

# Determining the playing 11 based on opposition squad: An IPL illustration

G. Gokul<sup>a</sup> and Malolan Sundararaman<sup>b,\*</sup>

<sup>a</sup>*Department of Electrical and Electronics Engineering, National Institute of Technology Tiruchirappalli, India*

<sup>b</sup>*Department of Management Studies, National Institute of Technology Tiruchirappalli, India*

Received 27 February 2022

Accepted 19 August 2023

Pre-press 13 September 2023

Published 9 November 2023

**Abstract.** Indian Premier League (IPL) is the most popular T20 domestic sporting league globally. Player selection is crucial in winning the competitive IPL tournament. Thus, team management select 11 players for each match from a team's squad of 15 to 25 players. Different player statistics are analysed to select the best playing 11 for each match. This study attempts an approach where the on-field player performance is used to determine the playing-11. A player's on-field performance in a match is computed as a single metric considering a player's attributes against every player present in the opposition squad. For this computation, past ball-by-ball data is cleaned and mined to generate data containing player-vs-player performance attributes. Next, the various performance attributes for a player-vs-player combination is converted into a player's performance rating by computing a weighted score of the performance attributes. Finally, an optimisation model is proposed and developed to determine the best playing-11 using the computed performance ratings. The developed optimisation model suggests the playing-11 that maximises the possibility of winning against a given opponent. The proposed procedure to determine the playing-11 for an IPL match is demonstrated using past data from 2008-20. The demonstration indicates that for matches in the league stage, the suggested playing-11 by model and the actual playing-11 have a  $\sim 7\%$  similarity across all teams. The remaining  $\sim 3\%$  are different from those selected in the actual team. Nevertheless, this difference approximately yields a  $\sim$  Indian Premier League (IPL) is the most popular T20 domestic sporting league globally. Player selection is crucial in winning the competitive IPL tournament. Thus, team management select 11 players for each match from a team's squad of 15 to 25 players. Different player statistics are analysed to select the best playing 11 for each match. This study attempts an approach where the on-field player performance is used to determine the playing-11. A player's on-field performance in a match is computed as a single metric considering a player's attributes against every player present in the opposition squad. For this computation, past ball-by-ball data is cleaned and mined to generate data containing player-vs-player performance attributes. Next, the various performance attributes for a player-vs-player combination is converted into a player's performance rating by computing a weighted score of the performance attributes. Finally, an optimisation model is proposed and developed to determine the best playing-11 using the computed performance ratings. The developed optimisation model suggests the playing-11 that maximises the possibility of winning against a given opponent. The proposed procedure to determine the playing-11 for an IPL match is demonstrated using past data from 2008-20. The demonstration indicates that for matches in the league stage, the suggested playing-11 by model and the actual playing-11 have a  $\sim 7\%$  similarity across all teams. The remaining  $\sim 3\%$  are different from those selected in the actual team. Nevertheless, this difference approximately yields a  $\sim 13.32\%$  increase in performance rating compared to the existing team.  $3.32\%$  increase in performance rating compared to the existing team.

Keywords: Cricket, IPL, analytics, optimisation, simulation

---

\*Corresponding author: Malolan Sundararaman, Department of Management Studies, National Institute of Technology Tiruchirappalli, India. E-mail: malolan@nitt.edu.

## 1. Introduction

Cricket is a famous team sport. With 2.5 million followers, cricket is the second most followed sport globally (Shvili 2020). The sport has evolved over the years, and presently it is played in three formats at the international level. The game's longest format is called "Tests", which can last for up to 5 days. The second format is the "One Day" game, where each side can play for a maximum of 300 legal balls. The most recent and popular format is the shortest version of the game "Twenty-Twenty" (T20), in which each team is allowed to play a maximum of 20 overs (120 legal balls). There are several T20 competitions played globally throughout the calendar year. The T20 competitions can be international games or professional leagues. In an international game, the teams represent their respective country of origin. In leagues, teams represent franchises who have acquired their player services through an auction/contract.

The Indian Premier League (IPL) is one of the most competitive and the most attended cricket leagues globally. It is a professional T20 cricket league conducted by the Board of Control for Cricket in India (BCCI). BCCI founded the league in 2007. The brand value of the IPL in 2019 was Rs 475 Billion, compared to Rs 418 Billion in 2018 (Vhora 2019). IPL 2020 set a viewership record with 31.57 million Average Impressions, with an overall consumption increase of 23 per cent from the 2019 season. In 2020, IPL was ranked fifth in Google's global trends (Hindustan Times 2020). Therefore, it is a highly anticipated event for cricket fans in India and one of the most-watched sporting events globally.

The IPL is a franchise-based competition with eight teams, each representing an Indian city. When the tournament was first founded, an auction was held to determine the cities that the teams would be based in, as well as the owners of each team. In 2020, the eight cities/ franchises part of the IPL are Bangalore, Chennai, Delhi, Hyderabad, Jaipur, Kolkata, Mohali, and Mumbai (ESPN cricinfo 2020).

The IPL contests are between these eight franchises, each of which fields a single team. The IPL auction is a yearly event conducted by BCCI to auction cricket players to the various franchise. There are three types of players in any IPL team, Capped players, Uncapped players and Foreign or Overseas players. Capped players are Indian players who have represented the India Men seniors' team in any inter-

national game format at least once. Uncapped players are domestic players in the Indian Men first-class circuit. These players have never represented India at the international level. All non-Indian players fall into the category of foreign players, who can either be capped or uncapped in their respective country. Players in the common pool are sold to the highest bidder as per the IPL rules and regulations for forming squads and playing teams. The rules (as of the 2020 season) for forming a franchise's squad and team can be found in IPL T20 2021.

The IPL tournament is conducted in two stages, "League" and "Playoffs". Currently, each team plays every other team twice in a home-and-away round-robin format in the league stage. After the league stage, the top four teams will qualify for playoffs. The playoffs stage consists of four matches, "First Qualifier", "Second Qualifier", "Eliminator", and "IPL Final". The top two teams from the league stage will play against each other in the first qualifier. The winning team of the First Qualifier qualifies for the IPL Final. The Eliminator is played between the league stage's third and fourth place teams. The winner of the Eliminator and the losing team of the First Qualifier play each other in the Second Qualifier. The winning team of the Second Qualifier qualify for the IPL Final and face the winner of the First Qualifier. Finally, the winning team of the IPL Final is crowned "IPL Champions".

For the IPL, this study attempts to identify the best team playing 11 for the eight IPL franchises so that each franchise may have its best chance of being crowned champions. In this study, each player in the squad of an IPL franchise is evaluated based on past IPL performance. The past data for evaluating performance are the various player statistics such as strike rate, wickets, etc. These statistics are obtained for a player against all other players in opposing IPL franchise squads. Based on this performance evaluation, the best playing team is predicted for each match of the various franchise. This predicted playing team will be the most effective, with the highest possibility of winning a match against another IPL franchise's opposition team.

In the performance evaluation stage, each player's player statistics are used to determine ratings against the players in the opposition squad. The ratings are determined using weighted formulas. The rating is obtained for players in different roles (Batsman, Bowler, Wicketkeeper, Allrounder). Next, the playing-11 is selected through an optimisation model. Based on the number of players required in each spe-

cific role and the rating of players in that role, the model optimises to identify the team that generates the maximum overall rating. This ensures that the suggested playing-11 has the maximum possibility to win against the specific opposition.

The following section provides a closely related literature survey that establishes the newness of the focused problem. In Section 3, the problem statement of the current work is elaborately described. Section 4 explains the data collection & processing, model development and demonstration aspects. The results and managerial implications from the demonstration on the problem addressed in this study is discussed in Section 5. The paper concludes in Section 6.

## 2. Literature review

In general, the use of data for analysis of player performance was first demonstrated by Chadwick 1861. Recent literature on data-driven decisions for sports is generally referred to as “Sports Analytics” (Davenport 2014). These literature focuses on both on-field and off-field aspects of various sports. On-field analytics deals with improving the on-field performance of teams and players (Munir et al. 2015, Bose & Chakraborty 2019, Bowala et al. 2021 and Brydges 2021) or predicting the results as the event progress (Shah et al. 2016, Nimmagadda et al. 2018). Off-field analytics focuses on helping a franchise surface patterns and insights through data that would help increase ticket and merchandise sales (Howard & Crompton 2004, Cisyk & Courty 2021) and improve fan engagement (Zadeh 2021) and improve franchise performance in the event.

The literature on franchise performance for IPL tournaments has had good focus in recent times. A general review of literature on data mining schemes adopted in cricket is presented in Raju et al. 2020. Among the various literature on analytics in cricket, Lemmer et al. 2014, Jayalath 2018, Vistro et al. 2019 and Kapadia et al. 2019 address the issue of predicting the result of an IPL cricket match using historical data. To review the literature, the studies are grouped into two themes based on the focus as “Literature on Evaluating Player Performance” and “Literature on Match Result Prediction”.

### 2.1. Literature on evaluating player performance

The studies by Davis et al. 2015, Passi & Pandey 2018, Patel & Pandya 2019, and Santra et al. 2021

focus on predicting players’ performance. Davis et al. 2015 introduce a new metric of “expected run differential” for player evolution in T20 cricket. This metric measures the additional runs a player contributes to his team compared to a standard player. Here the standard player performs the same role as the player. The study computes this new metric individually for batsmen, bowlers, and all-rounders.

Passi & Pandey 2018 focus on the problem of determining a player’s performance. In the study, several past performance attributes are computed based on the experience of a given player. These attributes are developed for bowlers and batsmen separately. Analytical Hierarchy Process (AHP) is adopted to assign weights to attributes considered. A composite score is determined for bowlers and batsmen by combining the attribute score and the weights obtained from AHP. Next, supervised learning algorithms are utilised to develop prediction models. The models considered in the study include Naïve bays, decision trees, random forest, and multiclass support vector machines. These models predict the number of runs a batsman may score and the number of wickets a bowler may take. For a batsman, the predicted runs that may be scored is classified into five classes. Similarly, for bowlers, the predicted wickets that may be taken is classified into three classes. The study concludes that random forest models with a 90% train data set and 10% test data set are the best for predicting the number of runs (the number of wickets) a batsman (bowler) may take.

Patel & Pandya 2019 is similar to the previous study. It focuses on player performance prediction for IPL to choose players to create a playing-11 for a fantasy cricket league. In this study, the fantasy points system determines the consolidated score. Similar supervised learning models are used, and the fantasy score is predicted for a player. Based on the predicted fantasy score, the playing-11 is determined.

Santra et al. 2021 predict bowler ranking by evaluating the player’s past profile. Data from IPL games between 2008-2018 is utilised in supervised machine learning models to predict top bowlers. The performance of players in IPL 2019 is season is used to validate the models. In the study, a multivariate regression model is developed to determine a player’s rank. The various predictors for the model are determined using visualisations (target vs predictors). The study concludes that the proposed model accurately predicts the rank with minimal deviations.

## 2.2. Literature on match result prediction

In this section, the first group of studies Lemmer et al. 2014, Jayalath 2018, Vistro et al. 2019 and Kapadia et al. 2019 address the problem of only determining the result of the match using parameters other than the playing XI. Lemmer et al. 2014 investigate the prediction of the outcome of matches in a T20 series. Here the investigation is focused on improving the prediction rate by overcoming the inconsistent outcomes when two teams play in a series where a few matches are won by one team and a few by another. The study utilises data from IPL 2012 and improves the performance of predicting the winner of a match. Jayalath 2018 focuses on the comprehensive study of popular variables that affect the outcome of an ODI match. The study reveals the importance of home-field advantage. Further investigation leads to several insights on how “home field advantage” impacts the result based on the geographical location of the opponent.

Vistro et al. 2019 use machine learning algorithms to predict the winner of an IPL match before the beginning of a match. Here five machine learning models are utilised for prediction, adopting IPL data between 2008 to 2017. From the data, features are selected through visualisation. The study found that decision tree models are best suited for predicting the winner of a match.

Kapadia et al. 2019 investigate four machine learning algorithms to predict the result of a match using historical data. The data is for ten years of IPL matches from 2008 to 2017, along with the result for each match along with 16 other features. Feature selection techniques were used to reduce the number of features. The feature selection is performed for two categories of modules (1) Home field features and (2) Toss-winner decision features. The study concluded that models built using Toss-winner decision features perform better than models built using home-field features.

In contrast to the above studies on match result prediction, the study by Jayanth et al. 2018 is distinct and very closely related to the current work as it focuses on the recommendation of a squad for winning matches in a tournament based on past player performances. The squad is suggested to maximise the chance of winning. Additionally, the study also specifies the roles of different players recommended in the squad based on player performance. Here, data from the 2011 cricket world cup is utilised. The clear distinction between Jayanth et al. 2018 and the

current work is that the current work suggests the playing-11 from a squad of players when the squad of players in the opposition team is available.

From the above brief survey of related literature, to the best of our knowledge, no past study has focused on creating a playing-11 from a squad by considering the opposition squad in a cricket match.

## 3. Problem statement

Eight teams contest in the IPL as of 2021. This will be increased to 10 from 2022 (Nagraj Gollapudi 2021). From an IPL team perspective, the goal is to be crowned champions. This translates to winning as many matches as possible in the league stage to make it to the top four. Later, in the playoffs stage, a team must win all their games to be crowned champions. To this end, for every match, a franchise (which owns the team) should project the best playing 11 from their squad for that opposition team. This leads to the question of determining the best playing-11 with the maximum possibility of winning against the opposition.

The choice of best playing-11 must be made from the franchise's squad. The squad consists of 18 to 25 players. The 18-25 players can be of three types, overseas players or capped Indian players or uncapped Indian players. Of the 18 to 25 players, a maximum of 8 players can be overseas players (players who are non-Indian nationals). Moreover, a squad will consist of players of different categories. The player category can be a batsman, bowler or all-rounder. A batsman is a player who is primarily picked to score runs. A bowler is a player who is primarily picked to take wickets while not giving away too many runs. An all-rounder can either be a wicketkeeper-batsman or a bowling all-rounder. A wicketkeeper-batsman is a player who keeps wickets and bats. A bowling all-rounder is a player who bowls and bats. These player categories apply to all three types of players in a squad.

While creating a playing-11, it is necessary to ensure a balanced mix of player categories. This mix should score a good number of runs and defend the same by taking wickets. Some general rules of thumb in forming a playing-11 are as follows:

1. There should be at-least one wicketkeeper.
2. There should be at least three bowlers.
3. There should be at-least four batsmen.

Additionally, as there are different types of players in an IPL franchise squad, the following rules are enforced on a playing-11 by the IPL organisers.

1. A team must consist of eleven players, one of whom shall be captain.
2. Each captain shall nominate 11 players plus a maximum of 4 substitute fielders in writing to the IPL Match Referee before the toss.
3. Only those nominated as substitute fielders shall be entitled to act as substitute fielders during the match unless the IPL Match Referee, in exceptional circumstances, allows subsequent additions.
4. Each team may not name more than 4 Overseas players in its starting eleven for any match.
5. If the team names the maximum 4 Overseas players in its starting XI, an Overseas player may only take the field as a substitute fielder if he is replacing an Overseas player.

Thus, given these considerations, an approach to find the best playing-11 for a match would be to evaluate all the players in a squad against all the players in the opposition squad. This evaluation can be done based on the performance of a player. The performance parameters are exclusive for batsmen and bowlers. Based on their expertise, the performance parameter combinations are specified appropriately for all-rounders.

Therefore, to analyse player performances in the squad of a franchise against all players in squads of other franchises and generate the best playing-11 during a particular match, it is proposed to collect the ball-by-ball data of past IPL matches. This includes the matches played from 2007 till 2020. From this data, player vs player performance parameter metrics are computed for all players in the IPL. Using these computations, the best playing-11 for every match for every franchise is determined using the methodology proposed in the next section.

## 4. Methodology

### 4.1. Data collection and pre-processing

The ball-by-ball data for IPL games from 2007 to 2020 is compiled by scraping it from “cric-sheet.org”. According to the rules of IPL 2021, only 8 teams are participating in the tournament. The teams participating in the tournament are Chennai Super Kings (CSK), Delhi Capitals (DC), Kolkata

Knight Riders (KKR), Mumbai Indians (MI), Punjab Kings (PBKS), Rajasthan Royals (RR), Royal Challengers Bangalore (RCB), Sunrisers Hyderabad (SRH). However, the scraped ball-by-ball data had two teams (“Rising Pune Super Giants” and “Lucknow Super Giants”) which were created and dissolved during intermediate periods. Moreover, two teams had changed team names. Furthermore, the scenarios of retirement, player release, player exchanges and new bidding every year led to constant changes in the player mix. Therefore, the ball-by-ball data is cleared to include only players who played IPL 2020.

From the cleaned data, the attributes that are computed for a batsman, bowlers and all-rounders are presented in Table 1. The table also describes the attribute and its method of computation.

It is to be noted that some bowler attributes will be similar to batsman attributes in terms of computations (e.g., “Batsmen Total runs” is always equal to “Bowler Total runs”). Nevertheless, these are kept separate for the sake of easy communication. All the considered attributes are computed for three over-classes: “Power Play Overs (Overs 1 to 6)”, “Middle Overs (Overs 7 to 15)”, and “Slog Overs (Overs 16 to 20)”. Thus, the cleaned ball-by-ball data scraped from “cricsheet.org” is processed to compute the individual player-vs-player attribute-over class data. Since the tournament involves an auction system, players are constantly transferred between teams across the years. Therefore, in the processed data set, players are allotted to the teams they are part of in IPL 2020 with the same individual data.

Next, weights are assigned to each attribute-over-class combination. This is performed to obtain a consolidated “Rating Score” for Batting and Bowling statistics for every player-vs-player combination. The weights assigned to each attribute-over class combination is provided in Table 2. These weights are obtained based on discussion with cricket enthusiasts, analysts, and self-examinations. Here the weights assigned to attributes for each player category will add up to one in each over-class. Hence, the final data set constitutes the complete list of all player-vs-player combinations (for all players in squads of different IPL franchises) along with the respecting “Batting Rating Score” and “Bowling Rating Score”.

The “Rating Score” is the summation of “Batting Rating Score” and “Bowling Rating Score” for a player across all players in the squad of an opposition franchise. The above processing of ball-by-ball data to the final data set constituting the scores for

Table 1  
Attributes Considered for Batsmen, Bowler and All-Rounder

Player Category	Attribute	Attribute Description for this study	Attribute Computation Formula
Batsman	Batting Strike rate	Batting strike rate for the batsmen, against the opposition bowler	(Batsmen Total runs against a bowler/ Balls faced by batsmen against the same bowler) * 100
Batsman	Batsmen Total runs	Total runs scored by the batsman against the opposition bowler	Sum of all runs scored against a bowler
Batsman	Batsmen Average runs	Average runs scored by batsmen, against the opposition bowler	Batsmen Total runs / (number of innings batted against a bowler - number of not outs against the same bowler)
Batsman	Dot balls by batsmen	Number of balls faced by batsmen without scoring any runs against the opposition bowler	Count of all balls with 0 runs against a bowler
Batsman	1 s scored	Number of 1 run scored in a delivery, in total innings of a batsmen, against the opposition bowler	
Batsman	2 s scored	Number of 2 runs scored in a delivery, in total innings of a batsmen, against the opposition bowler	
Batsman	4 s scored	Number of 4 runs scored in a delivery, in total innings of a batsmen, against the opposition bowler	
Batsman	6 s scored	Number of 6 runs scored in a delivery, in total innings of a batsmen, against the opposition bowler	
Batsman	Balls faced by batsmen	Number of deliveries faced by the batsmen, against the opposition bowler, including wides and no-balls	Count of all balls faced by a batsman against a bowler
Bowler	Balls bowled	Number of deliveries bowled by the bowler, against the opposition batsmen	Count of all balls bowled by a bowler to a batsman
Bowler	Wickets taken	Number of wickets taken by the bowler, against the opposition batsmen	Count of number of times a bowler has got a batsman out
Bowler	Dot balls by bowler	Number of deliveries bowled by the bowler without conceding any runs, against the opposition batsmen	Count of balls with 0 runs when a bowler has bowled to batsman
Bowler	Bowler Total runs	Total number of runs conceded by the bowler, against the opposition batsmen	Sum of all runs conceded to a batsman
Bowler	Average 4 s for bowler	Average number of 4 s conceded by the bowler, against the opposition batsmen	Count of 4 s conceded to a batsman / Total number of balls bowled to the same batsman
Bowler	Average 6 s for bowler	Average number of 6 s conceded by the bowler, against the opposition batsmen	Count of 6 s conceded to a batsman / Total number of balls bowled to the same batsman

Table 2  
Different attributes and their weightage

Attributes	Weightage for Over 1 to Over 6	Weightage for Over 7 to Over 15	Weightage for Over 16 to Over 20
Batting Strike rate	0.15	0.20	0.25
Batsmen Total runs	0.30	0.35	0.30
Batsmen Average runs	0.15	0.15	0.10
Dot balls by batsmen	-0.10	-0.20	-0.30
1 s scored	0.03	0.04	0.02
2 s scored	0.07	0.08	0.05
4 s scored	0.15	0.16	0.25
6 s scored	0.22	0.20	0.30
Balls faced by batsmen	0.03	0.02	0.03
Balls bowled	0.25	0.35	0.40
Wickets taken	0.60	0.70	0.80
Dot balls by bowler	0.50	0.60	0.70
Total runs by bowler	-0.1	-0.25	-0.30
Average 4 s for bowler	-0.1	-0.15	-0.25
Average 6 s for bowler	-0.15	-0.25	-0.35

Table 3

Snippet of the final data set depicting the performance of V Kohli of RCB vs DI Chahar of CSK

Batsman	Bowler	Total Batting Ratings	Total Bowling Ratings
V Kohli	DL Chahar	72.348	11.183

Table 4

Rating Score for RCB Squad players when performing against CSK squad players

S. No	RCB Players	Total ratings against CSK
1	A Zampa	2.075
2	AB de Villiers	645.080
3	AJ Finch	345.564
4	CH Morris	266.795
5	D Padikkal	141.375
6	DW Steyn	117.352
7	Gurkeerat Singh	180.049
8	I Udana	5.200
9	M Ali	283.492
10	Mohammed Siraj	45.553
11	N Saini	115.540
12	S Dube	176.040
13	UT Yadav	80.039
14	V Kohli	1009.314
15	Washington Sundar	235.953
16	YS Chahal	116.178
17	Philippe	0
18	Shahbaz Ahmed	0

a complete list of all player-vs-player combinations is done in Python. The interested reader may refer to the Python script provided in Annexure-1.

For the benefit of readers, a snippet from the final data depicting the performance of V. Kohli of RCB vs DI Chahar of CSK is presented in Table 3. Table 4 provides the “Rating Score” for all players in the RCB squad, based on their individual performance against players in the CSK squad.

#### 4.2. Model development and demonstration

For a match, the strongest playing-11 of a franchisee must be selected based on the opposition. If the opposition playing-11 is known, then the best 11 players from the franchise squad against the 11 opposing players can be chosen. However, only the opposition squad and not the opposition playing-11 is known. Thus, the franchise playing-11 must

be selected such that each player in the playing-11 has performed better than all other non-playing-11 players in the franchise squad against the entire combination of players in the opposition squad. Additionally, there are also requirements on the number of players required in each roll (batsman, bowlers, all-rounders, and wicketkeeper) in the playing-11. Therefore, for determining the playing-11, the following binary integer linear programming model is proposed.

The sets, uncontrollable variables and decision variables for the model are as specified below:

*Sets*

1.  $i$ : Set of players in the squad of the current franchise  $[1, \dots, n]$
- $j$ : Set of all opposition franchise  $[1, \dots, m]$

#### Uncontrollable Variables

$R_{ij}$  – Rating Score for player- $i$  when performing against player- $j$

$F_i$  – 1 if player- $i$  is a foreign player, 0 otherwise

$B_i$  – 1 if player- $i$  is a batsman, 0 otherwise

$O_i$  – 1 if player- $i$  is a bowler, 0 otherwise

$A_i$  – 1 if player- $i$  is an all-rounder, 0 otherwise

$W_i$  – 1 if player- $i$  is a wicketkeeper, 0 otherwise

$NB_j$  – Number of batsmen required in the playing-11 against franchise- $j$

$NO_j$  – Number of bowlers required in the playing-11 against franchise- $j$

$NA_j$  – Number of all-rounders required in the playing-11 against franchise- $j$

$NW_j$  – Number of wicket keepers required in the playing-11 against franchise- $j$

$MF_j$  – Maximum number of foreign players permitted in the playing-11 against franchise- $j$

#### Decision Variable

$$X_{ij} = \begin{cases} 1, & \text{if player } - i \text{ is in playing } - 11 \text{ against franchise } - j \\ 0, & \text{otherwise.} \end{cases}$$

Since the rating score depicts how successful a player has been against a given player, the objective is to maximise the score for all players in the playing-11. Accordingly, this objective is mathematically described in (1).

$$\text{Objective} = \text{Maximize} \sum_{i=1}^n \sum_{j=1}^m (R_{i,j} * X_{i,j}) \quad (1)$$

The above objective is maximised under the conditions that there must be at least  $NB_j$ ,  $NO_j$ ,  $NA_j$ ,  $NW_j$  number of batsmen, bowlers, all-rounders, and wicket keepers respectively. Among these,  $NW_j$  is usually 1. Thus, these are all “greater than or equal to” constraints. Additionally, there is also the IPL restriction that no team can feature more than 4 four players ( $MF_j = 4$ ) in their playing-11. The above constraints are represented mathematically in (2) to (6). Finally, a requirement that the playing-11 must have 11 players is enforced through the equality constraint (7).

$$\sum_{i=1}^n [X_{ij} * B_j] \geq NB_j, \forall j \in [1, \dots, m] \quad (2)$$

$$\sum_{i=1}^n [X_{ij} * O_j] \geq NO_j, \forall j \in [1, \dots, m] \quad (3)$$

$$\sum_{i=1}^n [X_{ij} * A_j] \geq NA_j, \forall j \in [1, \dots, m] \quad (4)$$

$$\sum_{i=1}^n [X_{ij} * W_j] \geq NW_j, \forall j \in [1, \dots, m] \quad (5)$$

$$\sum_{i=1}^n [X_{ij} * F_j] \leq MF_j, \forall j \in [1, \dots, m] \quad (6)$$

$$\sum_{i=1}^n X_{ij} = 11, \forall j \in [1, \dots, m] \quad (7)$$

To generate the model for any given data set, a LINGO Set Code is developed. The developed LINGO Set code is presented in Annexure-2. The final data set created in Section 4.1 is fed to the LINGO Set code and solved in a LINGO 11 solver to demonstrate the model. The model is solved to optimality.

The total number of scheduled matches in the league stage of the IPL 2020 tournament is 56. Each

match involves two franchises playing against each other. The complete solution report containing the playing-11 for each franchise against each opponent is specified in Annexure-3. This result demonstrates the workability of the model. The following section discusses the results in depth.

## 5. Results

For discussion, Table 5 summarises the suggested and actual playing-11 for RCB against all other franchises in 2020. First, to establish the performance parameters of the proposed methodology, the suggested playing-11 for RCB is compared with the actual playing-11 of RCB. This will establish the confidence of the algorithm to capture the current working of the team management. Additionally, the performance of those players who were not suggested but featured in the playing-11 is compared with the projected match performance. This analysis is presented in Table 6.

From Table 5 the following points are observed:

- There are five core players played by RCB in every match, AB de Villiers, D Padikkal, V Kohli, Washington Sundar and YS Chahal. Except for D Padikkal, Washington Sundar and YS Chahal, the model chose the rest of the core players. This implies that the model captures 40% of the core players.
- D Padikkal is preferred over AJ Finch, despite having a rating that is lower by 1116.69 rating points when totalled over all opposition teams for the same role as Batsmen. Similarly, Washington Sundar is preferred over CH Morris, despite having a rating score which is lesser by 518.31 for the same role of All-rounder when computed over all opposition squads. Both might be due to the constraint of 4 foreign players in a playing-11 team.
- AB de Villiers and V Kohli are the only players who were predicted and played all the matches in 2020.
- M Ali and UT Yadav are not picked in the actual team. This is despite the model suggesting them for 4 and 6 matches, respectively. Even having overall ratings of 1421.40 for 10 matches, they were picked for only 2 matches overall.

From Table 6, it is observed that.



Table 5  
The “Actual” and “Suggested” Playing-11 For RCB against Different Opposition IPL Franchise

RCB Squad	Role	Playing-11 (And Player Type) For RCB Against Different Opposition Franchise													
		CSK		KKR		MI		DC		RR		PBKS		SRH	
		Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested
A Zampa	Bowl	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AB de Villiers	WK	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AJ Finch	Bat	1	1	1	1	0	1	0	1	1	1	1	1	0	1
CH Morris	AR	1	1	1	1	1	1	1	1	1	1	0	1	1	1
D Padikkal	Bat	1	1	1	1	1	1	1	0	1	1	1	1	1	1
DW Steyn	Bowl	0	0	0	0	1	0	0	0	0	0	1	0	0	1
Gurkeerat Singh	Bat	1	1	1	1	1	0	0	1	1	1	0	1	1	1
I Udana	Bowl	0	0	1	0	0	0	1	0	1	0	0	1	1	0
M Ali	AR	1	1	0	1	0	0	0	1	0	1	0	0	0	0
Mohammed Siraj	Bowl	1	0	1	1	1	0	1	1	0	1	0	0	1	1
N Saini	Bowl	1	1	1	1	0	1	0	1	1	0	1	1	1	1
S Dube	AR	0	0	0	0	1	1	1	1	0	0	1	0	0	0
UT Yadav	Bowl	0	1	0	1	0	1	0	0	0	1	1	1	0	1
V Kohli	Bat	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Washington Sundar	AR	1	1	1	0	1	1	1	1	1	1	1	1	1	1
YS Chahal	Bowl	1	1	1	1	1	1	1	1	1	1	1	1	1	0
Philippe	Bat	0	0	0	0	1	1	1	0	0	0	1	0	1	0
Shahbaz Ahmed	AR	0	0	0	0	0	0	1	0	1	0	0	0	0	0

Key: AR → All Rounder, Bat → Batsman, Bowl → Bowler, WK → Wicketkeeper.

Table 6  
Comparison of Projected Match Performance for “Suggested” and “Actual” Playing-11 For RCB against Different Opposition IPL Franchise

Playing Position in RCB-XI	Ratings for RCB Playing-11 Against Different Opposition Franchise													
	CSK		KKR		MI		DC		RR		PBKS		SRH	
	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested	Actual	Suggested
1	645.08	645.08	699.52	699.52	777.66	777.66	798.56	798.56	856.95	856.95	598.22	598.22	930.12	930.12
2	345.56	345.56	467.25	467.25	376.28	404.20	133.06	302.79	284.11	284.11	299.89	299.89	372.73	576.14
3	266.80	266.80	153.52	153.52	395.16	376.28	197.60	133.06	280.89	280.89	94.26	235.37	221.49	372.73
4	141.38	141.38	199.07	199.07	100.03	395.16	33.94	216.71	314.30	314.30	59.39	94.26	374.62	221.49
5	180.05	180.05	164.46	164.46	65.06	29.91	67.77	65.02	234.00	234.00	106.58	21.68	5.45	103.38
6	283.49	283.49	3.50	228.87	26.81	177.79	224.26	67.77	5.10	103.46	156.82	101.75	108.08	374.62
7	45.55	115.54	31.50	31.50	177.79	141.74	600.11	118.77	36.19	88.48	74.82	106.58	78.00	108.08
8	115.54	80.04	29.05	29.05	669.80	669.80	168.61	224.26	833.45	168.98	413.88	74.82	716.06	78.00
9	1009.31	1009.31	679.21	44.87	139.58	139.58	114.11	600.11	182.29	833.45	322.74	413.88	226.87	230.11
10	235.95	235.95	24.29	679.21	61.52	61.52	37.31	168.61	75.64	182.29	57.26	322.74	75.46	716.06
11	116.18	116.18	67.14	67.14	203.50	203.50	34.00	114.11	1.40	75.64	-0.21	57.26	122.77	226.87
Total Ratings	3384.90	3419.38	2518.51	2764.46	2993.20	3377.13	2409.34	2809.77	3104.31	3422.55	2183.67	2326.46	3231.64	3937.59
% Change in Total Rating	1.02%		9.77%		12.83%		16.62%		10.25%		6.54%		21.84%	

1. The overall ratings of the team would have increased by 2231.79 ratings in the tournament's league stage if RCB had chosen the suggested Playing-11. This is a  $\sim 11.25\%$  increase in the team's rating.
2. For all opponents, the suggested playing-11 has a better overall rating than the actual playing-11 of RCB when the same team combination is considered for making the suggestions.
3. If the suggested playing-11 is played, then the minimal increase in overall rating would have been  $\sim 1.01\%$  (against CSK), and the maximum increase would have been  $\sim 21.84\%$  (against SRH).

The above points give significant insight into the composition of the team. A similar analysis can be performed for all teams. In general, for all matches in the league stage, the suggested playing-11 and the actual playing-11 have a 75.32 % similarity across all teams. Moreover, the suggested playing-11 on average has 11.25% more rating than the actual playing-11. This indicates the importance of determining playing-11 based on the opposition squad.

## 6. Conclusion

Sports analytics is an exciting avenue for research. In this work, the problem of determining the playing-11 from a side's squad in a cricket match given an opposition squad is addressed. This problem is explored from the perspective of player performance and is illustrated for the Indian Premier League (IPL). The squad size of an IPL franchise can vary from 18 to 25 players. Moreover, each squad will consist of players of different categories (batsman, bowler, all-rounder, wicketkeeper) and types (Foreign, Capped Indian, Uncapped Indian). A performance-based approach is proposed to select a playing-11 from the given squad.

Accordingly, a player's performance is measured using performance attributes. Fifteen performance attributes are considered across the different player categories in this study. Each attribute is measured for three over classes of an IPL game: power-play, middle-overs, and slog-overs. To compute the attributes (in every over class), ball-by-ball past data from IPL is utilised, and each attribute is computed for a player-vs-player combination. Next, weights are assigned to each attribute (computed for every over class) to arrive at an overall rating for a player-

vs-player combination. Thus, this computation will result in information that can be utilised to compare a player against another player.

With this information, given an opposition squad, the best playing-11 from the current squad is determined. The determination is done using a developed optimisation model. The optimisation model operates with an objective to maximise the overall rating from the chosen playing-11. This objective is constrained based on the IPL selection rules and the team management decision on team combination. The IPL selection rules impose restrictions on the maximum number of foreign players selected in a playing-11. The team combination specifies the number of batsmen, bowlers, all-rounders, and wicketkeepers to be included in the playing-11. For the IPL data, a demonstration taking the Royal Challengers Bangalore (RCB) squad as an example proved that a playing-11 created with the above approach would, on average, increase the overall team rating by approximately 11.25%. Moreover, considering the overall scenario across all league matches, 75% of the suggested playing-11 and actual playing-11 are in sync for RCB. Furthermore, in the league stage, across all teams, if the suggested playing-11 would, on average, increase a team's rating by  $\sim 13.32\%$ . This indicates that the proposed models also capture the team management's current work.

Nevertheless, the current study is limited as the performance rating is computed considering all past IPL data. It may be interesting to see how recent performance and form affect the selection. This would require the distribution of attributes weights for recent performance and overall performance. Also, the selection of players in the playing-11 is made considering only IPL encounters against the opposition players. Though this seems intuitively aggregable, future research can explore the inclusion of all past encounters between players to estimate performance ratings and then explore the playing-11 selection. Another limitation is the consideration of "Pinch Hitting" where a player is sent to deliberately hit every ball and not worry about the fall of wickets. This is not captured in the current study. Additionally, when choosing a player for the role of "wicketkeeper", the current work only considers the player's "Batting Ability". However, the ability of the player to keep wickets effectively (attributes such as "stumpings", "catches taken/ dropped", etc.) are critical and need to be considered. Finally, the current study makes a player selection only using "On-Field" performance attributes. The "Off-field" performance characteris-

tics are not considered. Therefore, future research can explore extensions of the current work to consider these limitations.

## Supplementary material

The Annexure part is available in the electronic version of this article: <https://dx.doi.org/10.3233/JSA-220638>.

## References

- Bose, D. & Chakraborty, S., 2019. Managing In-play Run Chases in Limited Overs Cricket Using Optimized CUSUM Charts, *Journal of Sports Analytics*, 5(4), DOI: 10.3233/JSA-190342
- Bowala, S.M.B., Manage, A.B.W. & Scariano, S.M., 2021. Modeling T20I cricket bowling effectiveness: A quantile regression approach with a Bayesian extension, *Journal of Sports Analytics*, 07(3), DOI: 10.3233/JSA-200556
- Brydges, C.R. 2021, Analytics of batting first Indian Premier League twenty20 cricket matches, SportRxiv, DOI: <https://doi.org/10.31236/osf.io/jq564>
- Chadwick, H. 1861, 'Beadle's Dime Base Ball Player', in H Chadwick (ed.), proceedings of the fifth annual base-ball convention, Beadle and Company, 141 William St., Northern Illinois University, United States, <https://dimenovels.lib.niu.edu/islandora/object/dimenovels%3A6302#page/1/mode/1up>
- Cisyk, J. & Courty, P., 2021, 'Stadium Giveaway Promotions: How Many Items to Give and the Impact on Ticket Sales in Live Sports', *Journal of Sport Management*, 35(6), DOI: <https://doi.org/10.1123/jism.2020-0322>
- Davenport, H. 2014, What Businesses Can Learn From Sports Analytics, MIT Sloan Management Review, month of publication June, viewed 21 Dec 2021, <http://mitsmr.com/1h4FHgs>
- Davis, J., Perera, H. & Swartz, T.B., 2015, Player evaluation in Twenty20 cricket, *Journal of Sports Analytics*, 1(1), DOI:10.3233/JSA-150002
- ESPN cricinfo 2020, Squads, ESPN cricinfo, n.d., viewed 21 December 2021, <https://www.espnricinfo.com/series/ipl-2020-21-1210595/squads>
- Gollapudi, N., 2021, IPL to become 10-team tournament from 2022', ESPN cricinfo, 31 August, viewed 21 December 2021, <https://www.espnricinfo.com/story/ipl-to-become-10-team-tournament-from-2022-1275505>
- Hindustan Times 2020, 'IPL emerges as top Google trend of 2020 in India, ranked fifth globally', Hindustan Times, 10 December, viewed 21 December 2021, <https://www.hindustantimes.com/cricket/ipl-emerges-as-top-google-trend-of-2020-in-india-ranked-fifth-globally/story-1YiwMUXKJaaM0vVKICeP.html>
- Howard, D.R. & Crompton, J.L., 2004, Tactics used by sports organisations in the United States to increase ticket sales, *Managing Leisure*, DOI: 10.1080/13606710410001709617
- IPL T20 2021, IPL 2021 Match Playing Conditions, IPL T20, 15 October, viewed 21 December 2021, <https://www.iplt20.com/about/match-playing-conditions>
- Jayalath, K.P., 2018, A machine learning approach to analyse ODI cricket predictors, *Journal of Sports Analytics*, 04(1), DOI: 10.3233/JSA-17175
- Jayanth, S.B., Anthony, A., Abhilasha, G., Shaik, N. & Srinivasa, G., 2018, A team recommendation system and outcome prediction for the game of cricket, *Journal of Sports Analytics*, 4(4), DOI: 10.3233/JSA-170196
- Kapadia, K., Abdel-Jaber, H., Thabtah, F. & Hadi, W., 2019, Sport analytics for cricket game results using machine learning: An experimental study, *Applied Computing and Informatics*, DOI: <http://doi.org/10.1016/j.aci.2019.11.006>
- Lemmer, H.H., Bhattacharjee, D. & Saikia, H. 2014, A Consistency Adjusted Measure for the Success of Prediction Methods in Cricket, *International Journal of Sports Science & Coaching*, 09(3), DOI: <https://doi.org/10.1260/1747-9541.9.3.497>
- Munir, F., Hasan, Md.K., Ahmed, S. & Md. Quraish, S., 2015, Predicting a T20 cricket match result while the match is in progress, BSc thesis, BRAC University, Bangladesh, <http://hdl.handle.net/10361/4372>
- Nimmagadda, A., Kalyan, N.V., Venkatesh, M., Teja, N.N.S. & Raju, C.G., 2018, Cricket score and winning prediction using data mining, *International Journal of Advance Research and Development*, 03, <https://www.ijarnd.com/manuscripts/v3i3/V3I3-1230.pdf>
- Passi, K. & Pandey, N., 2018, Increased Prediction Accuracy In The Game Of Cricket Using Machine Learning, *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, 08(2), DOI: 10.5121/ijdkp.2018.8203
- Patel, N., & Pandya, M., 2019, IPL Player's Performance Prediction, *International Journal of Computer Sciences and Engineering*, 07(5), DOI: <https://doi.org/10.26438/ijcse/v7i5.478481>
- Raju, V.S., Sethi, N. & Rajender, R., 2020, A Review of Data Analytic Schemes for Prediction of Vivid Aspects in International Cricket Matches', 5th International Conference on Computing Communication Control and Automation (ICCUBEA), DOI: 10.1109/ICCUBEA47591.2019.9128835
- Santra, A., Mitra, A., Sinha, A. & Das, A.K., 2021, Prediction of Most Valuable Bowlers of Indian Premier League (IPL), 4th International Conference on Data Management, *Analytics & Innovation*, 02, DOI: <https://doi.org/10.1007/978-981-15-5619-7>
- Shah, A., Jha, D. & Vyas, J., 2016, Winning and Score Predictor (Wasp) Tool, *International Journal of Innovative Research in Science and Engineering*, 02(6), <http://www.ijirse.com/wp-content/upload/2016/02/346ijirse.pdf>
- Shvili, J. 2020, The Most Popular Sports In The World, World Atlas, 16 October, viewed 21 December 2021, <https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html>
- Vhora, F.B. 2019, Hitting new boundaries: Brand IPL valued at Rs 47,500 crore, up 13.5% YoY, CNBC TV18, 19 September, viewed 21 December 2021, <https://www.cnbc18.com/sports/hitting-new-boundaries-brand-ipl-valued-at-rs-47500-crore-up-13-5-yoy-4382191.htm>

Vistro, D.M., Rasheed, F. & David, L.G., 2019, 'The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics, *International Journal of Scientific & Technology Research*, 08(9), <https://www.kclas.ac.in/wp-content/uploads/2021/01/The-Cricket-Winner-Prediction-With-Application-Of-Machine-Learning-And-Data-Analytics-.pdf>

Zadeh, A.H., 2021, Quantifying fan engagement in sports using text analytics, *Journal of Data, Information and Management*, DOI: <https://doi.org/10.1007/s42488-021-00052-4>