A

MINI PROJECT REPORT ON

"Cricket Players' Squad Recommendation System"
Prepared by

Mr. Sabnis Sushant Suhas   (Roll No : 125)

Mr. Shirke Aniket Bhausaheb   (Roll No : 138)

Mr. Somwanshi Adesh Sainath   (Roll No : 140)

Mr. Sonawane Anand Rajendra   (Roll No : 142)

(T.E. Computer)

SAVITRIBAI PHULE PUNE UNIVERSITY ACADEMIC YEAR : 2018-19

DEPARTMENT OF COMPUTER ENGINEERING

SANJIVANI COLLEGE OF ENGINEERING, KOPARGAON

P.O. SHINGNAPUR, PINCODE - 423 603

CERTIFICATE

This is to certify that

Mr. Sabnis Sushant Suhas   (Roll No : 125)

Mr. Shirke Aniket Bhausaheb   (Roll No : 138)

Mr. Somwanshi Adesh Sainath   (Roll No : 140)

Mr. Sonawane Anand Rajendra   (Roll No : 142)

Have successfully completed Mini Project

"Cricket Players' Squad Recommendation System"

Towards the completion of
Skill Development Lab in Computer Engineering
During the academic year 2019 - 2020

Prof. G. MADHUKAR RAO   DR. D. B. KSHIRSAGAR

[Guide]   [H.O.D. Computer Dept.]

Dr. D. N. KYATANAVAR

[Principal]

ACKNOWLEDGEMENT

First and foremost, we would like to give thanks to God Almighty for His mercy and beautiful blessings in our life. Even in the hardest times, He has always been there to guide and comfort us and to make sure that we never stray too far from His narrow way.

We express our sincere gratitude to Prof. D. B. Kshirsagar Head of Department(Computer) SRES's COE for his unending support and encouragement during the years we have studied under his tutelage.

Our sincere thanks go to all the teachers and staff for their help and understanding.

Faculty :-

Prof.M.A.Sale
Prof.G.Madhukar Rao

Mr. Sabnis Sushant Suhas   (Roll No : 125)

Mr. Shirke Aniket Bhausaheb   (Roll No : 138)

Mr. Somwanshi Adesh Sainath   (Roll No : 140)

Mr. Sonawane Anand Rajendra   (Roll No : 142)
(T .E Computer)

Dedication

To all the Cricket Lovers...

Contents

CHAPTER 1

INTRODUCTION

Cricket is the most popular game in Asian countries, a lot of series are played across the set of different countries during a year and World Cup is played once in four years across all the cricket playing nations. Cricket is played in different formats like one day international (ODI), T20 and Test matches.Apart from this many league matches at club level and national level are played within the country. For any such tournaments, series or World Cup a suitable team of playing 11 and 4 extra players(i.e., squad) needs to be selected to form a team.

Cricket team consists of a set of batsmen and bowlers with one wicket keeper who can also bat. The selectors and team captain have to select batsmen and bowlers in the team with a wicket keeper. Each batsman in the team will be specialized to bat at a different position in the playing eleven and there are varieties of bowlers like spin,fast and medium fast in the team. Team members need to be selected from a pool of players depending on the opponent team against which the team is playing. So selecting such a team by coaches and captain is biased and error prone, this decision depends on the experience and analytic skills of the individual by using the performance measures and statistics of the players. So automating the selection process by considering the performance measures helps coaches and captains to select the most efficient team which have higher chances of winning the game.

To automate such process, we propose a squad selection on the basis of statistics of the players.

CHAPTER 2

SCOPE AND OBJECTIVES

2.1 Scope

1. **Cricket:**
   In Sports field Cricket ,Select the players on the basis of performance , To understand the Performance of Players' , how that Player is best for Squad on the basis of Statistics

2. **Reviews:**
   This tool can be used for analyzing the properly Allocation of Players for the Squad

2.2 Objectives

1. The main objective of the project is to build an application program that provides a platform to to reduce the manual work for analysis the cricket data of players
2. To recommend players for the Cricket Squad on the basis of statistics of the players.
3. To show abilities of the players on the basis of statistics
4. To show comparisons between Players with stack plot

CHAPTER 3

<u>LIMITATIONS</u>

There are few limitations that this study has with respective to data and the methodology that can be used.

1. System is restricted to a particular structure of dataset
2. Pitch report and cliamatic conitions aren't considered
3. For bowlers (a specific category of cricket player) we haven't distinguished them as pacers and spinners
4. Wicket Keeper are cnsidered as the batsman who has the wicket keeping abilities
5. Only registered users can used this system

CHAPTER 4

<u>DATA & REQUIREMENT</u>

4.1 Data

Datasets were taken from online community, kaggle.com

4.2 Hardware Requirements

1. 2GB RAM
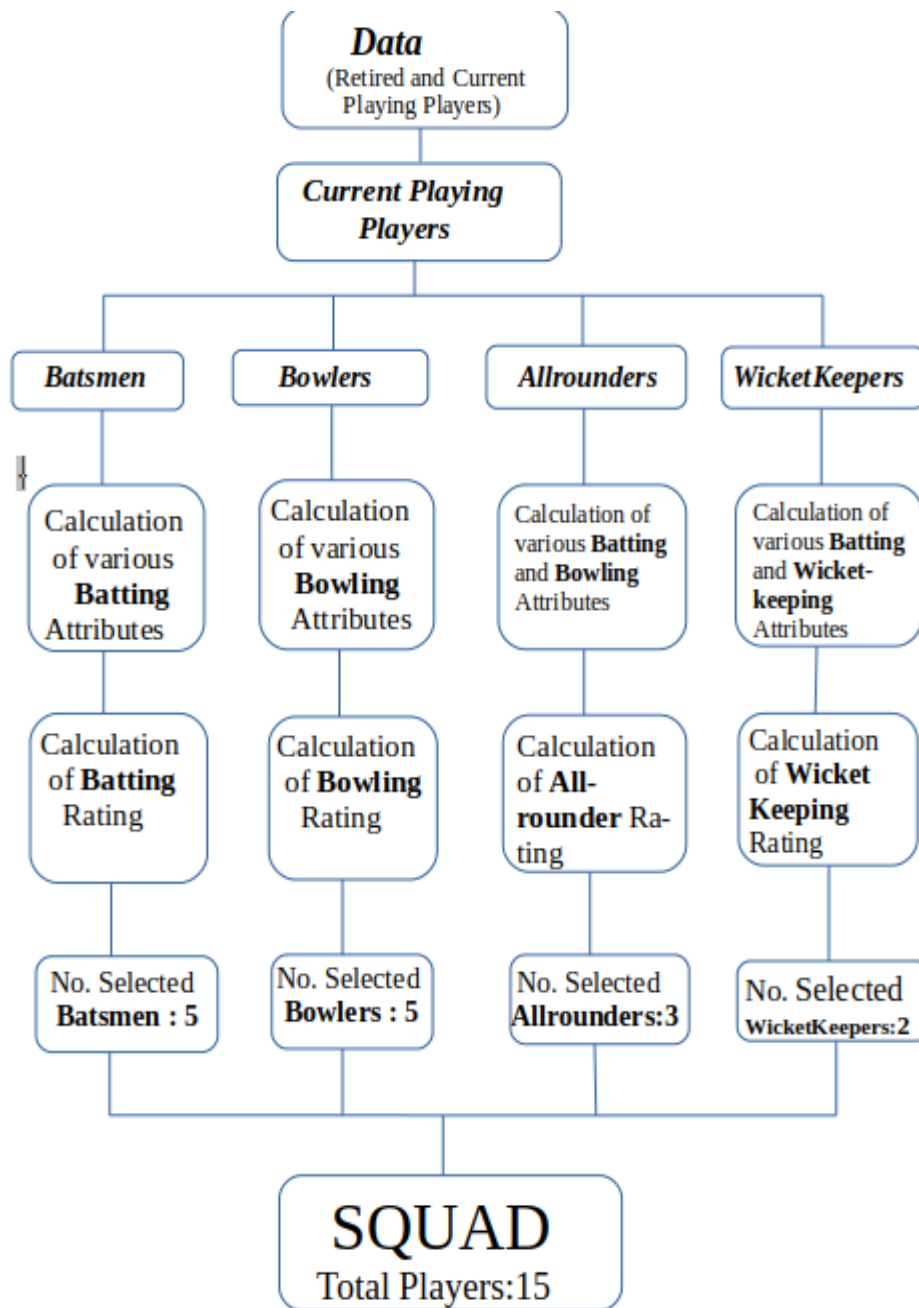2. Dual core Intel or AMD processor

4.3 Software Requirements

1. Python3 Programming Language
2.Windows or Linux operating system
3.*pandas* Python library

4.4 Tools and Techniques

1.We have used Python environment and CSV file for our analysis of cricket players data
2.There are some tools and libraries that are required while doing the analysis of the data and they are as follows:

1. matplotlib.pyplot
2. pandas
3. os
4. tkinter
5. time

3.We have used **Stack Plots** for visualization of the calculated ratings

CHAPTER 5

Breakdown Structure


```
                    Data
             (Retired and Current
               Playing Players)

              Current Playing
                  Players

    Batsmen      Bowlers     Allrounders   WicketKeepers

  Calculation   Calculation  Calculation of  Calculation of
  of various    of various   various Batting various Batting
  Batting       Bowling      and Bowling     and Wicket-
  Attributes    Attributes   Attributes      keeping
                                             Attributes

  Calculation   Calculation  Calculation    Calculation
  of Batting    of Bowling   of All-         of Wicket
  Rating        Rating       rounder Ra-     Keeping
                             ting            Rating

  No. Selected  No. Selected No. Selected   No. Selected
  Batsmen : 5   Bowlers : 5  Allrounders:3  WicketKeepers:2

                    SQUAD
                Total Players:15
```

CHAPTER 6

<u>DATA FETCHING</u>

First of all ,we download dataset from kaggle website.
Website:-
   https://www.kaggle.com/hverified/cricket-players-odi
There are lots of attributes in a dataset , and it is found that data is not clean.
Some samples of original dataset : -

NAME	COUNTRY	Full name	Born
PK Zutshi	India	PK Zutshi	date unknown
RN Zutchi	India	RN Zutchi	date unknown
Zulfiqar Ali Khan	India	Zulfiqar Ali Khan	December 21, 1969, Bikaner, Rajasthan
Zakaria Zuffri	India	Syed Zakaria Zuffri	October 12, 1975, Gauhati (now Guwahati), Assam
Zuber Khan	India	Zuber Khan	April 13, 1974, Indore, Madhya Pradesh

Died	Current age	BATTING_ODIs_Mat	BATTING_ODIs_Inns
Not Available	Not Available	0	
Not Available	Not Available	0	
Not Available	49 years 92 days	0	
Not Available	43 years 162 days	0	

BATTING_ODIs_6s	BATTING_ODIs_Ct	BATTING_ODIs_St	BOWLING_ODIs_Mat	BOWLING_ODIs_Inns	BOWLING_ODIs_Balls	BOWLING_ODIs_Runs	BOWLING_ODIs_Wkts	BOWLING_ODIs_
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	

BOWLING_ODIs_BBI	BOWLING_ODIs_BBM	BOWLING_ODIs_Ave	BOWLING_ODIs_Econ	BOWLING_ODIs_SR	BOWLING_ODIs_4w	BOWLING_ODIs_5w	BOWLING_ODIs_10	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	

CHAPTER 7

<u>DATA PRE-PROCESSING</u>

The data which we have Download from k*aggle.com* may contain some missing values,
some garbage values, extra characters, blank spaces, symbols which are not required for
analysis.
So in order to make the data readable and executable(able to be analysed) we are pre-
processing the data.

7.1 Data Pre-processing

Identify missing values, blank spaces and special symbols in the given data, then
put <NA> (Not Applicable) in the records where there are missing values and remove
special symbols as well as blank spaces.
Code (in python):-

```
import pandas as pd

df=pd.read_csv("cricket_data.csv",low_memory=False,na_values=[0])
new_df1=new_df.fillna({
            'Born':'Not Available',
            'Died':'Not Available'})
```

Current age	BATTING_ODIs_Mat	BATTING_ODIs_Inns	BATTING_ODIs_NO	BATTING_ODIs_Runs	BATTING_ODIs_HS	BATTING_ODIs_Ave	BATTING_ODIs_BF	BATTING_ODIs_SR	BATTING
Not Available	0	0	0	0	0	0	0	0	
Not Available	0	0	0	0	0	0	0	0	
49 years 92 days	0	0	0	0	0	0	0	0	
43 years 162 days	0	0	0	0	0	0	0	0	

Chapter 8

<u>Data Processing</u>

In Data-Processing section of this project, the system calculates various statistical attributes related to the **PlayingRole** of each player.Then, on the basis of these calculated attributes,the system computes the **ratings** for **each of these attributes**.The final rating of each player is computed, which depends completely on the **previously computed ratings**.

```
┌─────────────────┐
│    Raw Data     │
└─────────────────┘
         │
┌─────────────────┐
│   Statistical   │
│   Attributes    │
└─────────────────┘
         │
┌─────────────────┐
│ Ratings  of Vari│
│ -ous  Attribute │
└─────────────────┘
         │
┌─────────────────┐
│  Final Rating   │
│ (based on Role) │
└─────────────────┘
```
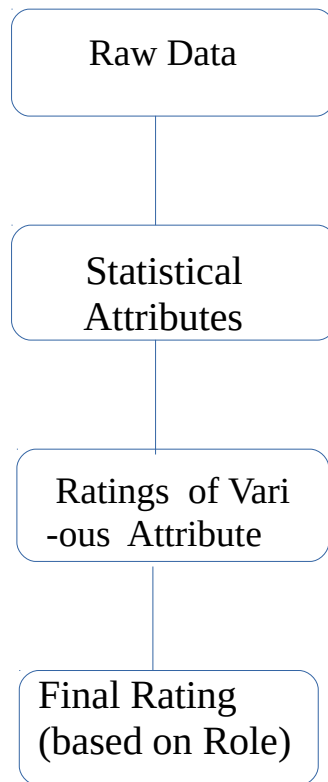
Here are various formulae for computing **Statistical Attributes , Ratings of Statistical Attributes** and **Final Rating of Each Player** (for comparison).

Formulae are the basis of our projetcs and these computations are totally dependent on the mentioned formulae.These formulae are deviced by our team and we have tried hard to acheive maximum accuracy.

Data Processing Formulae

8.1.Statistical Attributes :

(a) For Batting Qualities:

$$BattingAverage = \frac{RunsScored}{InningsPlayed - NotOuts}$$

$$BattingStrikeRate = \frac{RunsScored}{BallsFaced} * 100$$

$$MilestoneReachingAbility(MRA) = \frac{CenturiesScored + HalfCenturiesScored}{InningsPlayed}$$

$$HittingAbility = \frac{FoursScored + SixesScored}{BallsFaced}$$

$$BRPI(BoundaryRunsPerInnings) = \frac{4 * FoursScored + 6 * SixesScored}{InningsPlayed}$$

$$OutRate = \frac{InningsPlayed - NotOuts}{BallsFaced}$$

(b) For Bowling Qualities:

$$BowlingEconomy = \frac{RunsGiven}{BallsBowled} * 6$$

$$BowlingAverage = \frac{RunsGiven}{WicketsTaken}$$

$$BattingStrikeRate = \frac{RunsScored}{BallsFaced} * 100$$

$$WicketsPerInnings(WPI) = \frac{WicketsTaken}{InningsPlayed}$$

(c)For WicketKeeping Qualities:

$$WicketKeepingAbility(WKA) = \frac{CatchesTaken + StumpingsDone}{InningsPlayed}$$

8.2.Ratings of Attributes :

(a)Ratings For Batting Quality Attributes:

$$BattingAverageRating = \frac{BattingAverage}{\frac{\sum BattingAverage}{\sum 1}} * 100$$

$$StrikeRateRating = \frac{StrikeRate}{\frac{\sum StrikRate}{\sum 1}} * 100$$

$$MRARating = \frac{MRA}{\frac{\sum MRA}{\sum 1}} * 100$$

$$OutRateRating = \frac{OutRate}{\frac{\sum OutRate}{\sum 1}} * 100 \qquad \text{\#(Negative Attribute)}$$

$$BRPIRating = \frac{BRPI}{\frac{\sum BRPI}{\sum 1}} * 100$$

(b)Ratings For Bowling Quality Attributes:

$$EconomyRating = \frac{Economy}{\frac{\sum Economy}{\sum 1}} * 100 \quad \text{\#(Negative Attribute)}$$

$$BowlingAverageRating = \frac{BowlingAverage}{\frac{\sum BowlingAverage}{\sum 1}} * 100$$

\#(Negative Attribute)

$$BowlerStrikeRateRating = \frac{BowlerStrikeRate}{\frac{\sum BowlerStrikeRate}{\sum 1}} * 100$$

\#(Negative Attribute)

$$WPIRating = \frac{WPI}{\frac{\sum WPI}{\sum 1}} * 100$$

(c)Ratings For Wicketkeeping Quality Attributes:

$$WKARating = \frac{WKA}{\frac{\sum WKA}{\sum 1}} * 100$$

8.3.Final Rating (playing-role based) :

(a)Batsman Rating:

$$Batsman_{Rating} = \frac{StrikeRateRating_{Batting} + AverageRating_{Batting} + MRA_{Rating} + BRPIRating + HittingAbility_{Rating} - OutRate_{Rating}}{6}$$

(b)Bowling Rating:

$$BowlerRating = \frac{EconomyRating + BowlingStrikeRateRating + BowlingAverageRating - WPIRating}{4}$$

#(Negative
Attribute)

(c)Wicketkeeper Rating:

$$WicketKeeper_{Rating} = \frac{StrikeRateRating_{Batting} + AverageRating_{Batting} + MRA_{Rating} + BRPI_{Rating} + HittingAbility_{Rating} + WKA_{Rating} - OutRate_{Rating}}{7}$$

(d)Allrounder Rating:

$$AllrounderRating_{Batsman} = \frac{StrikeRateRating_{Batting} + AverageRating_{Batting} + MRA_{Rating} + BRPI_{Rating} + HittingAbility_{Rating} - OutRate_{Rating}}{6}$$

$$AllrounderRating_{Bowling} = \frac{EconomyRating + BowlingStrikeRateRating + BowlingAverageRating - WPIRating}{4}$$

#(Negative Attribute)

$$AllrounderRating = \frac{AllrounderRating_{Batting} - AllrounnderRating_{Bowling}}{2}$$

Note: Negative attributes are those ,whose value should be less in order to denote a better player.
For example , there are two bowlers who are having **BowlerStrikeRateRating** 100 and 90
respectively.

Chapter 9

Result of Processed Data

Squad Of Fifteen :-

	NAME	BattingRating	AllRoundersRating	BowlingRating	WicketKeeperRating	PlayingRole
56	Virat Kohli	97.423627	NaN	NaN	NaN	[Batsman]
27	Shikhar Dhawan	90.663617	NaN	NaN	NaN	[Batsman]
54	Shreyas Iyer	87.023447	NaN	NaN	NaN	[Batsman]
103	Rohit Sharma	82.391658	NaN	NaN	NaN	[Batsman]
40	Gautam Gambhir	71.005801	NaN	NaN	NaN	[Batsman]
121	Hardik Pandya	NaN	41.642982	NaN	NaN	[Allrounder]
11	Stuart Binny	NaN	36.604860	NaN	NaN	[Allrounder]
62	Ravindra Jadeja	NaN	28.022957	NaN	NaN	[Allrounder]
53	Kuldeep Yadav	NaN	NaN	22.863333	NaN	[Bowler]
77	Amit Mishra	NaN	NaN	29.116424	NaN	[Bowler]
7	Jasprit Bumrah	NaN	NaN	29.168689	NaN	[Bowler]
33	Yuzvendra Chahal	NaN	NaN	32.300374	NaN	[Bowler]
76	Mohammed Shami	NaN	NaN	34.807140	NaN	[Bowler]
26	MS Dhoni	NaN	NaN	NaN	111.900614	[Batsman]
134	KL Rahul	NaN	NaN	NaN	74.714895	[Batsman]

fig:9.1

From these calculated ratings , we have selected the players accordingly.For example, a batsman having highest batting rating(final rating) would be selected first as a batsman .We have selected 5 such batsman ,5 such bowlers ,3 allrounders(who can bat as well as bowl) and 2 such wicketkeepers.

CHAPTER 10

<u>Data Visualization</u>
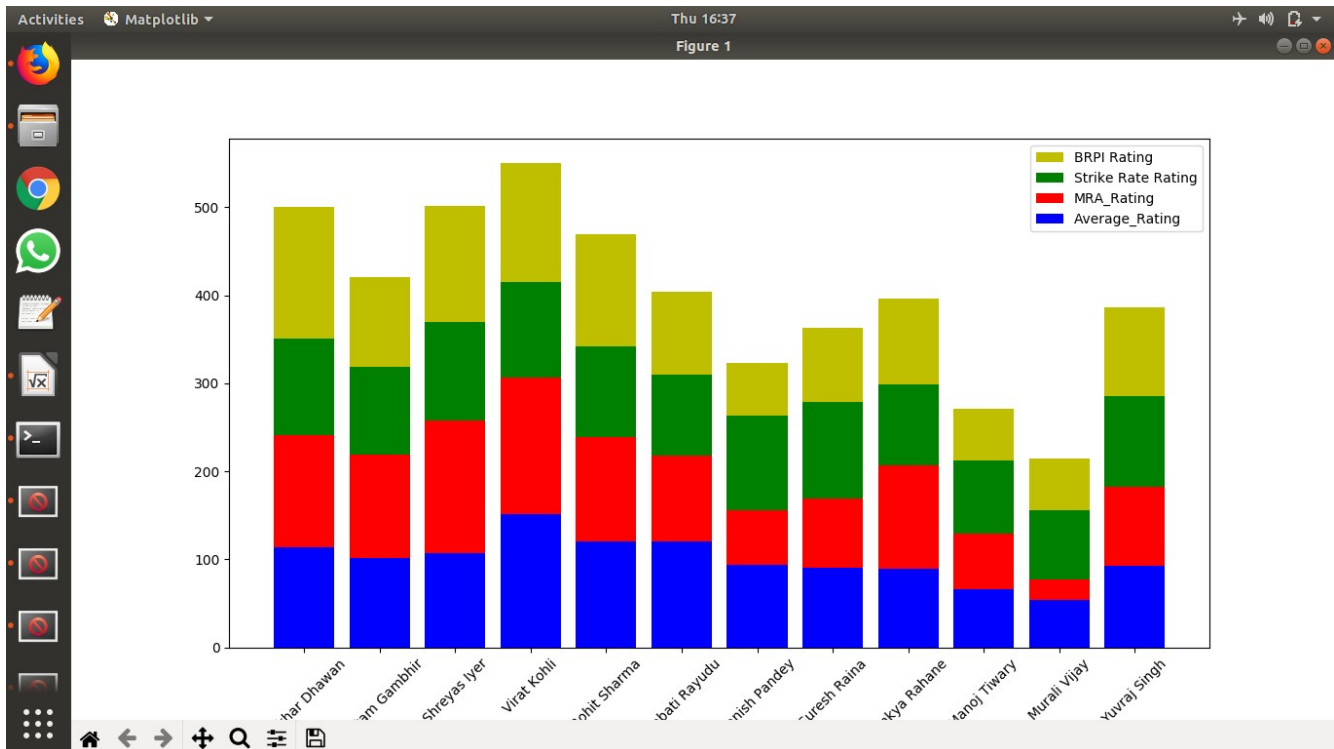
1. Batsmen :-

fig:10.1
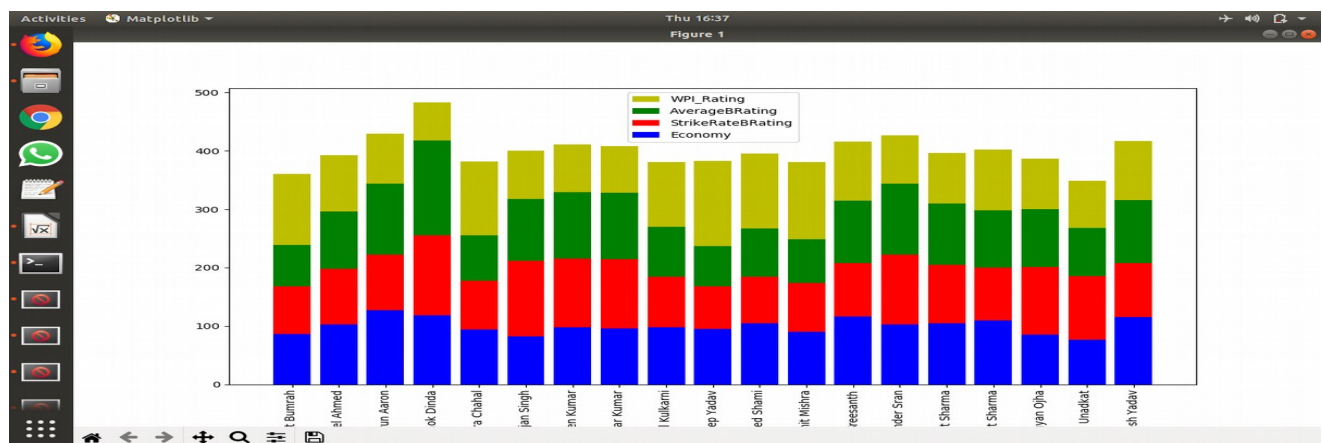
2. Bowler :-

fig:10.2

3.Batsman as all-rounder
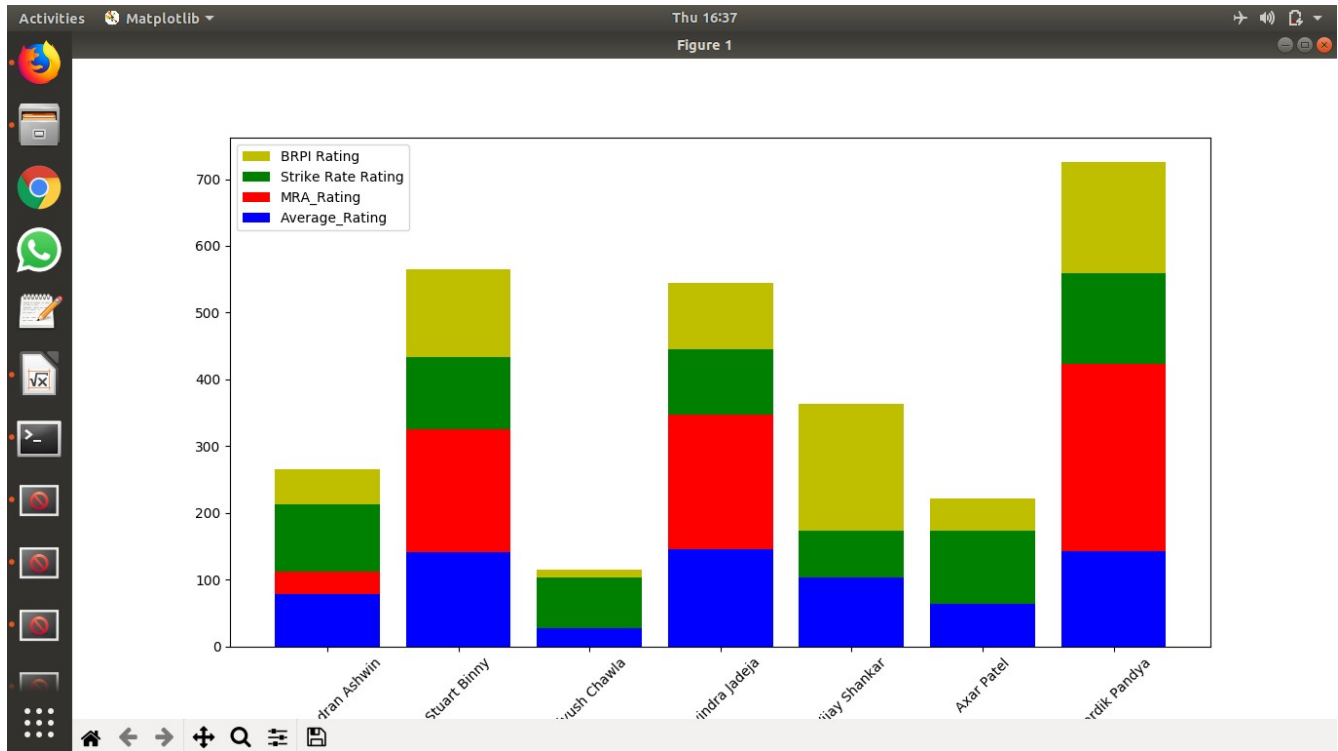

fig: 10.3

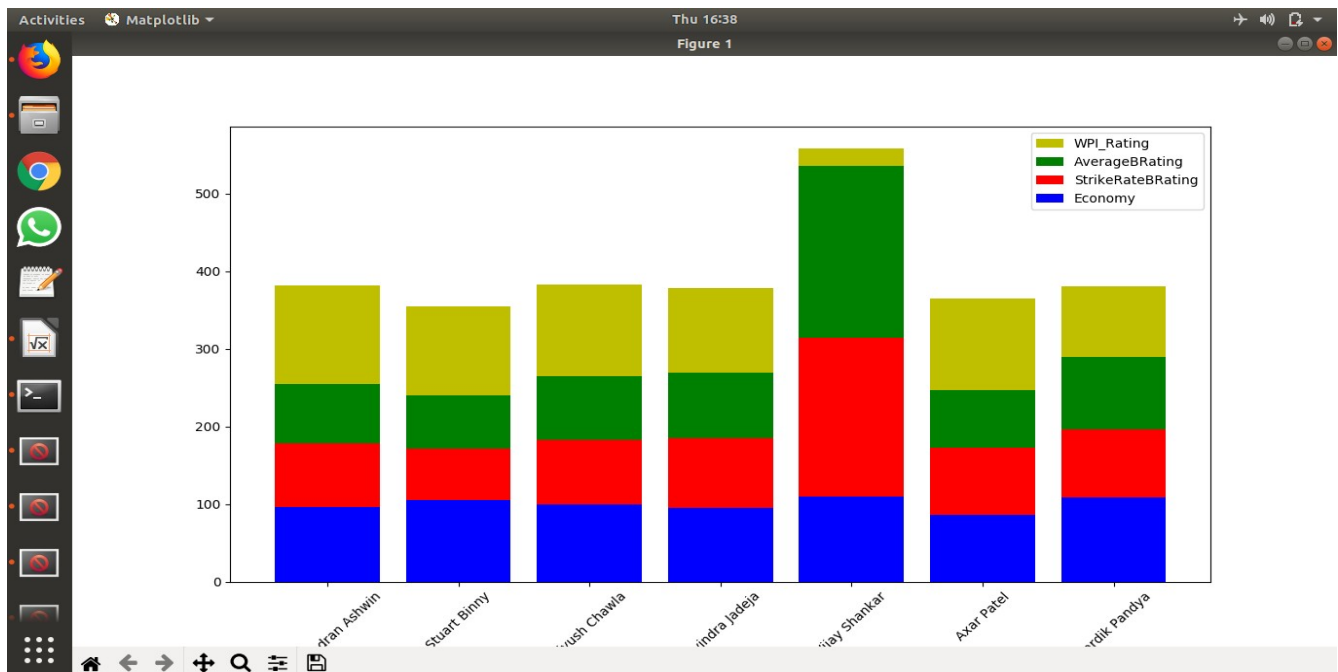# 4.Bowler as all-rounder

fig:10.4

CHAPTER 11

System Interface (GUI)

1.Register/Login Screen

fig:11.1
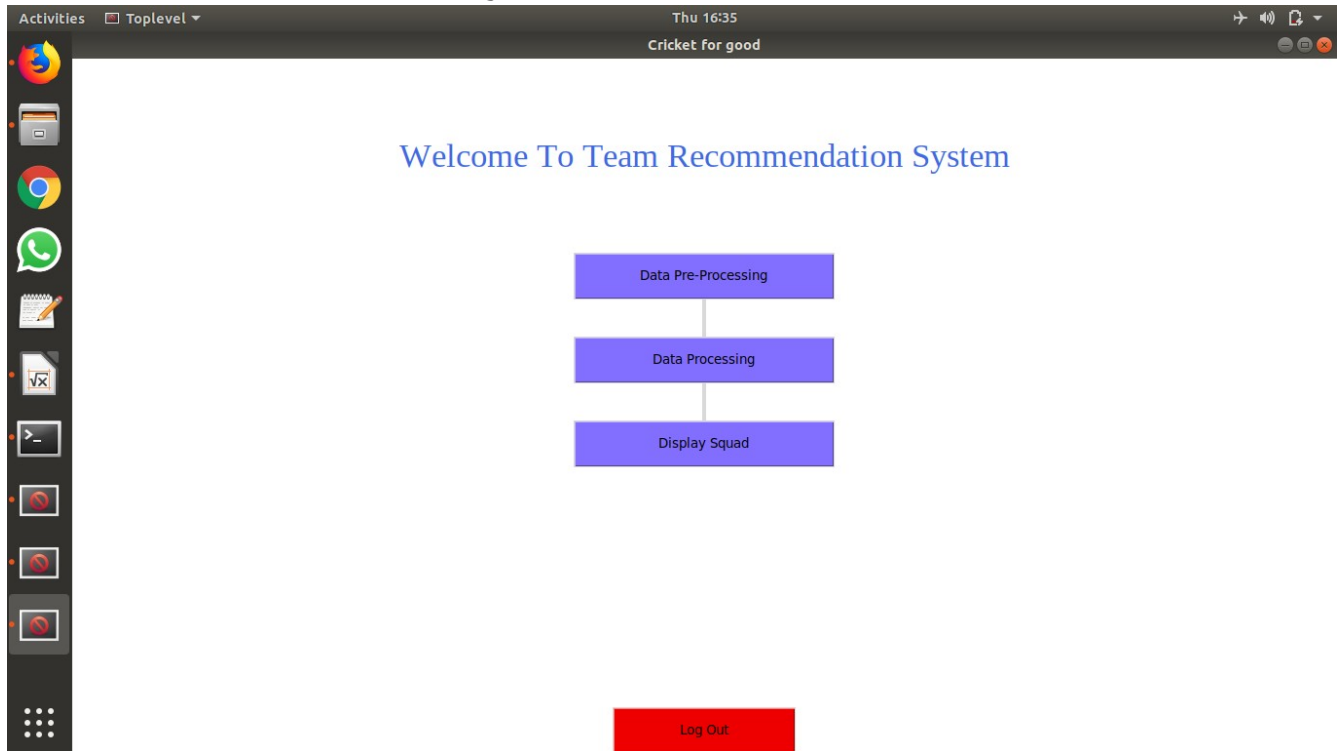
2.Login Section

fig:11.2

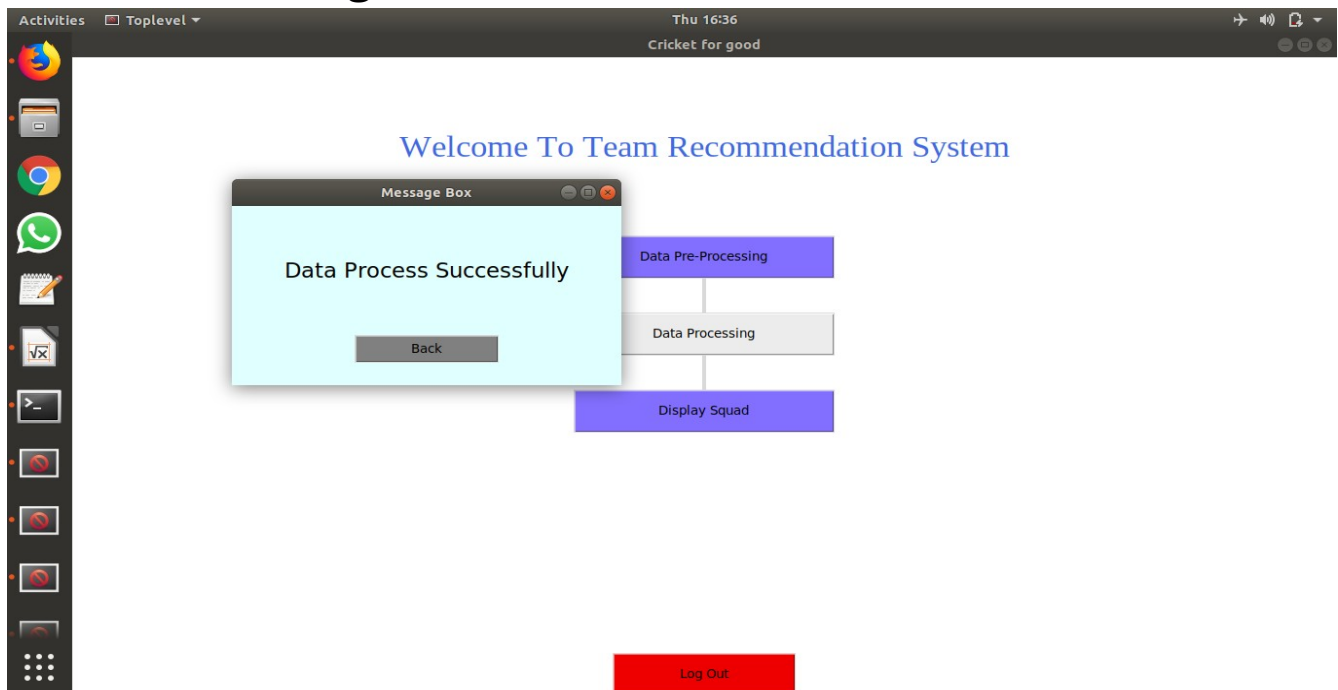# 3.Main task screen of system

fig:11.3

4.Data Processing

fig:11.4

5.Squad of fifteen(Final Output):

CHAPTER 12

<u>Conclusion</u>

In this project, we have proposed a model for cricket players squad recommendetion system using the statistics of the players, for a particular tournament or a particular series. We have devised a method to find the player ranking system using overall playing data of the players and selected players accordig to their ranks .

In our player recommendation system, we have used player performance measures by statistics and various mathematical formulae for recommending the players for the squad and reducing the loss due to the biased selection procedure of cricket players.

We have spotted some inferences from our report that there is some kind of gerrymandering in the selection process and we have trodden a step forward in order to make this process clear as crystal.

Though we cannot assure 100% accurate and efficient system,we have picthed the idea of doing something like that.

CHAPTER 13

<u>References</u>

1. For pandas references :

     https://pandas.pydata.org/

2. For datasets :

     https://www.kaggle.com/hverified/cricket-players-odi

3.For concepts related to cricket statistics and attributes calculation:

     https://content.iospress.com/download/journal-of-sports-analytics/jsa196?id=journal-of-sports-analytics%2Fjsa196

4. For Python3:

     (a)https://www.python.org/download/releases/3.0/
     (b)https://data-flair.training/blogs/python-tutorials-home/
     (c)https://www.edureka.co/blog/python-tutorial/

5.For User Interface (Graphical):

     <u>tkinter :</u>
     https://docs.python.org/2/library/tkinter.html