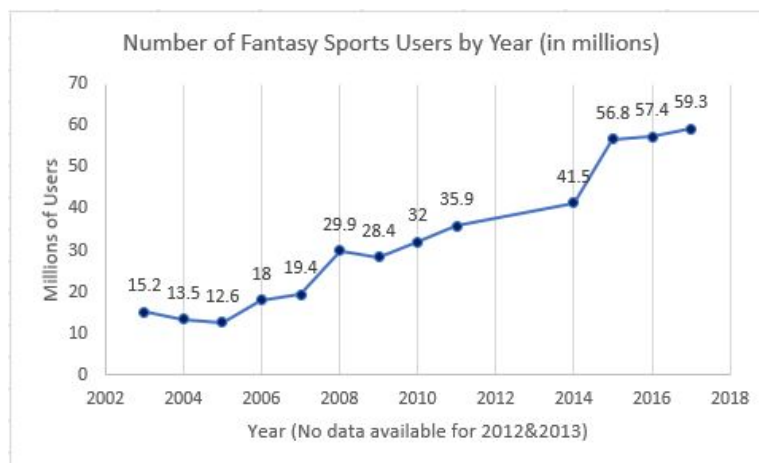


Figure 1 : Fantasy Sport users (2002-2018) [13]

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 till 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.

Introduction

The aim of our project is to select a team of players for each match of the IPL (Indian Premier League) season. 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 on Kaggle (<https://www.kaggle.com/manasgarg/ipl>) containing ball by ball statistics for matches between 2008-2016.

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. However, they fail to provide any substantial model for selecting players based on match conditions.

There has been only one group [2] that has explored integer programming model for selecting an 11 person Cricket team. But, no prior research in this area can be practically applied for Fantasy league environment for a 20 over match. Also, due to a budget restriction, this problem is somewhat similar to a portfolio selection problem in finance. So, we will explore optimization techniques like Markowitz problem and find scenarios where can apply such algorithms or tweak them for our problem.

Significance

- People spend hours drafting their team for the league.
- Novice has a chance to win contests using this software.
- Our approach suitable for player selection in dynamic auction environments.

Literature Survey

Limited research work has been carried out in prediction in Cricket compared to other sports like Soccer, American Football and Baseball. This is quite surprising since there is a huge betting market for Cricket. For example, the total amount bet for a One-Day International (ODI) Cricket match between Pakistan and India is in the neighbourhood of 1 billion USD.

Among the work done on forecasting in Cricket is done with pre-match forecasting, for example; The paper analyses the prediction of test Cricket outcomes using an ordered response model. 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 labour inputs [3]. Allsopp and Clarke determine the Cricket teams' relative strength using multivariate regression. 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. This technique works well for all formats of Cricket[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 to evaluate the player performance during the game[9] like “totalWickets”, “singlesInMiddleOvers”, etc. Paul et al propose a granular level(ball by ball) approach to analyse the performance of the Cricket players using isotonic regression to determine cumulative impact of every player through the course of the game[10]. Financial modelling techniques(like stocks expectation maximization, linear programming and budget restriction optimization) can be modified to suit Cricket fantasy leagues. We utilize the techniques proposed in the paper to determine the player's score at the end of each game[8].

Risks and Payoff

We are not considering the weather and Cricket pitch conditions during team selection and player performance evaluation. This is a calculated risk we are taking with the assumption that good players are adaptable and play well in all conditions.

If our project is successful, inexperienced fantasy Cricket league players can use our software to accurately predict the best Cricket team in the given constraints and make a good amount of money playing fantasy Cricket leagues.

Outcomes and Deliverables

We aim to build an algorithm to assign a single metric (player's cost) ranging from 5.0 to 11.0 to each of the 22 players participating in the upcoming match. The cost of each player is calculated by applying varying weights to a list of attributes but not limited to-

- Player's performance in recent most matches
- Player's past performances in the stadium where the upcoming match is being hosted
- Player's past performances against the opponent team.

The above algorithm uses data analytics techniques to assign high costs to players with great record. Next, we aim to build an Optimizer function to predict the best 11 players (a combination of batsman, bowlers and all-rounders) within a limited budget of 100 points - which roughly translates to 9 points for each player.

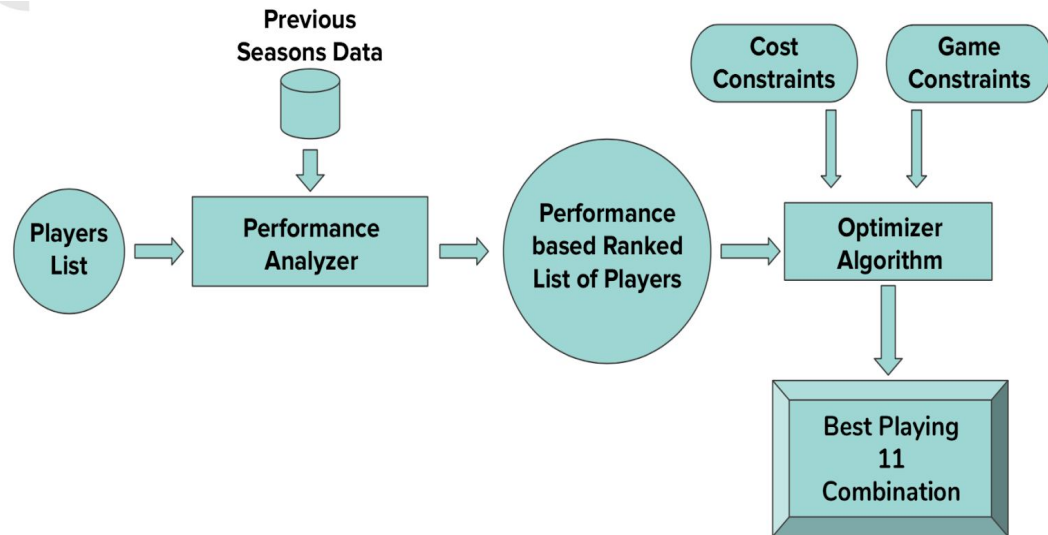


Fig 2 : High Level Design Overview of the System

Deliverables

- An intelligent algorithm to calculate a player's cost based on previous and current season's performance data.
- An optimizer function to produce the best combination of 11 players within a limited budget of 100 points.
- A web portal that lets users enter the match details and outputs the best playing 11 for that particular match.
- A match simulator to compare the performance of the team produced by optimizer function against various teams recommended by Cricket experts.
- Dynamically produced graphs and charts to support our claim of best playing 11 players.

Technology Stack

- Visualisations : D3, Tableau
- OpenRefine and Scrapy for dataset cleaning and preparation
- Database management software (MySQL Workbench)
- Web technology stack : AngularJS, Ruby on Rails
- Languages : Python, Javascript, HTML, CSS

Responsibilities

Table 1 : Responsibilities of Each Team Member

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

Timeline

The project will take approximately 200 man-hours(50 hours for each team member). Our midterm milestone is to complete the design and implementation of the training model to evaluate team performance. The final milestone is to develop a minimum viable product with acceptable accuracy post fine-tuning the prediction model.

Table 2 : Timeline Summary

Due Date	Work Expected to be Complete by the Due Date
7th Oct 2019	Literature Survey & Project Proposal
21st Oct 2019	Design Finalization & Data Pre-processing
4th Nov 2019	Evaluate Multiple Data Analysis Models & Interpreting Results
8th Nov 2019	Design & Implementation of Training Model for Team Performance Midterm Milestone Presentation & Report
18th Nov 2019	Implementation of Prediction Model, UI Development & Data Visualisations
3rd Dec 2019	Final Project Poster Presentation, Evaluation of Prediction Model & Fine Tuning Prediction Model
9th Dec 2019	Final Project Report

Cost

Apart from an estimated value of 200 man-hours, this project will not require any monetary costs. We intend to use our laptops for code development, testing and fine-tuning.

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., Kasimbeyli, R. ,New mathematical models for team formation of sports clubs before the match, *Central European Journal of Operations Research* 2019
- [6] FaezAhmed ·KalyanmoyDeb· AbhilashJindal , 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