

# **Sports Analytics (Cricket) Using The 3 Factor Algorithm Based on Team Performance**



**Institute of Engineering & Management  
Kolkata, India**

**MAY,2018**

A PROJECT REPORT ON

**Sports Analytics (Cricket)**

Presented to the faculty of Engineering,



Institute of Engineering & Management, Kolkata.

In partial fulfilment of requirements for the degree of

B.Tech (I.T. Engineering)

Submitted By,

**Dhairyya Agarwal**

ROLL NO: 10400214112

**&**

**Rishav Kumar**

ROLL NO: 10400214086

Under the Guidance of

**Prof. Mr. Satyasan Changdar**

**Department of Information Technology, IEM**

# **PREFACE**

The thesis has been submitted towards the partial fulfilment of the requirement for the **AWARD** of the **DEGREE of BACHELOR OF TECHNOLOGY IN INFORMATION TECHNOLOGY** of **INSTITUTE OF ENGINEERING & MANAGEMENT, MAULANA ABULA KALAM AZAD UNIVERSITY OF TECHNOLOGY, KOLKATA.**

**Dated: May, 2018.**

**Place: Salt Lake, Kolkata**

.....

**(Dhairyya Agarwal)**

.....

**(Rishav Kumar)**

# **Declaration of Originality and Compliance of Academic Ethics**

We hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of **Bachelor of Technology in Information Technology**.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

**Name : Dhairyya Agarwal & Rishav Kumar**

**Exam Roll Number : 10400214112 & 10400214086**

**Thesis Title : Sports Analytics (Cricket)**

**Date : May, 2018.**

.....

**(Dhairyya Agarwal)**

.....

**(Rishav Kumar)**

# **CERTIFICATE OF RECOMMENDATION**

I hereby recommend that the thesis entitled “**Sports Analytics(Cricket)**” prepared under my supervision and guidance by **Dhairyya Agarwal** and **Rishav Kumar** be accepted in partial fulfilment of the requirement for awarding the degree of **B.Tech in Information Technology** from **Institute of Engineering and Management, Kolkata, India**. The project, in our opinion, is worthy for its acceptance.

---

**ASST. PROF. SATYASARAN CHANGDAR**

(THESIS SUPERVISOR)

I.T Dept. of IEM, Kolkata, India.

**COUNTER SIGNED BY**

---

**PROF. DR. MOHUA CHAKRABORTY**

(HEAD OF THE DEPARTMENT)

I.T Dept. of IEM, Kolkata, India.

**COUNTER SIGNED BY**

---

**PROF. DR. AMLAN KUSUM NAYAK**

(PRINCIPAL)

I.T Dept. of IEM, Kolkata, India.

## **ACKNOWLEDGEMENT**

Every project big or small is successful largely due to the effort of a number of wonderful people who have always given their valuable advice or lent a helping hand. I sincerely appreciate the inspiration; support and guidance of all those people who have been instrumental in making this project a success.

We, Dhairyya Agarwal & Rishav Kumar, the student of **Institute of Engineering and Management (I.T.)**, are extremely grateful to mentors for the confidence bestowed in us and entrusting our project entitled “**Sports Analytics (Cricket)**” with special reference.

At this juncture we feel deeply honored in expressing our sincere thanks to **Satyasaran Changdar** for making the resources available at right time and providing valuable insights leading to the successful completion of our project who even assisted us in compiling the project.

We would also like to thank all the faculty members of **IEM, Kolkata** for their critical advice and guidance without which this project would not have been possible.

Last but not the least we place a deep sense of gratitude to our family members and friends who have been constant source of inspiration during the preparation of this project work.

**DATE –** May, 2018.

**NAME –** Dhairyya Agarwal & Rishav Kumar

**PLACE –** Kolkata, India.

## **ABSTRACT**

In this assignment we have developed a simple algorithm inspired by Coupled Matrix Factorization using which we find the cricketing value of the player (not the monetary value of the player) which could be used by analysts and managers to build and manager teams. With the advent of IPL, many major and minor T20 league have sprang across the club. All clubs are trying to estimate the value of players and build a team which can win them the competition using the resources they have. As team building is a fundamental task in club based sports and huge amount of money is involved in this field, many algorithm and predictive models have been tried in this field. We have proposed a simple algorithm which takes into the performance of a player in only T20 matches with respect to the team performance as well. The algorithm also considers the fielding skills of the player which have never been considered before. This simple equation, taking into account a holistic view of the T20 landscape, is able to answer many questions about why a particular team won or lost, which is the best player for a scenario, replacement for a particular player and the composition of the team, etc.

# CONTENTS

<i>Preface</i> .....	[iii]
<i>Declaration of Originality and Compliance of Academic Ethics</i> .....	[iv]
<i>Certificate of Recommendation</i> .....	[v]
<i>Acknowledgement</i> .....	[vi]
<i>Abstract</i> .....	[vii]
<i>Contents</i> .....	[viii]
<b>1.INTRODUCTION</b> .....	<b>1</b>
1.1 The Importance of T20 format.....	1
1.2 Importance of IPL.....	2
1.3 Auction.....	2
1.4 Importance of analytics into auction.....	2
<b>2.FUNDAMENTAL STEPS IN ANALYTICS</b> .....	<b>3</b>
2.1 Identifying the Problem.....	3
2.2 Designing Data Requirement.....	4
2.3 Pre-processing the data.....	4
2.4 Performing Analytics Over Data.....	5
2.5 Visualizing Data.....	5
<b>3.BACKGROUND &amp; MOTIVATION</b> .....	<b>5</b>
<b>4.NOVELTY</b> .....	<b>6</b>
<b>5.ALGORITHM</b> .....	<b>6</b>



6.DATA STORAGE.....	9
7.RESULTS.....	9
7.1 Fielding Value 2017.....	9
7.2 Fielding Value 2016.....	10
7.3 Fielding Value 2018.....	10
7.4 Fielding Value 2016 & 2017.....	11
7.5 Bowling Value 2017.....	11
7.6 Bowling Value 2016.....	12
7.7 Bowling Value 2018.....	12
7.8 Bowling Value 2016 & 2017.....	13
7.9 Batting Value 2018.....	13
7.10 Batting Value 2017.....	14
7.11 Batting Value 2016.....	14
7.12 Batting Value 2016 & 2017.....	15
7.13 Why Mumbai Indians won in 2017.....	15
7.14 Why Hyderabad won in 2016.....	16
7.15 Replacement For Rabada.....	16
7.16 Replacement For Russell.....	17
7.17 Replacement for Warner.....	17
7.18 Replacement For Smith.....	18
8. ALGORITHM VALIDATION.....	18
9. CONCLUSION.....	18

10. FUTURE SCOPE.....	19
11. REFERENCES.....	21

# **CHAPTER 1**

## **Introduction**

Before we go into the details of the algorithm devised for the analytics of the game, we need to understand the game. Cricket is a bat and a ball game which is played by two teams comprising eleven players each. One team bats, and the other team fields, the team with the maximum runs at the end of the game wins. The game has evolved and developed many formats but the most interesting of them is the T20 over format.

The format is generally played as a league based sports event. Before the beginning of the league teams are formed via drafts/auctions which is based after analysing how valuable a player is. For this purpose, large amount of data is analysed based on predictive and statistical models as a large amount of money is at stake.

So, we have devised an algorithm that also ascertains how valuable a player is in cricketing terms based on ones batting, bowling and fielding facets involving team performance and importance of the match to get more insights into this field.

### **The Importance of T20 format**

With the objective of further increasing the audience base, the format was invented in England. The game became shorter as each team had to face only 120 balls, with the objective being same to score the maximum runs. This made the game more entertaining by tweaking the nature of the game completely. The game became more appealing to the people all around the world.

Based on the popularity and financial interests, more and more leagues were started around the world. But the first one to make this a major financial and sporting success was the Indian Premier League (IPL).

## **Importance of IPL**

Started in 2007, based along the lines of English Premier League, IPL is one of the most commercialised sports league in the world joining the likes of NBL and UEFA premier league. The league comprises of 8 T20 teams with each team playing 14 matches each with a game away and at home format against each opponent. The IPL is one of the foremost T20 laboratory where new talent as well as old warhorses showcase their talent. The league brought into the concept of auction to form the teams before the league gets underway thus promoting analytics into sports, especially Cricket. With a huge prize pool at stake along with heavy spending purses, analytics became way more important when cricket teams were picked.

This trend continued all over the globe.

## **Auction**

Here each player is assigned a base price based on ones' skill sets and the past performance of the player. The player are then grouped into sets based on the skills player has and every set has a few number of players. The auction is designed as an open cry English auction and is conducted by a well-known, independent, professional auctioneer with each player being individually put up for bidding by the franchises.

Each player bidding starts at the base or reserve price and goes on till there is the highest bid and no one wants to bid further. The franchise with the highest bid is accepted as the franchise for which the player will play and the bid as the winning bid, the fee that the franchise will pay to the player for the season. The fee assumes the player is available for the entire season and is adjusted on a pro-rata basis depending on the player's availability during the season.

The players are auctioned in a linear and sequential manner going from one set to another and auctioning players in a set based on their names drawn from a bag by the auctioneer.

## **Importance of analytics into Auction**

The trend started in the year 2014 when KKR employed SAP analytics to help them put up a team for 2014 and KKR won on to win the championship. This move prompted other teams to tie up with other analytics firms for the future

auctions. Each analytic and firm moving onto use various softwares such as SAP, Hadoop, R along with various predictive models and AI algorithms to evaluate player on the basis of ton of data available.

Some franchises using the tools mentioned above devised their own metrics to evaluate a player and were quite successful in doing so. This year again in 2018 when the major IPL auction took place, franchises were using analytics to choose players for their teams.

Analytics is not only used in auctions, but also to sign up players for injured players while formulating match strategies against players and also when the actual game is under play.

These analytics move is now no longer unique to IPL but has caught trend all over the globe wherever T20 leagues are under way and this trend is slowly catching on with International teams as well.

## **CHAPTER 2**

### **Fundamental Steps in Analytics**

---



## **IDENTIFYING THE PROBLEM**

The first step involved is identifying the problem one wants to solve using analytics. Today, analytics trends change by performing analytics over web datasets and other sources. With the help of web analytics; one can solve huge range of problems in various verticals and horizontals, i.e., domains.

The problem we want to solve is to identify the cricketing value of each player under question so perform better at the auctions, player replacement scenarios, why a team won the league previously, a player for a match situation.

## **DESIGNING DATA REQUIREMENT**

To perform the data analytics for a specific problem, it needs datasets from related domains. Based on the domain and problem specification, the data source can be decided and based on the problem definition; the data attributes of these datasets can be defined.

For our problem we need data regarding the matches held in the IPL both in the past and in the present. We have decided to scrap the data of 2016, 2017 from [www.iplt20.com](http://www.iplt20.com) and the current edition, 2018, data from the cricbuzz API. We have decided to get data of 2016 and 2017 as many new players have come into the game and this is quite recent to consider as various things have changed compared to the other editions.

The scraping is done using python and selenium web driver for the past editions and for the current edition we use the PyCricbuzz library to make the API calls.

## **PREPROCESSING THE DATA**

The data we obtained from scraping might not be useful in the raw format. In layman terms, pre-processing is used to perform data operation such as data cleansing, data aggregation, data augmentation, data sorting, and data formatting, to translate data into a fixed data format before providing data to algorithms or tools. The data analytics process to derive knowledge from data will be initiated with this formatted data as the input.

The data is scrapped in such a way that no pre-processing is required and the data can be fed into the algorithm giving us the knowledge required straight away. The API call also provides us the data in JSON format from which we extract the data using slicing and indexing operations so that the algorithm can process the data.

## **PERFORMING ANALYTICS OVER DATA**

After data is available in the required format for data analytics algorithms, the data analytics operations are performed for discovering meaningful information from data to take better decisions towards business and problem in question. Analytics can be performed with various machine learning as well as custom algorithmic concepts, such as regression, classification, clustering, and model-based recommendation

In our case, the scraped data and data from the API call are fed into the algorithm giving values of a particular skill such as batting, bowling and fielding for a particular game and then stored into a database. The algorithm mentioned above is explained in detail in the subsequent pages.

The collective information for a player can be calculated via obtaining data from the individual database tables and merging it to get the knowledge.

## **VISUALIZING DATA**

Data visualization is used for displaying the output of data analytics. Visualization is an interactive way to represent the data insights. This can be done with various data visualization softwares, R packages and Python packages.

We use data visualization to show the top 10 batters, bowlers and fielders. Data visualization is an excellent way to show complicated and lot of data at the same time in an easy concise manner.

# **CHAPTER 3**

## **Background and Motivation**

---

After reading the paper on Coupled matrix Factorization which highlighted how a combination of matrices was giving better insight into recommendation system during the time of IPL auctions 2018 an idea developed into my mind.

Taking a leaf out of coupled matrix factorization, I wanted to employ player performance with respect to team performance and analyse how valuable a player is in cricketing terms. I also wanted to consider other facets of the sport such as the fielding and importance of the game as a whole on the league.

Based on that motivation, our team performed background research providing insight that such a thing have never been done before in the scientific community and we decided to develop a simple algorithm based on the idea and put it into practise. The developed algorithm was able to give satisfactory results for the purpose it was developed for.

## **Novelty**

---

The algorithm which will be discussed in subsequent pages has few novelties. They are as follows: -

1. The algorithm takes into account the team performance as a whole
2. The algorithm compares the importance of individual performance with team performance
3. The algorithm takes into account fielding aspects of the game
4. The algorithm takes into account the importance of the game



# CHAPTER 4

## Algorithm

---

The algorithm termed by us as 'The three factor algorithm based on team performance' is based on three facets of the game.

The three facets are:-

1. The batting skills of the player
2. The bowling skills of the player
3. The fielding skills of the player

Summing up the three skills of the player we get the total skill set or the value of the player in cricketing terms. Each skill is calculated individually and independently of the other and added at the end and then multiplied by the importance of the match. In this way we develop a standardized metric to compare each player against one another. Each skill is calculated by a simple equation.

First, the batting skill/value of the player

**Bat Factor1** = Batsman Runs - [out status \* (Team Score/Total Batsman Played)]

**Bat Factor2** = Batsman Boundaries Count/ Team Boundaries Count

**Bat Factor3** = Batsman Runs /(Balls Faced \*10)

**Batting Value**= **Bat Factor1** + **Bat Factor2** + **Bat Factor3**

where,

Bat Factor1 tells us whether the number of runs the batsman has scored more than the team's average runs or not. If batsman has scored less than the team's average, his factor goes below negative because he has performed below the team's average for the game.

Bat Factor2 is the relative boundaries count of the player compared to the team's boundaries count. As boundaries play a critical role in T20 games.

Bat Factor3 is the strike rate of the player divided by 1000.

The summation of all these above 3 factors give us the batting value of a player for a T20 game.

Second, the bowling skill/value of the player.

**Bowl Factor1** = wickets Taken by the bowler \* (Runs given by the team/ Wickets taken by the team) – runs given by the bowler

**Bowl Factor2** = Runs given by the bowler / wickets taken by the bowler

**Bowl Factor3** = Runs given by bowler / overs thrown by the bowler

**Bowling Value** = **Bowl Factor1 / (Bowl Factor2 \* Bowl Factor3)**

where,

Bowl Factor1 is the runs given by the player for the wickets which he took compared to the team's average. If the team gave 150 runs and took 10 wickets, out of which the player took 2 wickets. So the team overall gave 15 runs per wicket, so for 2 wickets team gave 30 runs. Now, the player who took 2 wickets gave only 25 runs he performed better than the team average by 5 but if he gave 40 runs than he performed worse by 10 runs compared to team's average.

Bowl Factor2 is the bowling average of the player for the match

Bowl Factor3 is the bowling economy of the player for the match

The multiplication of the two factors along with the division as showed above give us the bowling value of a player for a T20 game.

Finally, the fielding skill/value of the player.

**Fielding Value** = Run outs caused by the player + Catches taken by the player + Stumping performed by the player

This equation is the summation of the run outs, catches and stumping performed by the player for a T20 match giving his fielding value for that game.

To calculate the overall value of a player for a match we use,

**Value = (Batting Value + Fielding Value + Bowling Value) \* Importance of the match**

where,

importance of the match depends on the game which the player is playing. A knockout league game has a different importance compared to a normal league game. The same logic applies for semi-finals, finals and eliminator as this games are of varying importance level. So a player performing in this games has to be given more importance compared in other games of less importance as this games have much more at stake.

Finally, Total normalized Value of a player for a season is given by

$$\text{Total Value} = \frac{\sum_i^n (\text{Value}_i * \text{Importance}_i)}{\sum_i^n (\text{Importance}_i)}$$

This gives us the normalized value of a player in cricketing terms for any number of games or season or a particular season in question. Thus forming a metric to compare players with one another.

## **DATA STORAGE**

The values calculated via the algorithm is stored in database tables so that these values can be used for various processing and insight to answer the problems such as a player replacement, etc.

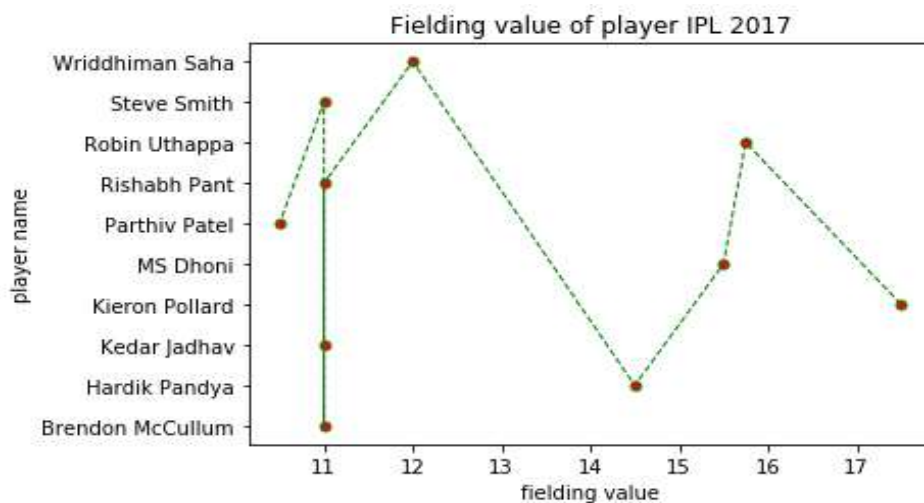
The data is stored in three tables. One for fielding value, one for batting value and another for the bowling value. The data can be fetched from the tables easily using the select statement and easily used to answer the solutions to the problems.

# CHAPTER 5

## Results

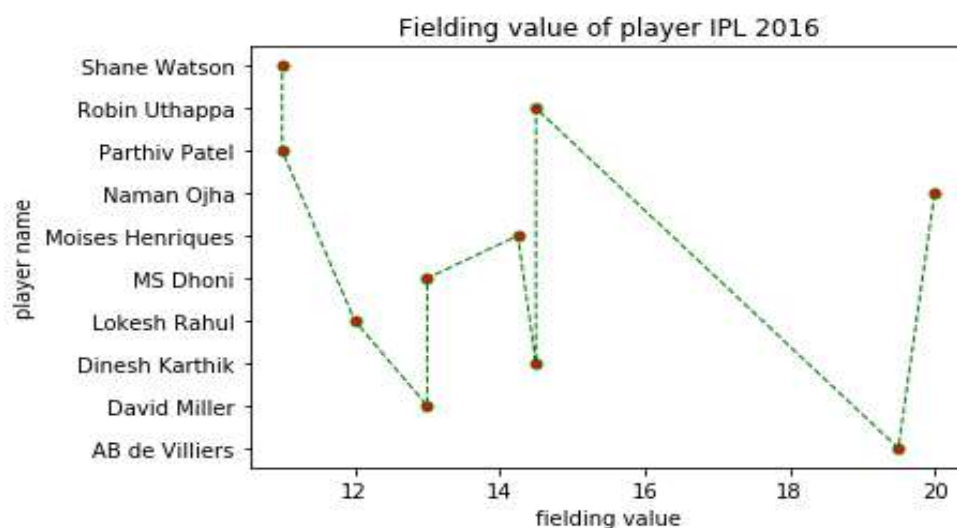
### FIELDING VALUE 2017

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2017. It provides insight by stating the top 10 fielders of the year 2017 and also states how they compare graphically against each other. Kieron Pollard was the top fielder for the year 2017.



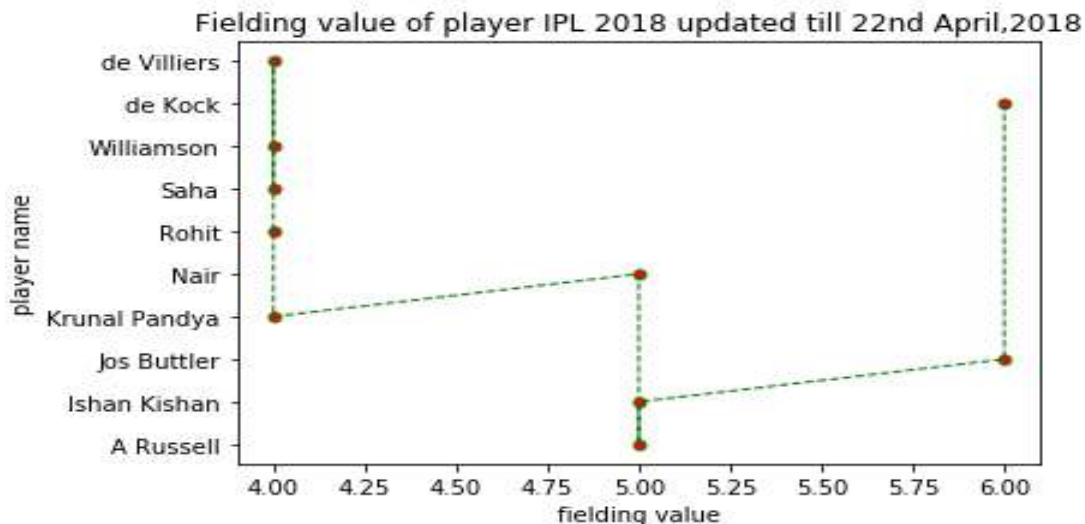
### FIELDING VALUE 2016

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2016. It provides insight by stating the top 10 fielders of the year 2016 and also states how they compare graphically against each other. Naman Ojha was the top fielder for the year 2016.



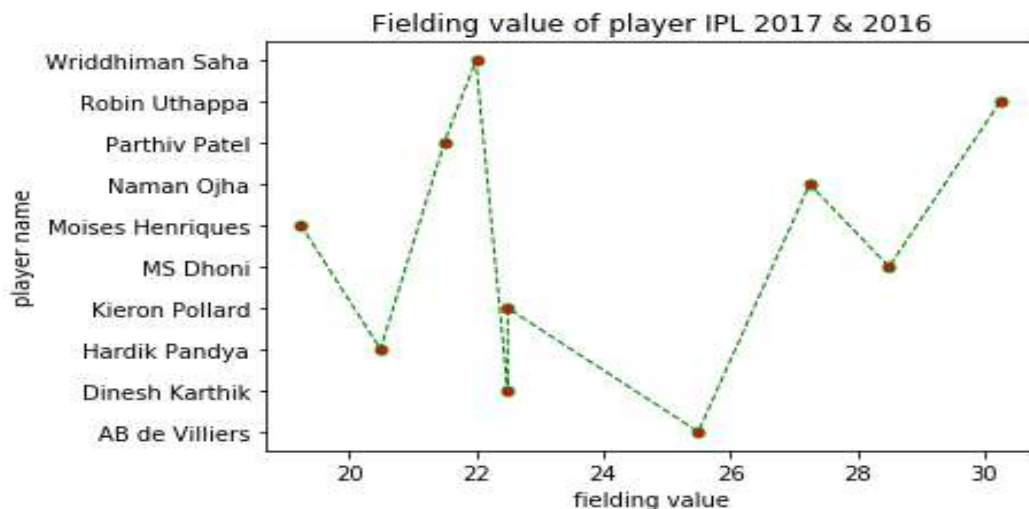
## FIELDING VALUE 2018

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2018. It provides insight by stating the top 10 fielders of the year 2018 up to a certain date and also states how they compare graphically against each other. De kock was the top fielder for the year 2018 up to 22<sup>nd</sup> April, 2018.



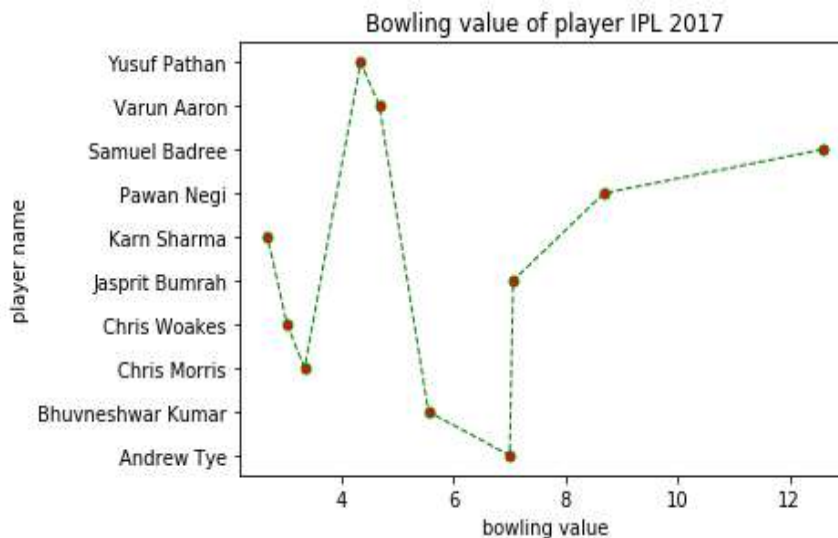
## FIELDING VALUE 2016 & 2017

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the years 2017 & 2016. It provides insight by stating the top 10 fielders of the years 2017 & 2016 and also states how they compare graphically against each other. Robin Uthappa was the top fielder for the year 2017 & 2016 combined.



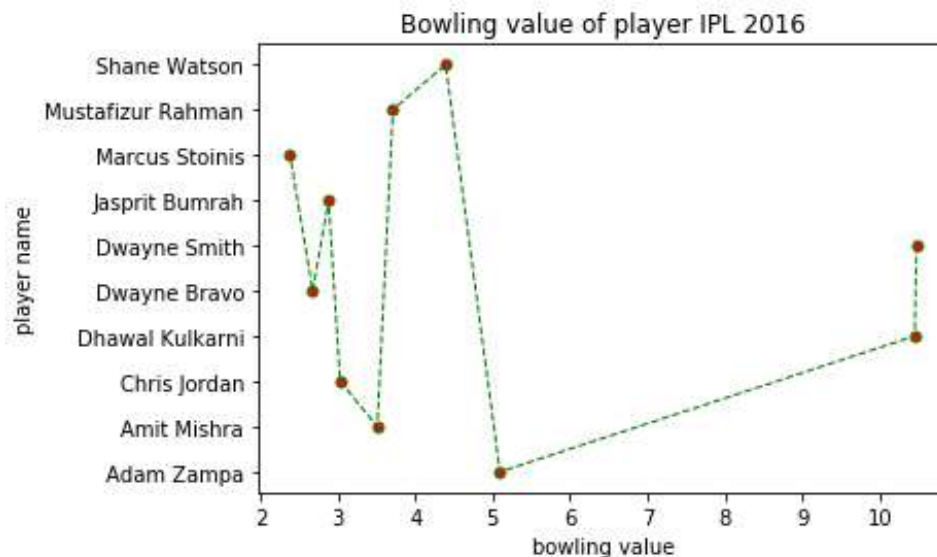
## **BOWLING VALUE 2017**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2017. It provides insight by stating the top 10 bowlers of the year 2017 and also states how they compare graphically against each other. Samuel Badree was the top bowler for the year 2017.



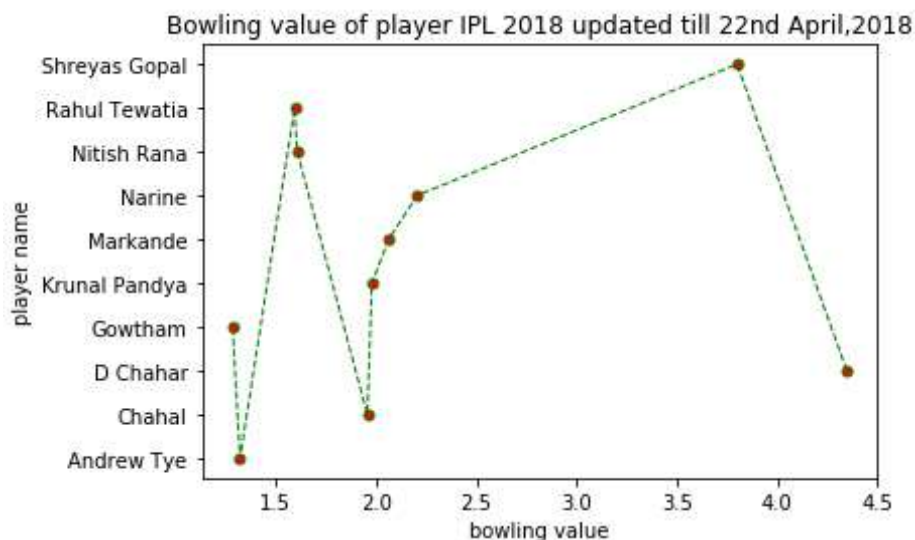
## **BOWLING VALUE 2016**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2016. It provides insight by stating the top 10 bowlers of the year 2016 and also states how they compare graphically against each other. Dwayne Smith was the top bowler for the year 2016.



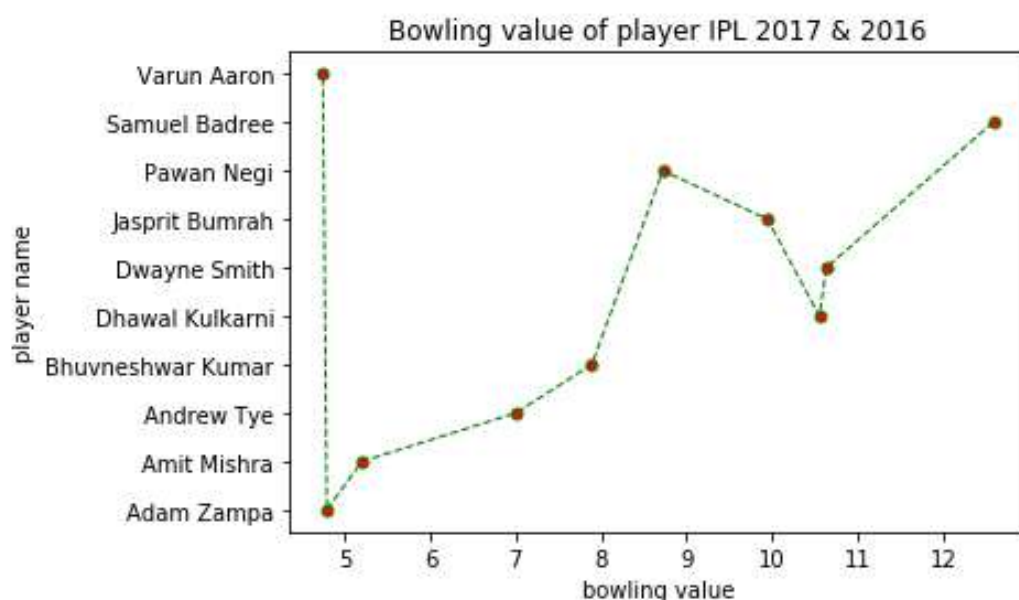
## **BOWLING VALUE 2018**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2018. It provides insight by stating the top 10 bowlers of the year 2018 up to a certain date and also states how they compare graphically against each other. Deepak Chahar was the top bowler for the year 2018 up to 22<sup>nd</sup> April, 2018.



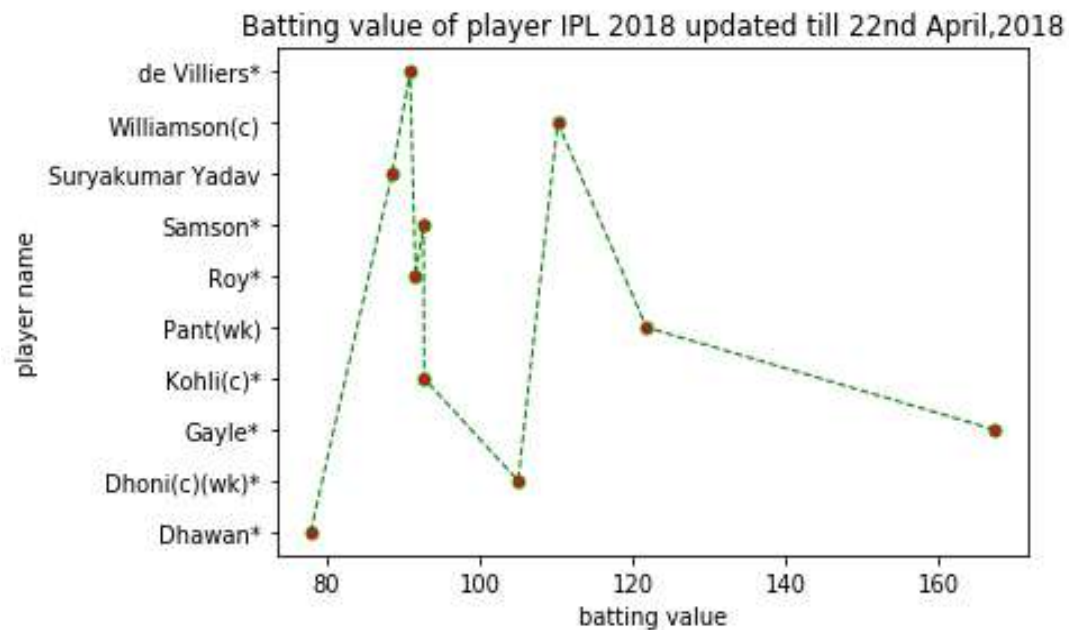
## **BOWLING VALUE 2016 & 2017**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the years 2017 & 2016. It provides insight by stating the top 10 bowlers of the years 2017 & 2016 and also states how they compare graphically against each other. Samuel Badree was the top bowler for the year 2017 & 2016 combined.



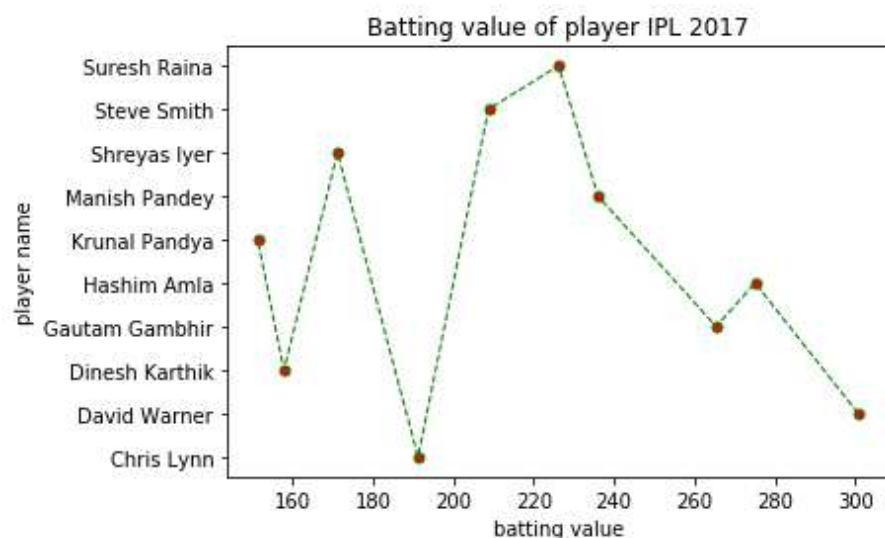
## **BATTING VALUE 2018**

The program outcome shown below is generated from the knowledge derived from the API data via the algorithm for the years 2018. It provides insight by stating the top 10 batsman of the year 2018 up to a certain date and also states how they compare graphically against each other. Chris Gayle was the top batsman for the year 2018 up to 22<sup>nd</sup> April, 2018.



## **BATTING VALUE 2017**

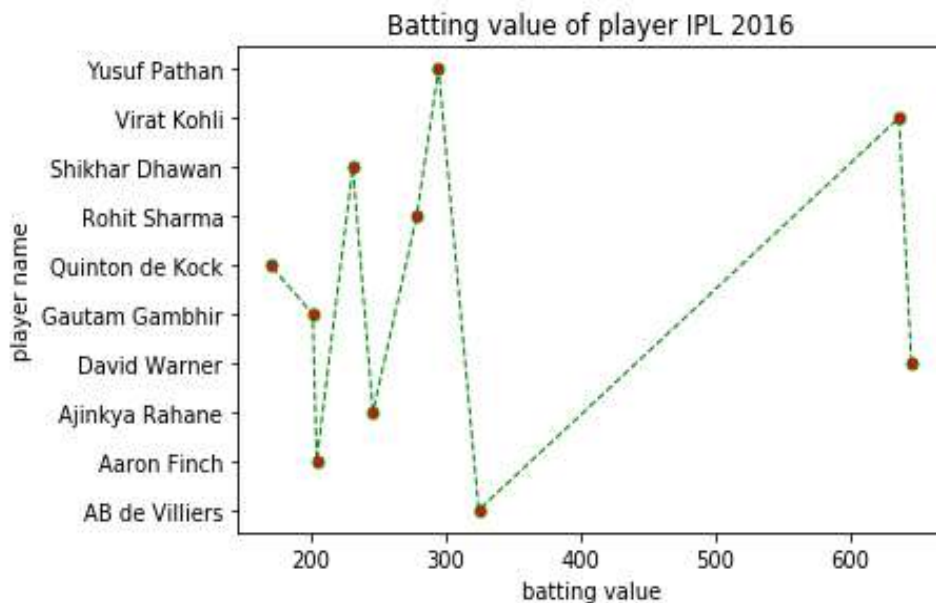
The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2017. It provides insight by stating the top 10 batsman of the year 2017 and also states how they compare graphically against each other. David Warner was the top batsman for the year 2017.





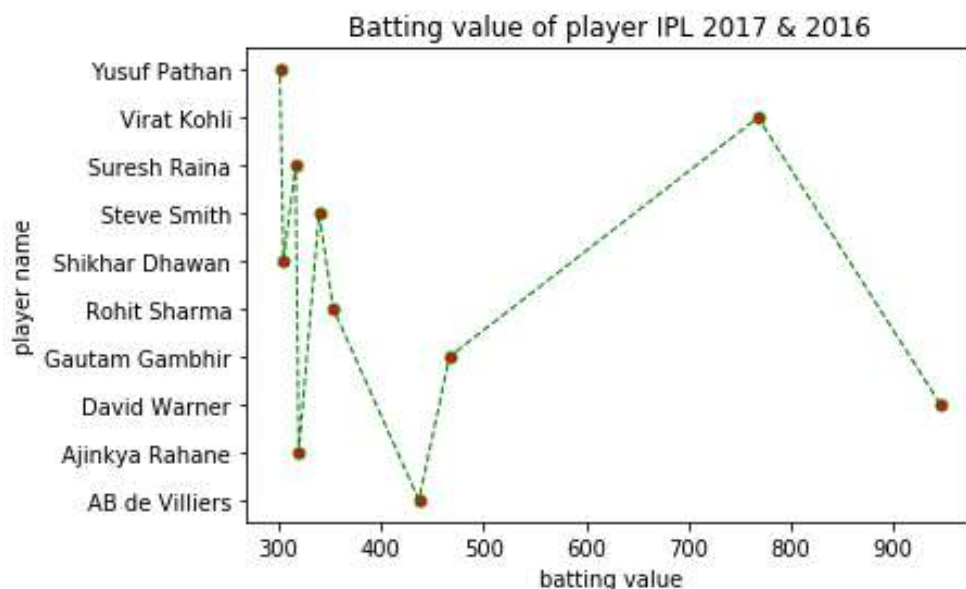
## **BATTING VALUE 2016**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the year 2016. It provides insight by stating the top 10 batsman of the year 2016 and also states how they compare graphically against each other. David Warner was the top batsman for the year 2016.



## **BATTING VALUE 2016 & 2017**

The program outcome shown below is generated from the knowledge derived from the scrapped data via the algorithm for the years 2016 and 2017. It provides insight by stating the top 10 batsman of both the years combined and also states how they compare graphically against each other. David Warner was the top batsman for both the years.



## Why Mumbai Indians Won in 2017?

The outcome of the program based on the knowledge of the 2017 scrapped data fed into the algorithm shows why Mumbai Indians won against Supergiants in the final that year validating the logic of the algorithm.

```
Mumbai Indian Batting Score is 49.9379328807
Pune Supergiants Batting Score is 10.8904650082
Mumbai Indians is superior in batting department

Mumbai Indian Bowling Score is 0.945627844843
Pune Supergiants Bowling Score is 0.521431643135
Mumbai Indians is superior in bowling department

Mumbai Indian Fielding Score is 11.0
Pune Supergiants Fielding Score is 9.0
Mumbai Indians is superior in fielding department

Mumbai Indian team score is 61.8835607255
Pune Supergiants team score is 20.4118966514
Diff is 41.4716640741
Mumbai Indians is superior

So Mumbai Indians Won IPL 2017
```

## Why Hyderabad won in 2016?

The outcome of the program based on the knowledge of the 2016 scrapped data fed into the algorithm shows why Sunrisers Hyderabad won against Royals in the final that year validating the logic of the algorithm.

```
Sunrisers Hyderabad Batting Score is 49.0316347485
Royal Challengers Bangalore Batting Score is 43.808031988
Sunrisers Hyderabad is superior in batting department

Sunrisers Hyderabad Bowling Score is 0.517056669895
Royal Challengers Bangalore Bowling Score is 1.16232429313
Royal Challengers Bangalore is superior in bowling department

Sunrisers Hyderabad Fielding Score is 11.0
Royal Challengers Bangalore Fielding Score is 11.0
Both teams are equal in this aspect

Sunrisers Hyderabad team score is 60.5486914184
Royal Challengers Bangalore team score is 55.9703562811
Diff is 4.57833513735
Sunrisers Hyderabad is superior

So Sunrisers Hyderabad Won IPL 2016
```

## **REPLACEMENT FOR RABADA**

The program outcome shown below shows the players who are capable of replacing Kagiso Rabada as he is ruled out of IPL 2018. For the year 2017, Rabada's performance was analysed and player whose bowling skills were nearer to Rabada were suggested as his replacement as shown below in the list.

Kagiso Rabada batting value is 9.803944

Kagiso Rabada bowling value is 0.143441

Kagiso Rabada fielding value is 1.0

Kagiso Rabada total value is 10.947385

Players that can replace Kagiso Rabada are:

Shane Watson

Washington Sundar

Mitchell McClenaghan

Nathan Coulter-Nile

Mitchell Marsh

Pradeep Sahu

## **REPLACEMENT FOR RUSSELL**

The program outcome shown below shows the players who are capable of replacing Andre Russell as he is ruled out of IPL 2017. For the year 2016, Russell's performance was analysed and player whose bowling skills were nearer to Russell were suggested as his replacement as shown below in the list.

Andre Russell batting value is 0.467251375

Andre Russell bowling value is 0.1677135

Andre Russell fielding value is 1.0

Andre Russell total value is 1.634964875

Players that can replace Andre Russell are:

Karn Sharma

Chris Morris

Mustafizur Rahman

Dwayne Bravo

Gurkeerat Mann Singh

## REPLACEMENT FOR WARNER

The program outcome shown below shows the players who are capable of replacing David Warner as he is ruled out of IPL 2018. For the two years 2016 and 2017, David's performance was analysed and player whose batting skills were nearer to david were suggested as his replacement as shown below in the list.

David Warner batting value is 30.7618676504

David Warner bowling value is 0

David Warner fielding value is 1.0

David Warner total value is 31.7618676504

Players that can replace David Warner are:

Chris Lynn

Hashim Amla

AB de Villiers

Vijay Shankar

Virat Kohli

## REPLACEMENT FOR SMITH

The program outcome shown below shows the players who are capable of replacing Steven Smith as he is ruled out of IPL 2018. For the two years 2016 and 2017, Steve's performance was analysed and player whose batting skills were nearer to smith were suggested as his replacement as shown below in the list.

Steve Smith batting value is 15.0798402667

Steve Smith bowling value is 0

Steve Smith fielding value is 1.0

Steve Smith total value is 16.0798402667

Players that can replace Steve Smith are:

Ben Cutting

Krunal Pandya

Gautam Gambhir

Iqbal Abdulla

Chris Morris

Manish Pandey

Yusuf Pathan

AB de Villiers

Vijay Shankar

Ankit Bawne

Ishank Jaggi

Quinton de Kock

Murali Vijay

JP Duminy

# **CHAPTER 6**

## **ALGORITHM VALIDATION**

The algorithm is validated by the results shown above especially the results showing why Mumbai Indians and Sunrisers Hyderabad won the IPL respectively in 2017 and 2016 as indicated by their collective team's batting, bowling and fielding value for that particular year. The algorithm also gives insight why several players for sold such a high amount this year as their performance in the last two years 2016 and 2017 have been among the very best in the league. So this proves the validity of the algorithm.

# **CHAPTER 7**

## **CONCLUSION**

In this paper, we present a comprehensive work of sports analytics conducted on IPL this year and past two years based on the simple mathematical equation which we have devised. Different applications of the algorithm have been shown above. The key issues of this work can be summarized as follows:

1. Scraping data and storing in a usable format is very essential for any data analytics application as without it no analytics can be performed. The scraping process and the storage of the data is very vital for the success of this experiment. Extensive care was taken during this step as we used selenium for scraping and SQL for storing the data.
2. One challenge for data analytics is when data is there in useful form but some of the data have anomalies or exception to the ideal form, as the equations of the algorithm are prone to data where a division by zero can occur. We so have to clean the data and use for this exceptional data before applying the mathematical equations of the algorithm to get the desired results. In this manner the algorithm performs in the correct manner without throwing any unexpected errors.
3. We then store the values computed from the algorithm into the database. The database has some processed knowledge which can be used for

generating insights based on the application as we have shown in some examples above. However, this method has limitation as programs have to be tweaked from application to application and a single uniform algorithm won't work in this regard.

## **FUTURE SCOPE**

This project has lot of future scope and some of them are indicated as follows:-

1. Scraping and storing data of all the T20 matches held till now irrespective of their level and the country in question which will make it a universal database with information about the oldest and newest player and also providing a greater amount of insight. This can easily help new talent coming up and a suitable replacement of a player can be a player who has not been heard in this part of the world.
2. Taking the commentary of live matches into a program as an input feed and analysing it using the algorithm to better estimate the fielding capabilities as it can track the diving efforts, catches dropped, stumping left, miscommunication between batsman, match situations and much more making the algorithm much more complex and dynamic in nature and offering more insight into the game.
3. The leadership skills of the player can be considered as well to have a greater peek into the world of cricket and a player cricketing skills as some players are in demand because of their leadership skills. So modifying the algorithm to incorporate this feature can help us have greater analytical insight about the player.
4. Building machine learning models such as neural networks using this equation to have greater insight into the game and predict results and outcomes of a game with ease. Also the neural network can be trained for match situations and in order to buy players at the auction. The applications for the machine learning model can be far and wide.

## **REFERENCES:**

- [1] Prince Kansal, Pankaj Kumar, Himanshu Arya, Aditya Methaila, Player valuation in Indian premier league auction using data mining technique, in 2014 International Conference on Contemporary Computing and Informatics (IC3I).
- [2] Pranavan Somaskandhan, Gihan Wijesinghe, Leshan Bashitha Wijegunawardana, Asitha Bandaranayake, Sampath Deegalla, Identifying the optimal set of attributes that impose high impact on the end results of a cricket match using machine learning: Techniques and Applications, in 2017 IEEE International Conference on Industrial and Information Systems (ICIIS).
- [3] Rajdeep Paul, Big data analysis of Indian premier league using Hadoop and MapReduce, in: 2017 International Conference on Computational Intelligence in Data Science (ICCIDS).
- [4] Jayant Prakash, Mayank Khandelwal, Tribikram Pradhan, Evaluation of IPL teams and players using association, correlation and classification rules, in: 2015 International Conference on Computer, Communication and Control (IC4).
- [5] Chuyuan Wei, Fangfang Li, Qiang Zhan, Xiongzhong Fan, Coupled Matrix Factorization for Question Similarity, in: Chinese Journal of Electronics, 2016, Volume 25, pp. 665–671.
- [6] Ramanna, Vishwanath, Indian Premier League (March 6, 2009). Available at SSRN: <https://ssrn.com/abstract=1354534> or <http://dx.doi.org/10.2139/ssrn.1354534>
- [7] David Parker, Phil Burns AND Harish Natarajan, Player valuations in the Indian Premier League, in: October 2008 Frontier Economics.
- [8] C. Petersen, D.B. Pyne, M.J. Portus, B. Dawson, Analysis of Twenty/20 Cricket performance during the 2008 Indian Premier League, in: International Journal of Performance Analysis in Sports, 03 April 2017, pp. 63-69.
- [9] Sanjeet Singh, Measuring the Performance of Teams in the Indian Premier League, in American Journal of Operations Research, August 2011, pp. 180-184.
- [10] Hemanta Saikia, Dibyojyoti Bhattacharjee, On Classification of All-rounders of the Indian Premier League (IPL): A Bayesian Approach, in Vikalpa: The Journal for Decision Makers., Oct 2011.
- [11] Kalgotra, Pankush and Sharda, Ramesh and Chakraborty, Goutam, Predictive Modeling in Sports Leagues: An Application in Indian Premier League (April 28, 2013). Available at SSRN: <https://ssrn.com/abstract=2465300> or <http://dx.doi.org/10.2139/ssrn.2465300>.
- [12] Liam J.A. Lenten, Wayne Geerling, László Kónya, A hedonic model of player wage determination from the Indian Premier League auction: in: Sport Management Review, Feb 2012, pp 60-71.

[13] Depken, Craig A. and Rajasekhar, Ramakrishna, Open Market Valuation of Player Performance in Cricket: Evidence from the Indian Premier League (April 1, 2010). Available at SSRN: <https://ssrn.com/abstract=1593196> or <http://dx.doi.org/10.2139/ssrn.1593196>.