Massachusetts Institute of Technology

Sloan School of Management

15.093 - Optimization Methods

Optimizing Cricket Team Composition: An Indian Cricket Team Case Study

Authors
Vidushi Gupta, Joseph Kajon

December 06, 2023

Introduction

Cricket is a sport that is deeply ingrained in the culture of several countries. It has a massive global following and holds significant economic and symbolic value. The success of a national cricket team is not just a matter of national pride but also a catalyst for economic benefits, international recognition, and societal morale. The factors of the game, with its rich history and evolving strategies, make it an exciting opportunity for analytical exploration and innovation.

Team composition is a key indicator of success in cricket. Each player brings unique skills and attributes to the team, and the right combination can be the difference between victory and defeat. However, cricket, as with many sports, is a game of complex variables and unpredictable dynamics. Player performance varies with conditions, match formats (like ODIs, Test matches, and T20s), and opponents, making the selection of the optimal team a challenging task.

By leveraging statistical data, teams can gain insights into player performances, team dynamics, and strategic approaches against various opponents in different match conditions. This project seeks to tap into this rich data opportunity to optimize cricket team composition, blending the art of cricket with the science of data analytics.

Problem Statement

The core challenge addressed in this project lies in effectively utilizing the wealth of cricket statistics to formulate the best possible team composition. Cricket, with its diverse player roles and match formats, presents a complex array of data points – from batting averages and strike rates to bowling economy and fielding records. The objective is to break down this complexity into actionable, holistic insights that can guide the selection of an optimal team.

A key aspect of this challenge is the need to balance various elements of the game – batting, bowling, fielding, and player roles such as wicket-keepers and all-rounders. Each match scenario, whether an ODI (One Day International - 50 overs), a Test match (450 overs), or a T20 (Twenty20 - 20 overs), demands a different strategy and thus a different team composition. Moreover, the performance of players is not static; it evolves over time and in response to different conditions and opponents. Our project aims to create a dynamic model that not only captures the essence of each player's performance but also adapts to various match scenarios.

In our project, we go beyond just enhancing a cricket team's performance by selecting an optimal team; we introduce a comparative analysis by measuring the actual performance scores of teams in historical matches against our calculated optimal performance scores. This approach provides a baseline performance score, serving as a real-world benchmark to assess the potential improvements offered by our optimal team compositions. Comparing these optimal scores to actual match outcomes allow us to highlight the efficacy of a data-centric approach in team selection and offer critical insights into the impact of analytics-based decision-making in cricket.

Why we care?

The strategic selection of players based on analytical data can revolutionize team management in cricket. By employing a data-driven approach, coaches and team managers can make informed decisions that enhance team performance, ultimately leading to increased success and competitiveness on the international stage. This international notability can positively impact the nation as it can boost national morale and the economy.

Datasets

The main data comes from a Kaggle dataset which covers 2850 cricket matches with varying match types and countries. The dataset is divided into three files:

1. *Bowling statistics* = individual player metrics specific to the "bowler" (defensive) position like runs conceded, maiden, economy, and wickets
2. *Batting statistics* = individual player metrics specific to the "batting" (offensive) position like runs scored, balls faced, strike rate, and reason for out
3. *Match statistics* = match information, such as date of the match, match number, and format.

After identifying India as having the most matches in the dataset, we made it the focus of the project. The data was filtered to only include matches where India played in. This was joined with another Kaggle dataset, which contained the list of Indian players and some additional statistics, to ensure only Indian players were considered for the model.

Optimization Model Formulation

We employed Gurobi in Python to develop three optimization models: one maximizing batting performance, another for bowling performance, and the last aiming to optimize both. These models take into account various constraints such as team size, player roles, and match types. Decision variables represent player selections, with the objective functions designed to maximize the respective performance metrics, derived from normalized and combined player statistics. The detailed optimization formulations are in Figure 1 below.

<u>Variables</u>

- x_p is a binary decision variable (1 if player p is chose, 0 if not)
- t_p represents the batting score for player p
- l_p represents the bowling score for player p
- b_p represents the normalized batting score for player p

- w_p represents the normalized bowling score for player p
- T = set of batsmen
- L = set of bowlers
- R = set of all rounders
- A = set of all players ($T \cup L \cup R$)

Constraints

- (1) Team cannot exceed 11 players

$$\sum_{p \in A} x_p = 11$$

- (2) Must be at least 4 batsmen on team

$$\sum_{p \in T} x_p \geq 4$$

- (3) Must be at least 4 bowlers on team

$$\sum_{p \in L} x_p \geq 4$$

- (4) Cannot be more than 2 all rounders on team

$$\sum_{p \in R} x_p \leq 2$$

Objective Function

Model 1

- Maximize team batting score

$$\max \quad \sum_{p \in T \cup R} x_p * t_p$$

Model 2

- Maximize bowling score

$$\max \quad \sum_{p \in L \cup R} x_p * l_p$$

Model 3

- Maximize combined normalized batting and normalized bowling scores

$$\max \quad \sum_{p \in T \cup R} x_p * b_p + \sum_{p \in L \cup R} x_p * w_p$$

Figure 1: Optimization Models maximizing batting score, bowling score, and both

Baseline comparison

To encapsulate the overall performance of a team, we created two metrics: batting score and bowling score. To calculate the batting score, the key batting statistics like strike rate, fall of wicket over, runs scored, run rate and total runs were summed and normalized. Similarly the bowling score is a summation of the important bowling statistics like the wickets per over, economy rate (which was inverted as lower the economy rate, better the bowler is), maiden over ratio, dot ball ratio and a penalty for wide and no balls. The higher these metrics, the better the team's performance. These metrics were key to finding the optimal team for our models.

When determining the impact of our optimization model, we used the historic batting and bowling scores as a baseline. For example, selecting an Indian team that maximizes bowling score against Australia in test matches increases bowling score from 0.52 to 0.59. This increase has the potential to add a strategic edge over India's competition as their overall bowling performance would be enhanced. Figure 2 highlights the baseline and optimal batting/bowlings scores for all three models in India vs Australia Test Matches. One can see that in each model, the optimal batting and bowling scores are higher than the baseline. To equally weigh their impact, the differences are slightly less in Model 3.

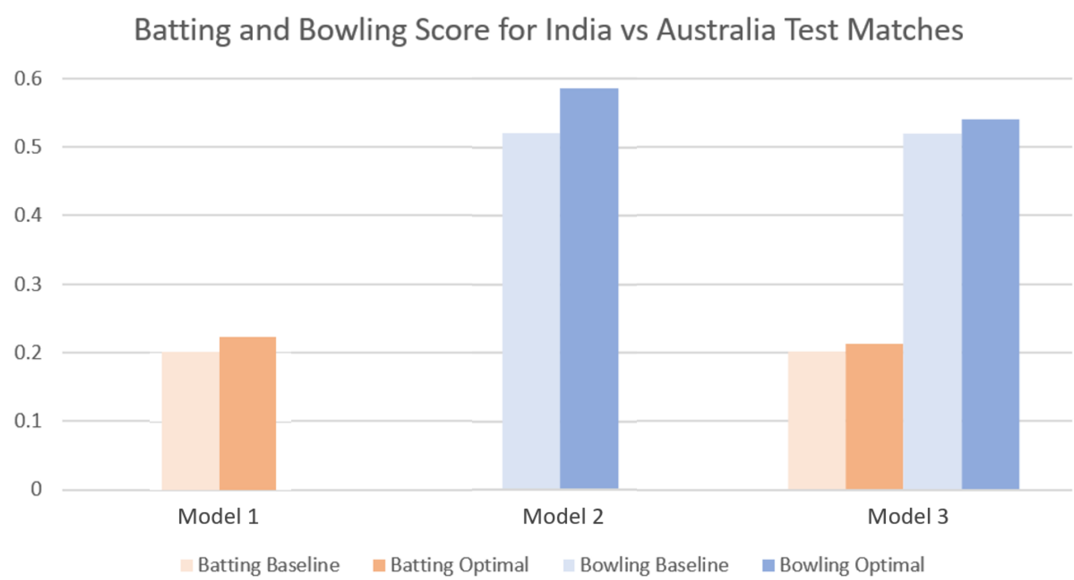

Figure 2: Batting and Bowling Score for India vs Australia Test Matches

Another way to visualize the impact of the different models is to look at the optimal team selected. In Figure 3, one can see if a player is on the optimal team when maximizing batting score (model 1), bowling score (model 2), or both (model 3). This gives interesting insights as there are some players that are well-rounded, such as Ravindra Jadeja, while others seem to excel only in one area, such as Bhuvneshwar Kumar. This provides insight into the impact of players and highlights those rare players that excel in both categories.

India vs Australia Test Optimal Team			
Player	Model 1	Model 2	Model 3
Amit Mishra	■		■
Bhuvneshwar Kumar	■		
Cheteshwar Pujara	■		
Jasprit Bumrah		■	■
Karn Sharma	■		
Karun Nair	■		■
Kuldeep Yadav		■	
Mayank Agarwal	■		
Mohammed Shami		■	■
Mohammed Siraj		■	■
Ravichandran Ashwin		■	
Ravindra Jadeja	■	■	■
Rohit Sharma		■	
Sachin Tendulkar		■	■
Shardul Thakur	■	■	■
Sourav Ganguly	■		■
Suresh Raina	■	■	■
Virender Sehwag		■	
Washington Sundar	■		■

Legend:	
■	Player on Team

Figure 3: Table of Optimal Teams for Models for India vs Australia Test Matches

Final Product

The culmination of our project involves a flask API that converts our optimization model into a user-friendly application and allows users to interact with our models. A video demonstration can be accessed [here]() or by scanning the QR code. This tool enables coaches and analysts to determine the optimal team composition for different match scenarios by simply inputting their requirements (opponent team name, match type, and type of model). The application also highlights the comparison between the baseline and optimal metric and the percent improvement that the optimization model can provide with. The display is shown below in Figure 4.

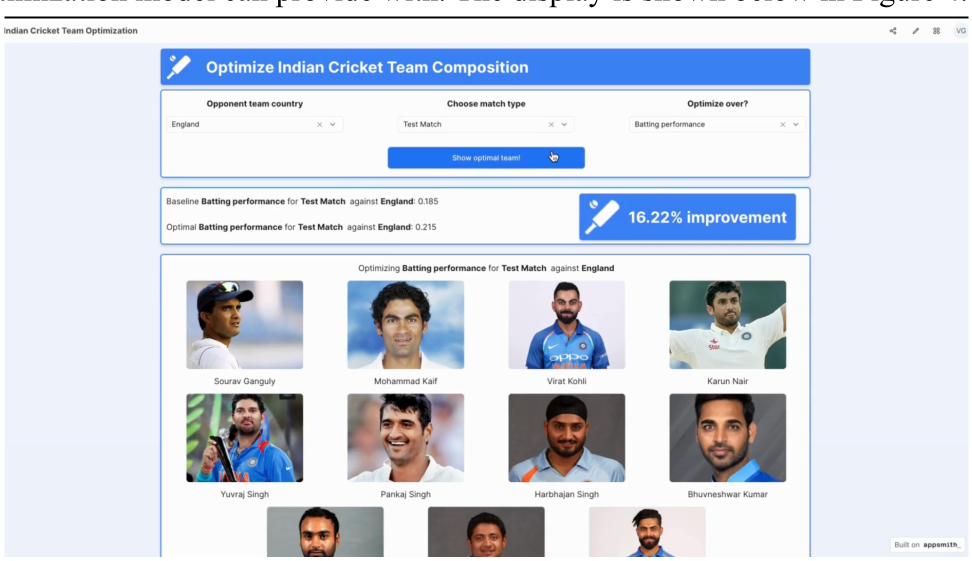

Figure 4: Look at the user-friendly application to visualize the results of the model

Impact

This project sets a precedent for analytics-based team management in cricket, potentially revolutionizing how teams are formed and strategies are devised. It offers a systematic and data-driven approach to team selection, which can lead to increased team success and reshape the future of cricket strategy.

Future Scope

The scope can potentially expand our model to encompass a wider array of teams and match formats, thereby broadening its applicability and relevance across the cricketing landscape. Additionally, an important aspect of our future work involves regularly updating the model to reflect the current cricketing landscape. This includes adding metrics for emerging players and phasing out those who have retired. Such updates ensure that the model stays relevant and accurate, reflecting the ever-evolving nature of the sport. Furthermore, the model can also incorporate player injury and fatigue metrics into our model. This addition will significantly enhance the model's accuracy and practicality by accounting for player fitness levels, which are crucial for peak performance. Recalculating the batting and bowling performance metrics will also help with understanding and factoring in these physical aspects will lead to more realistic and effective team compositions.