

# Cricket Team Selection and Analysis by Using DEA Algorithm in Python

V.Rajyalakshmi<sup>1</sup>, M.S. Vani<sup>2</sup>, M.Narendra<sup>3</sup>, M.Pavan Kumar<sup>4</sup>, D.Charan Kumar Reddy<sup>5</sup>, P.vennela<sup>6</sup>

<sup>1,2</sup> Assistant Professor, <sup>3,4,5,6</sup>B.Tech Students  
Computer Science and Engineering,

Mother Theresa Institute Of Engineering and Technology, Palamaner, Andhra Pradesh, India

\*\*\*

**Abstract :** A new method for cricket team selection using Data Envelopment Analysis (DEA) is proposed. DEA algorithm is proposed, which is the formulation for evaluation of cricket players in different capabilities using multiple outputs. This evaluation process determines efficient and inefficient cricket players and ranks them on the basis of DEA scores. The ranking can be used to decide the required number of players for a cricket team in each cricketing capability. A real dataset 2018 year data, Cricket players having various capabilities is used to choose the best cricket team. The estimated method has the advantage of considering multiple factors related to the performance of players in multiple capabilities collected from ESPN cricinfo and aggregates their scores using a linear programming DEA model. This DEA Aggregation gives the scores of players objectively instead of using subjective computations. The estimated DEA method can be used to form a national cricket team from several clubs or a team of top cricketers.

**Key Words:** Data Envelopment Analysis (DEA), Entertainment and Sports Programming Network (ESPN), Linear Programming

## 1. INTRODUCTION

Studies addressing different research issues related to various dimensions of the cricketing sport can be found in the literature. The studies are limited in numbers and few of them have addressed some issues of cricket using graphical methods. The batting strategies in limited over cricket are discussed in Preston and Thomas. Cricket is a relatively new and promising research area in comparison with other sports such as baseball, soccer, etc. Cricket is an immensely popular sport in the Indian subcontinent (India, Pakistan, Sri Lanka and Bangladesh) and its popularity sometimes affects other sports. Crickets newest innovation is Twenty20, essentially an evening entertainment also known as T20. Based on the popularity of the Twenty20 format, the International Cricket Council (ICC) organized a Twenty20 World Cup in 2007, which was played in South Africa and won by India. Now, India is being considered as the hub of international cricket in the world. To those not familiar with cricket, Preston and Thomas (2000) have introduced the notion of limited over cricket in detail.

### 1.1. PROBLEM DEFINITION

Considering multiple factors related to the performance of players simultaneously and comparing the players, with different capabilities, based on these outcomes needs an aggregation method. The contribution of proposed work is to propose a DEA model, which can be used to find efficiency of cricket players without using subjective opinion. Also, for improving the performance of inefficient players the proposed DEA method can suggest solutions. One important issue related to contribution of proposed work is the first attempt of using DEA method to form a cricket team from players of different clubs.

## 2. EXISTING SYSTEM

- ❖ Although, there are various similarities between limited over cricket and Twenty20 cricket but some important differences are pointed out in Bhattacharya, Gill, and Swartz (2011).
- ❖ Existing study on cricket team selection by Brettenny, Gonsalves, it is often assumed that selectors choose a team of 11 players.
- ❖ Ahmed, Jindal, and Deb (2011) proposed a method of cricket team selection from the perspective of multi-objective genetic optimization whereas Iyer and Sharda (2009) used a neural network approach in cricket team selection.
- ❖ Sharp et al. (2011) and Lemmer (2011) suggested an integer programming model in cricket team selection.

## DISADVANTAGES

- ❖ Work by Brettenny has drawback, it is common to choose a team of 15 players in order to provide flexibility to the captain and the coach of the cricket team. Selection of a cricket team under various constraints such as number of batsmen, bowlers, all-rounder and a wicket keeper is a complex task. This task can be made easier using mathematical modelling.

## PROPOSED SYSTEM

- ❖ A novel application of Data Envelopment Analysis (DEA) is proposed for team selections in cricket.
- ❖ A DEA model, which is a linear programming technique, is used for evaluation of cricket players in different abilities.
- ❖ A real dataset of cricket players from Indian Premier League 11 (IPL 2018) in different capabilities is used to choose the best cricket team.
- ❖ Proposed work first attempt using DEA method for cricket team selection, however, DEA has been applied to team sports particularly to baseball.

## ADVANTAGES

- ❖ This provides a way to choose the required number of players for a cricket team.
- ❖ The proposed method finds efficient and inefficient players for each capability and ranks them on the basis of DEA scores.

## 3. MODULES

- ❖ Team Selection
- ❖ Data Collection
- ❖ Finding one year average
- ❖ Data Envelopment Analysis

### Team Selection:

One important issue in the cricket team selection is the way of aggregation of multiple performance measures related to players' available statistics. This paper deals with the benefit of introducing a linear programming based DEA model for the aggregation. This avoids the inclusion of subjective computation for finding players' scores. In previous sections, we obtained top best cricket players who have the highest performance. According to the results, finding a cricket team is selected based on efficiency score arrived from above module.

### Data Collection

The dataset is collected with 15 number of players who is available in current. The dataset is collected from espnricinfo.com, in which all the 15 players one year statistics data is considered. The three types of data collected are for batsman, bowler and allrounder. The players given for dataset collection is shown below.

Table -1: Players list

	A	B	C	D
1	player	id	country	type
2	Rohit Sharma	34102	India	Batsman
3	Ambati Rayudu	33141	india	Batsman
4	Shikhar Dhawan	28235	india	Batsman
5	Virat Kohli (c)	253802	India	Batsman
6	MS Dhoni (wk)	28081	India	WK-Batsman
7	Kedar Jadhav	290716	India	Batsman
8	Hardik Pandya	625371	India	Allrounder
9	Bhuvneshwar Kum	326016	India	Bowler
10	Kuldeep Yadav	559235	India	Bowler
11	Yuzvendra Chahal	430246	India	Bowler
12	Jasprit Bumrah	625383	India	Bowler
13	Umesh Yadav	376116	India	Bowler
14	Mohammed Sham	481896	India	Bowler
15	Ravindra Jadeja	234675	India	Allrounder

**Finding one year average:**

From the above module, the players one year data is collected from espnricinfo.com, in which all the 15 players one year statistics data is considered.

**BATSMAN**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Name	Mat	Inns	NO	Runs	BF	SR	100	50	0	4s	6s	Ave
2	Rohit Shar	19	19	5	1030	1029	100.1	5	3	1	104	39	73.57
3	Ambati Ra	11	10	3	392	432	90.74	1	3	0	31	11	56
4	Shikhar Dh	19	19	1	897	877	102.3	3	2	0	127	11	49.83
5	MS Dhoni	20	13	2	275	385	71.42	0	0	1	19	2	25
6	Kedar Jadh	11	5	3	87	90	96.66	0	0	0	6	2	43.5

Table -1: List of Batsman's

**BOWLERS**

	A	B	C	D	E	F	G	H	I	J	K	L
1	Name	Mat	Inns	Overs	Mdns	Runs	Wkts	Ave	SR	4	5	Econ
2	Bhuvneshwar Kumar	14	14	101	6	537	11	48.8	55	0	0	5.31
3	Kuldeep Yadav	19	19	172.2	4	800	45	17.8	22.9	2	1	4.64
4	Yuzvendra Chahal	17	17	156.1	3	754	29	26	32.3	1	1	4.82
5	Jasprit Bumrah	13	13	100.5	9	366	22	16.6	27.5	1	0	3.62
6	Umesh Yadav	4	4	39.5	0	275	4	68.8	59.7	0	0	6.9
7	Mohammed Shami	2	2	20	0	140	3	46.7	40	0	0	7

Table -2: List of bowlers

**ALLROUNDER**

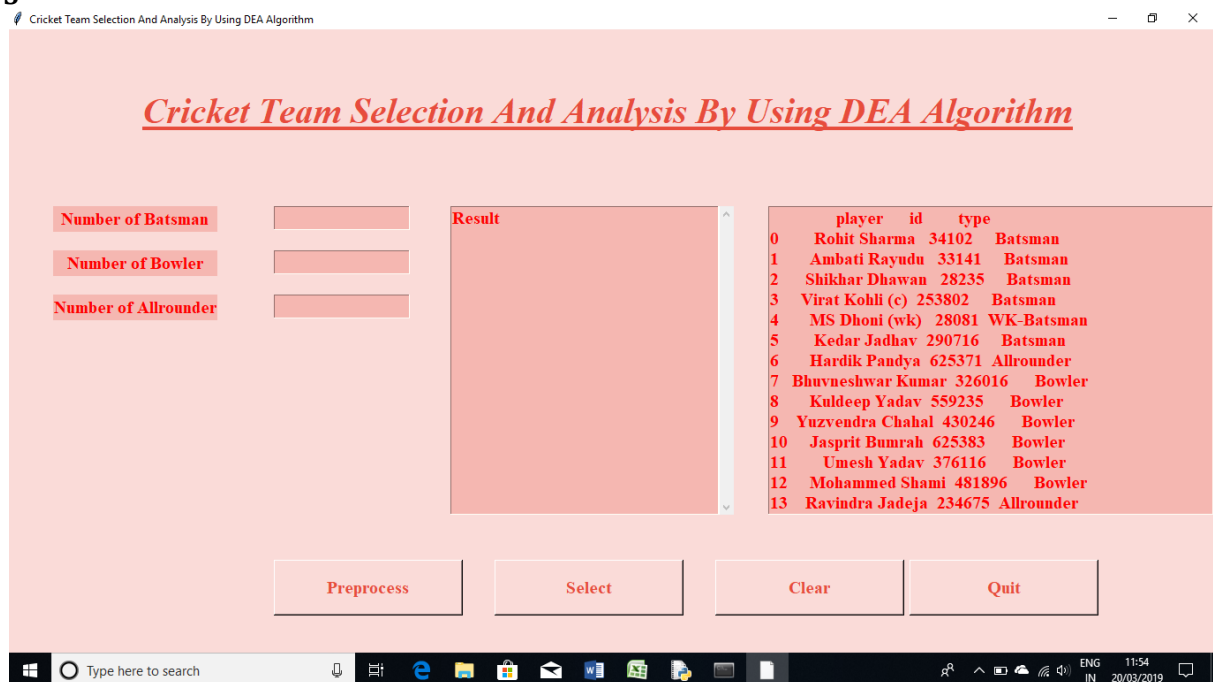
	A	B	C	D	E	F	G	H	I	J	K	L
1	Name	Mat	Runs	HS	100	Wkts	Bowl Av	5	Ct	St	Ave Diff	Bat Av
2	Hardik Pandya	10	68	21	0	5	79.2	0	7	0	-65.59	13.6
3	Ravindra Jadeja	8	68	25	0	14	24.57	0	0	0	-1.9	22.66

**Table -3:** List of Allrounders

**Data Envelopment Analysis:**

- In this study, we consider five important measures of batting statistics such as highest individual score (HS), average batting performance, strike rate (S/R), numbers of fours (4s) and number of sixes (6s). These parameters have been taken from dataset. These factors are DEA outputs for each batsman and can be defined as follows:
- HS: For a batsman, the highest individual score (HS) is the maximum number of runs scored in one match during a tournament.
- Average: The average batting performance is expressed by  $R/m$  where R denotes the number of runs scored and m the number of times the batsman was out.
- S/R: The batting strike rate can be expressed as  $R/b$  where R denotes the number of runs scored and b denotes the number of balls faced by a player.
- 4s: The number of 4s hit by a batsman during matches in 2018.
- 6s: The number of 6s hit by a batsman during matches in 2018

**4. RESULTS**



**Fig -1:** Home Page



Fig -2: Giving Input

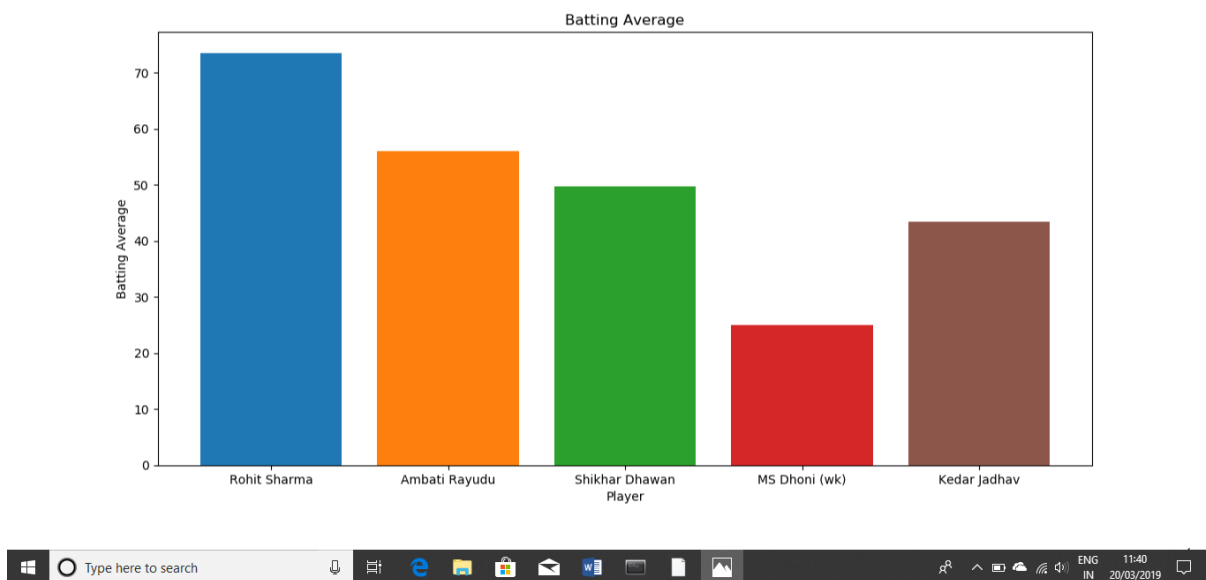


Fig -3: Batsman Average

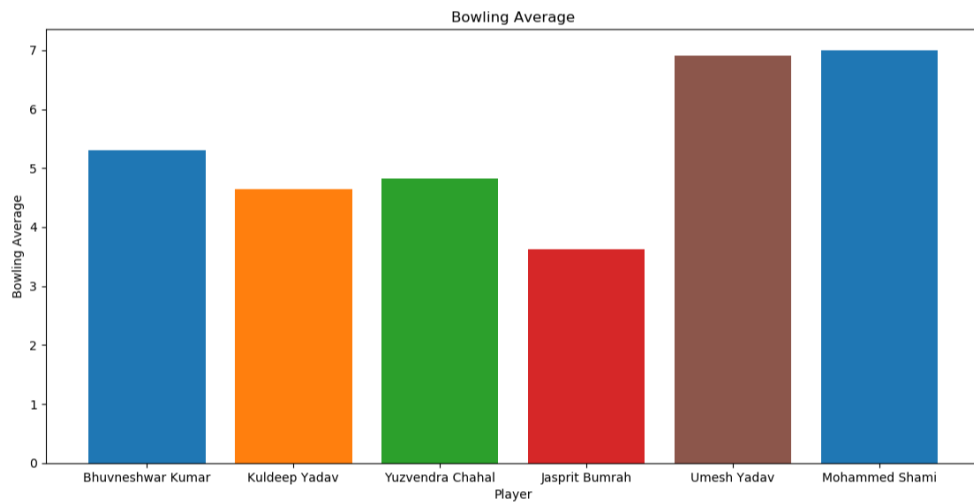


Fig -4: Bowlers economy

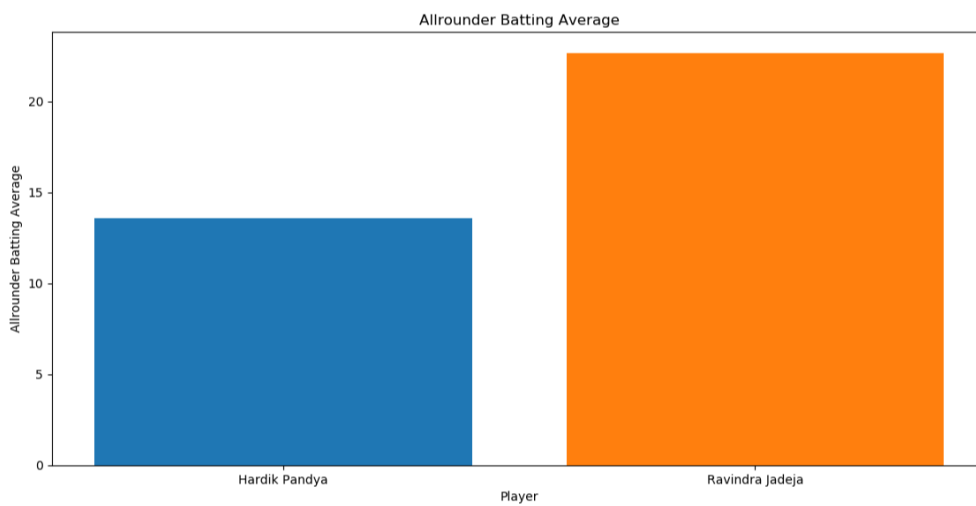


Fig -5: Allrounders Average and economy



Fig -6: Final Selected Team

## 5 CONCLUSIONS

The proposed work suggested a novel application of DEA for selection of a cricket team. We proposed a multiple outputs DEA formulation to assess cricket players indifferent capabilities. A real dataset consisting of cricket players in for year 2018 is considered to find the best cricket team. The proposed method has the benefit of using an optimization DEA model for aggregation of multiple performance measures. This avoids the inclusion of subjective computation in the cricket team selection. This project is the first attempt to form a cricket team from players who played in a tournament for different clubs. Further research needs to compare and evaluate the performance of a cricket team from performance of individual team's players.

## 6 FUTURE ENHANCEMENT:

In this project in the input should be limited that is the total batsmen, bowler and all rounder input should be 10. This is the drawback in this project, In future we can overcome this drawback

## REFERENCES

- [1] Jblonsky, j.,2016,Ranking models in Data Envelopment Analysis
- [2] Cooper, w., seiford, l. and tone, k., 2007 Data Envelopment Analysis. A comprehensive text with models, and DEA-solver software
- [3] <http://araith.github.io/pyDEA/>
- [4] <http://www.python.org>

**Authors**

V. Rajya Lakshmi,  
Assistant Professor,  
Dept. of CSE, MTIET, Palamaner.  
rajyalv@gmail.com



M.S. Vani,  
Assistant Professor,  
Dept. of CSE, MTIET, Palamaner.  
msvani1992@gmail.com



M.Narendra,  
15HR1A0573,  
Dept. of CSE, MTIET, Palamaner.  
maraninarendra@gmail.com



M.Pavan Kumar,  
15HR1A0571,  
Dept. of CSE, MTIET, Palamaner.  
Pavankumar.m1432@gmail.com



D. charan Kumar Reddy,  
15HR1A0563,  
Dept. of CSE, MTIET, Palamaner.  
dcharankumarr@gmail.com



P.Vennela,  
15HR1A0579,  
Dept. of CSE, MTIET, Palamaner.  
Vennela610@gmail.com