**INTERNATIONAL JOURNAL
FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Super Predictor of Indian Premier League (IPL) using Various ML techniques with help of IBM Cloud

K Rushikesh Reddy[1], A. Prem Sai[2], Chandrakanth V[3], Shiva Sumanth Reddy[4]

[1, 2, 3]BE Students, Department of Computer Science and Engineering, [4]Asst. Professor, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India

Abstract: Indian Premier League matches are one in every of the foremost important events in India. it's professional cricket league in India contested by eight teams representing the various cities in India. The paper focuses on the performance analysis of the eight contesting IPL teams supported the runs of the team, wickets, decisions usurping winning the toss and Duckworth Lewis rule analysis. The IPL data from 2008 to 2019 is used for the player analysis. The team performances are visualized graphically using data analytics to render the interpretation in an exceedingly good manner. The performance data using visual analytics helps in selecting players for future matches and provides additional information on player yet as team profiles. Almost every IPL team's Management use Analytics for better games. Not only team's owners, there are several betting and fantasy cricket platform, which are highly rely upon analytics for his or her success. Analytics can help all of them for his or her success. The research paper tries to predict the IPL matches using machine learning models with variables like match id, inning, batting team, bowling team, over, ball, batsman, non-striker, bowler is super over, wide runs, bye runs, leg by runs, bowling runs, penalty runs, batsman runs, extra runs, total runs, player dismissed, dismissal kind fielder. to appear out the result it uses different machine learning models like statistical method model, Random Forest. The results of the study shows that for IPL game, Teams, Venue, Winning Toss, Venue of the Match and Decision after winning the toss are important influencers to win a match. Different Machine Learning helps to predict outcome of a match. Right selection of Machine Learning Model helps to extend Accuracy of Prediction. From Different Classification Models, Support Vector Machine, Decision Tree and Random Forest are best to predict outcome of an IPL games. All of the subsequent gives almost 88% accuracy Level. The study has been conducted from data of Kaggle. Secondary data has been used for the analysis.

I. INTRODUCTION

A. Overview of our project:

Creating a legendary milestone in the world of a sport like cricket, the BCCI introduced Indian Premier League (IPL). As the volume of cricket enthusiasts across the globe increases, the curiosity on prediction of the future matches led to creation of many websites like Dream11, IPL fantasy league etc. IPL predictions can be right or wrong. Sports analytics is not new in cricket. Cricket being an attractive and profitable sports, there are so many stakeholders, involved in these sports. That's why decision-making process is very critical in cricket. Analytics is a supporting pillar for every decision maker in this game. The IPL predictor visualizes useful insights [4] and predicts outputs for instances provided by the user. The product caters to the needs of sports analysts, broadcasters, sports enthusiasts, business corporates [3], team management and so on [3]. The objective of this research is to observe impact of different Machine learning models in Prediction of an IPL match. Another objective of this study is to explore information, pattern related to Matches, Player etc. using descriptive analysis so as to increase the decision-making effectiveness. The main objective of our project is to analyse the IPL statistics of various matches and teams. Dashboards [4] are created for individual teams to provide an insight of their performance from the beginning of IPL. Future scores pertaining to a match are predicted using our web application. The project, developed as a web application, will satisfy the needs of a cricket expert as well as a beginner.

II. SCOPE OF THE PROJECT

IPL predictor provides useful insights obtained by performing exploratory and predictive analysis. The product serves both commercial and entertainment purpose. The product is user friendly where the insights are presented in comprehensible manner and the user can navigate freely through the pages, improving the reliability of the product. Features present in the product will give the user a better knowledge about IPL and features incorporated are:

A. Analysis about overview of IPL

Set of eleven questions were framed in such a way that it covers all the aspects of the game. The answers for the questions are found by performing exploratory analysis [1] on the dataset containing attributes like (toss winner, batting team, bowling toss, match winner and so on) with the help of SQL queries and python. The insights obtained are visualized in the form of graph using Cognos feature provided by IBM. .[1][2][3][9]

B. Team wise Analysis

Team wise analysis provides insights about the strengths and weaknesses of the team. The analysis includes happenings in different phases of game, highest run scorer and highest wicket taker and so on. The happenings in different phases of game (powerplay, middle over and death over) can help the team management and opponents to strategize their game plan, and the users an idea about the proceedings of the game. The analysis is done in IBM Watson and insights obtained are visualized using IBM Cognos.[2][3][9]

C. Teams' performance in particular season:

The feature gives an idea to the user about the proceedings of the game in a particular year. Data mining is performed to fetch results relevant to the particular team in the year prescribed by the future.[2][9]

D. Predicting Model

An important type of analysis in data analytics is predictive analysis. It provides insights about what is going to happen in the near future. This is predicted by a model trained using machine learning algorithms. Various ML algorithms like Random Forest were applied and linear regression...etc. [1][5][7][8]

E. Chatbot

Chatbot is created to answer minimal queries asked by the users. The chatbot is created using Watson assistant by adding relevant dialog and entities that can handle irrelevant questions and answer relevant questions. The chatbot can solve queries related to the structure of IPL, format of the game and so on. The above are features of the product that add value to the product and the user-friendly interface improves the reliability of the product [6].

III.DATASET DESCRIPTION

The section describes the IPL dataset providing information on team performance from 2008 to 2019. The section provides details on the attributes, instances, missing values in the dataset.

Number of attributes: 18

Number of instances: 756

ID –The attributes contain the information about the unique id for a match.

SEASON –The attribute contains the information about the year when the match has been conducted.

CITY - The attribute holds the information about the city where the match took place.

DATE – The attribute holds the information about the date when the match has been held.

TEAM 1 – The attribute describes that which team is going to bat first.

TEAM 2 – The attribute describes that which team is going to bat second.

TOSS_WINNER – The attribute holds the information about who wins the toss in that match.

TOSS_DECISION – The attribute contains the information about the decision (bat/field) taken by the toss winner. RESULT – The attribute contains information about the result (normal/tie) of the players.

DL_APPLIED – The attribute describes whether the. Duckworth Lewis (DL) rule is applied.

WINNER – The attribute holds the information about the winner of the match.

WIN_BY_RUNS – The attribute describes that which team had win by runs.

WIN_BY_WICKETS – The attribute describes that which team had win by wickets.

PLAYER_OF_MATCH – The attribute contains information about the man of the match.

VENUE – The attribute contains information about in which place the match has been played.

UMPIRE 1 – The attribute contains information about the names of the umpire 1.

UMPIRE 2 – The attribute contains information about the names of the umpire 2.

UMPIRE 3 – N/A.

IV. LITERATURE SURVEY

In this there are many past research paper which has been discussed. Predicting the Indian Premier League using Rabindra Lamsal machine reading, Ayesha Choudhary 2018 where there are variables used as home team, missing team, tos winner, location, referee, home team points, outside team points, power game. points, playing 11 players, Number of wickets taken, Number of balls with dots awarded, Number Four, Number Six, Hold number, stamp number. The Multilayer Vision segment has done much better than other class designers by accurately predicting 43 of the 60, 2018, Indian Premier League games. The Twenty 20 cricket format has a lot of unexpected features because one over can change the continuous pace of the game. Increased prediction accuracy in the game of cricket using the Kalpdrum Passi learning machine, Nirav Kumar Pandey in 2018 where the variables were the number of Innings, Batting Average, Strike Rate, Highest Score, Overs, Bowling Average, Bowling Strike Rate, Four / Five Wicket Haul, Location, Hundreds, Fifty, Hitting, Game Time, Hand, Game Type, Hitting, Hand Bowling where the analysis was a random jungle that created the most accurate guessing models in both hitting and throwing in all directions. Also, the accuracy of the models increases as it increases the size of the training data for all increases the size of the training data set in all algorithms except in the case of the Naïve Bayes bit where accuracy decreases as we increase the size of the training. set. The selection of the right players for the entire game played a big part in the team's victory. Accurate predictions about how much the batsman can gain and how many sticks a player may need in a game will help the team select players for a particular game. Live Cricket Score Prediction with Pramila M. Chawan Winner of 2018 where the variables used were Pitch, Toss, team strength, Home Ground Advantage where the result was a predictable model, the user predicts every game, and ends up watching the game checking that prediction it goes well. The project will therefore not only improve the popular cricket system but will also increase access to cricket in India. Cricket Analytics and Mr. Predictor Suyash Mahajan, Mr. Gunjan Kandhari, Mr. Salma Shaikh, Mr. Rutuja Pawar, Mr. Jash Vora, Ms. AR Deshpande in 2019 where other uses were City, Venue, Toss Result, Home Team, Away Team where previous data, it is beneficial for the owner to get details of the IPL game played by users who predict the winning percentage of the team and get player stats.

S.no	Name of the Paper	Author(s) & Journal name	Year	Variable Used	Result
1	Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning	Rabindra Lamsal, Ayesha Choudhary

Research chain | 2018 | home team, away team, toss winner, venue, umpires, home team score, away team score, power play score, playing 11 players, Number of wickets taken, Number of dot balls given, Number of fours, Number of sixes, Number of catches, Number of stampings | The Multilayer perception classifier outperformed other classifiers with correctly predicting 43 out of 60, 2018 Indian Premier League matches. The Twenty20 format of cricket carries a lot of randomness, because a single over can completely change the ongoing pace of the game. Indian Premier League is still at infantry stage, it is just a decade old league and has way a smaller number of matches compared to test and one-day international formats. |
| 2 | Increased prediction accuracy in the game of cricket using machine learning | Kalpdrum Passi, Nirav Kumar Pandey

International Journal of Data Mining & Knowledge Management Process. 8. 19-36. 10.5121/ijdkp.2018.8203 | 2018 | No. of Innings, Batting Average, Strike Rate, Highest Score, Overs, Bowling Average, Bowling Strike Rate, Four/Five Wicket Haul, Venue, Centuries, Fifties, Batting, Match Time, Hand, Match Type, Batting Position, Bowling Hand | Random Forest builds the most accurate prediction models for both batting and bowling in all the cases. Also, the accuracy of the models increases as increase the size of the training dataset for all algorithms except in case of Naïve Bayes for batting where the accuracy decreases as we increase the size of the training set. Selection of the right players for each match plays a significant role in a team's victory. An accurate prediction of how many runs a batsman is likely to score and how many wickets a bowler is likely to take in a match will help the team management select best players for each match. |
| 3 | Cricket Analytics and Predictor | Mr. Suyash Mahajan, Ms. Gunjan Kandhari, Ms. Salma Shaikh, Ms. Rutuja Pawar, Mr. Jash Vora, Ms. A. R. Deshpande

https://academia.edu/ | 2019 | City, Venue, Toss Result, Home Team, Away Team | From the previous data, it is beneficial to the owner to get the details of the IPL match played and the users who predict the winning percentage of the team and get the statistics of the player. |
| 4 | Prediction of Live Cricket Score and Winning | Pramila M. Chawan

Research gate | 2018 | Pitch, Toss, Team strength, Home Ground Advantage | a predictive model, a user makes a prediction on every game, and ends up watching that game to check if his prediction is going right Thus the project will not only improve the existing system of Fantasy Cricket, but will also augment the reach of Cricket in India |

## V. RESEARCH METHODOLOGY

A. Research Objective

The objective of this research is to observe impact of different Machine learning models in Prediction of an IPL match. Another objective of this study is to explore information, pattern related to Matches, Player etc. using descriptive analysis so as to increase the decision-making effectiveness.

B. Purpose of Study

Betting is an illegal activity for some region like India, but most of the country, is profitable business because Betting is not the only skill gambling, it is also a game of mathematical skills. No bets only, there are many arenas of the dream game like Dream11, my cricket circle where millions of users invest to get a good value profit where mathematical skills and data are present it is important. This study primarily aims to find out different statistical measures from IPL historical data and predict outcome of a match based on important factors to help users of betting sites and fantasy cricket league with scientific proof to support in their decision-making process.

VI.METHODOLOGY

Flow Chart

Overall Project Plan (Context Diagram):
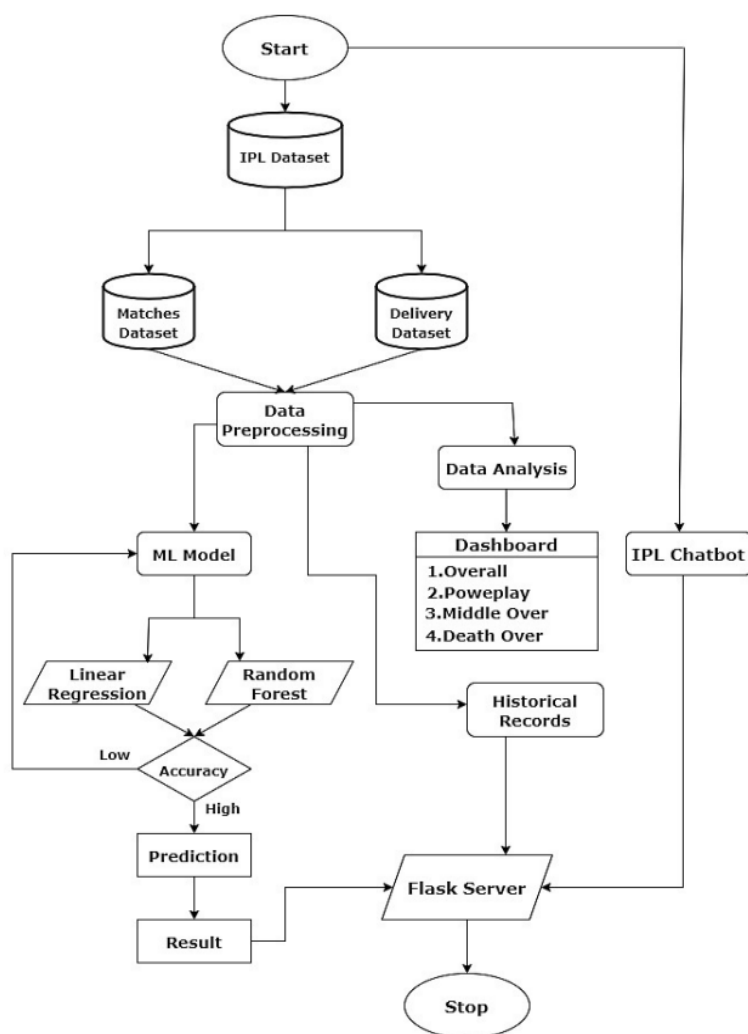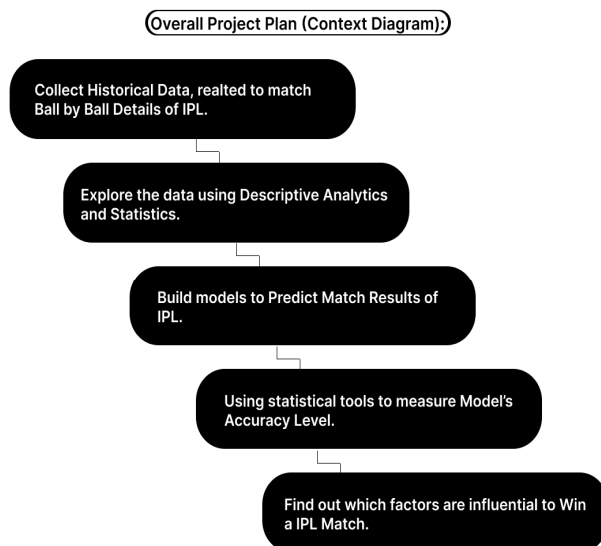
A. Sample Design

Secondary data is what is collected by someone. Some Common sources of secondary data include government public services department's Repository, libraries, internet searches and censuses. For this project, I have used Secondary data source to collect data.

B. Data Source

For this project work, data has been taken from Kaggle.com. Kaggle is subsidiary of Google LLC. It is an online community of data scientists and machine learning aspirant. It is also a repository of open-source data.

C. Analytical Methodology

This Project work focus on following two Analytical Methods-

1) *Descriptive Analytics:* Descriptive Analytics is a Method use in primary stages of any Analytics project to create a summary of historical data to mine useful knowledge, based on which further analysis can be done. In simple language, Descriptive analytics answered question like "what happened?". In our Project work, Descriptive model focus on two aspects:

a) Describe the data statistically.

b) Describe important factors.

2) *Predictive Analytics:* Predictive Analytics is a Method use in Advance stages of Analytics Projects to Predict Unknown future events based on different factors. Predictive Analytics use different Algorithms to build predictive models. Some of popular Algorithms, used in Predictive Modelling, are –Linear Regression, Random Forest, etc. For our project work, we will use predictive analytics to predict result of any matches.

VII. DATA FLOW DIAGRAM (DFD)
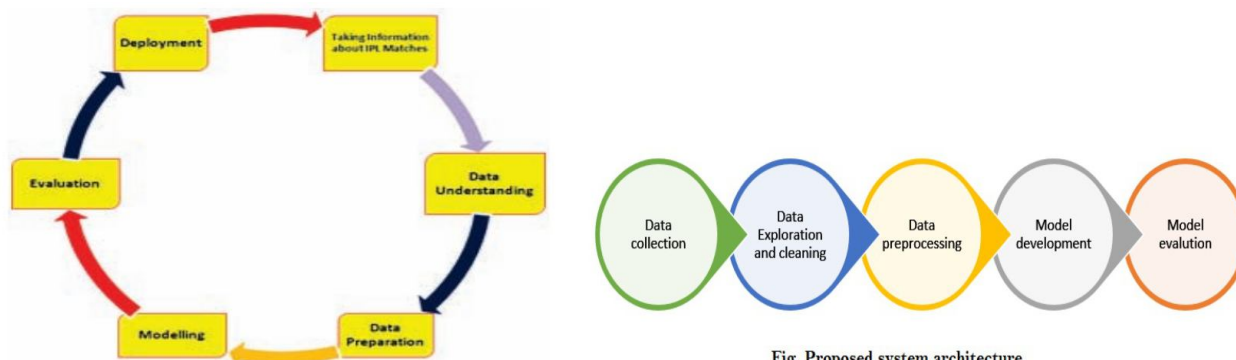
Fig. Proposed system architecture

VIII. TOOLS USED

A. IBM Cognos Tool

Cognos is a web-based Business Intelligence platform created by IBM. Provides a set of tools for analysing, reporting, and monitoring various metrics. IBM Cognos contains various components that are used to meet the needs of different information in any company. IBM Cognos has components such as IBM Cognos Framework Manager, IBM Cognos Cube Designer, IBM Cognos Transformer to help analyse data easily. Cognos will use this project as a descriptive mathematical tool.

B. Python

Python is an interpreted, high-level, general-purpose language. It's language uses an object-oriented approach that helps programmers write clear and logical code for any type of project. Thanks to its extensive library, Large Community, memory management, python is very popular among the Machine Learning community. In this Project Report, Python will be used for both descriptive and predictive analysis.

C. IBM Watson Studio (Watson Assistant)

IBM Watson Studio is software that simplifies developing, training, managing models, and deploying AI applications and used purpose, business and communication. It develops with many new features to build Artificial Intelligence applications. IBM Watson Assistant is a cloud service that allows for the development of a visual assistant in the software they are developing and branding the assistant as its own. Watson's assistant will help develop a visual assistant for this project.

IX. ANALYSIS AND INTERPRETATIONS

A. Descriptive Analysi

1) *Overall Statistics*

- Team that secured the greatest number of wins are Mumbai Indians. Similarly, the team secured the least number of wins are Kochi Tuskers Kerala. Furtherly, there are 52% chances of victory of winning the match if the team wins the toss. These analytics can be split furtherly into: winning toss, batting first and winning the match- where there are 82.3% chances of facing a defeat; winning toss, fielding first and winning the match where there are 65.7% chances of facing a defeat.
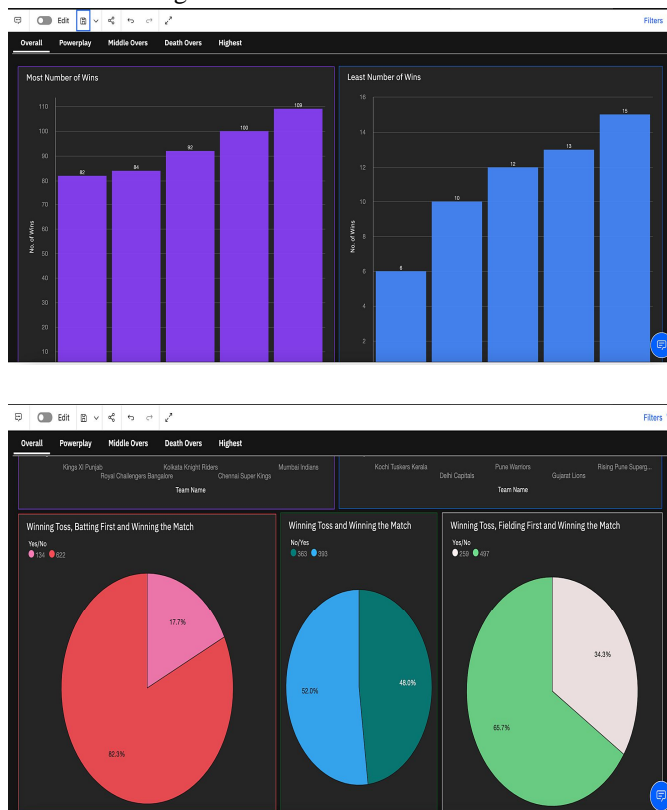
- Year in which each team secured maximum number of wins: CSK-2013 2. DC (Deccan Chargers)-2009 3. DC(Delhi Capitals)-2019 4. GL-2016 5. KXIP-2014 6. KTK-2011 7. KKR-2014 8. MI-2013 9. PWI-2011,2012,2013 10. RR-2018.

- Considering all the seasons, the stadium that hosted the greatest number of matches are Eden Gardens. Maximum number of MoM was won by CH Gayle. The largest margin of victory while defending a total was 146 runs, where MI won against DD in the year 2017.
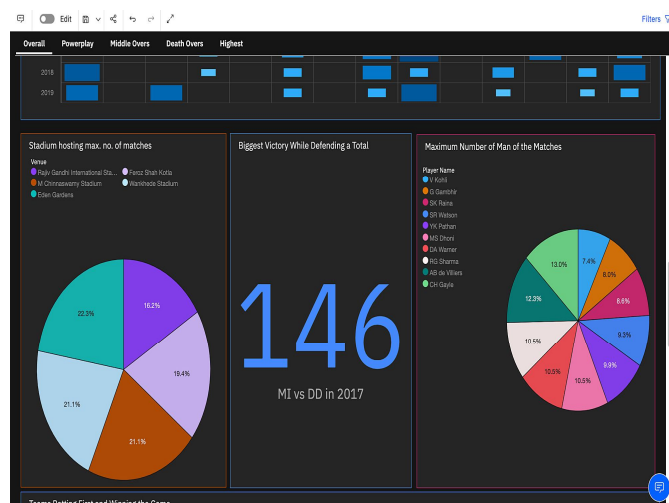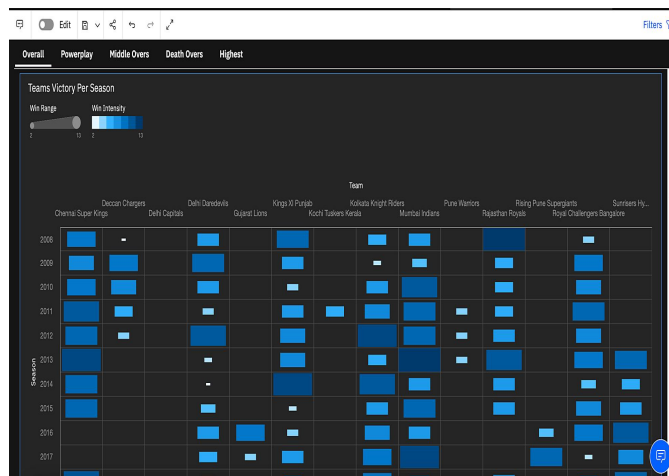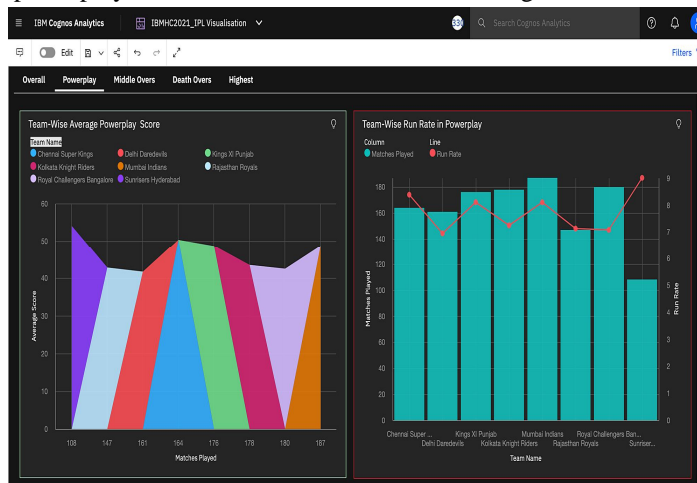

- Team that won the greatest number of matches batting first are Mumbai Indians. Again, team that won most number of matches fielding first are Mumbai Indians.
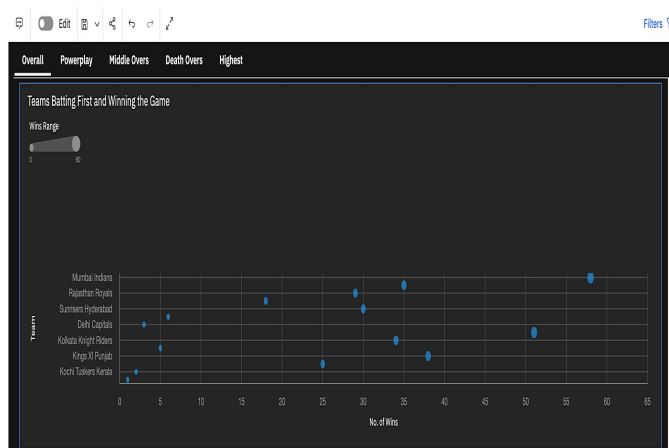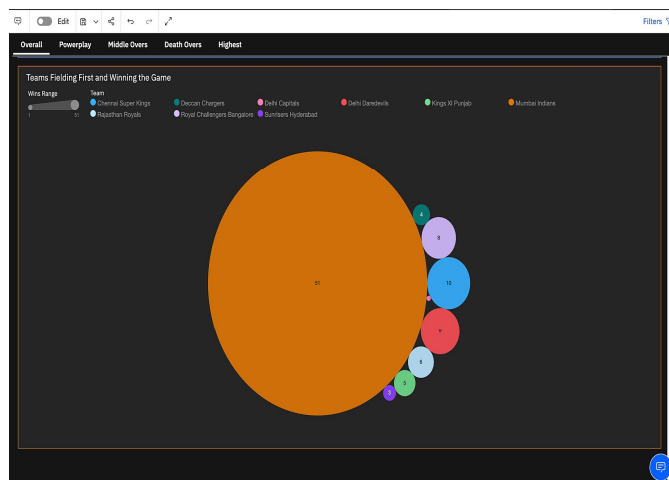
2) *Powerplay Statistics*

- CSK had the highest average powerplay score of 50.24 while SRH had the highest run rate in powerplay of 9.01.

- KKR, DD and KXIP had the highest average number of dismissals in powerplay, with KKR having the highest number of dismissals.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com
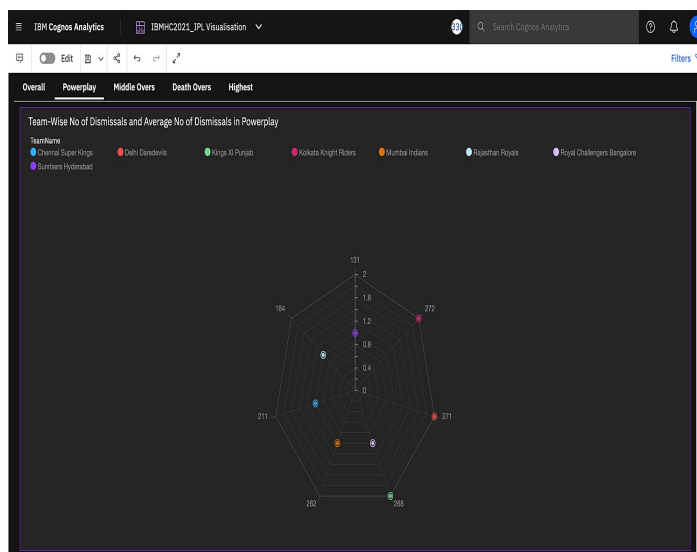
- CSK and RR had the highest average number of wickets taken in powerplay and the highest number of wickets taken.
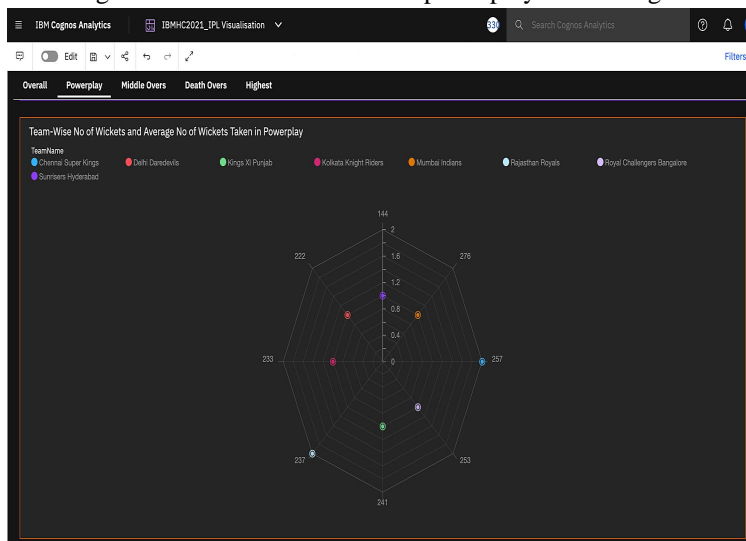
- Furtherly, two graphs with the ability to fitler team-wise, showed the number of times a player was dismissed and number of wickets taken by a player in powerplay.
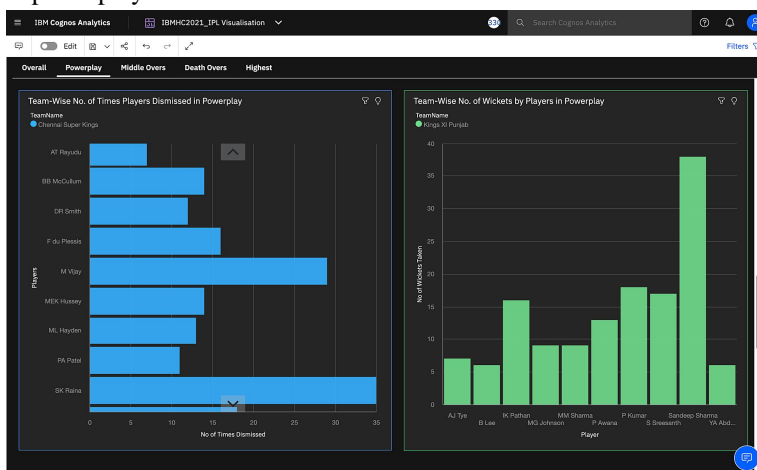
3) Middle Over Statistics

- CSK had the highest average middle over score of 64.74 while SRH had the highest run rate in middle overs of 68.56.
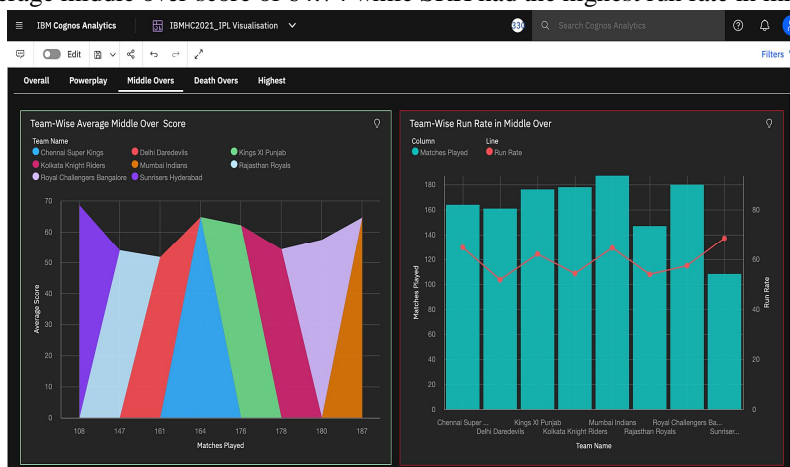
- All teams had the same average number of dismissals in middle overs while Mumbai Indian had the most number of dismissals.
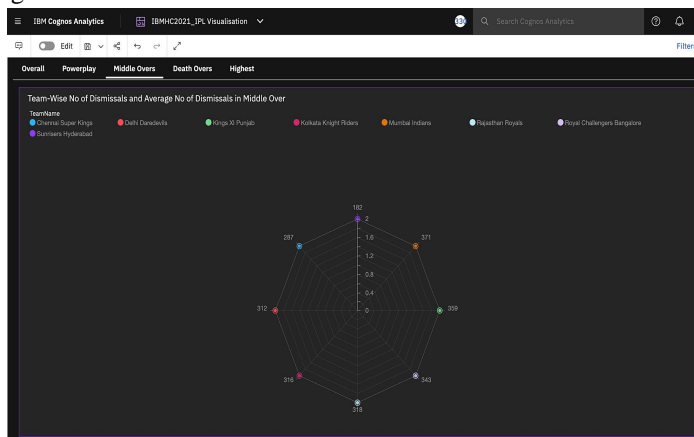
- All teams had the same average number of wickets taken in middle overs while Mumbai Indian had the most number of wickets taken.

- Furtherly, two graphs with the ability to fitler team-wise, showed the number of times a player was dimissed and number of wickets taken by a player in middle overs.
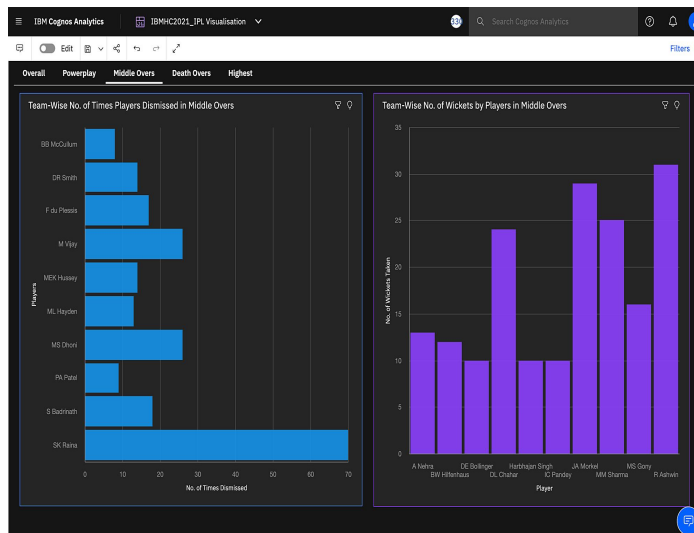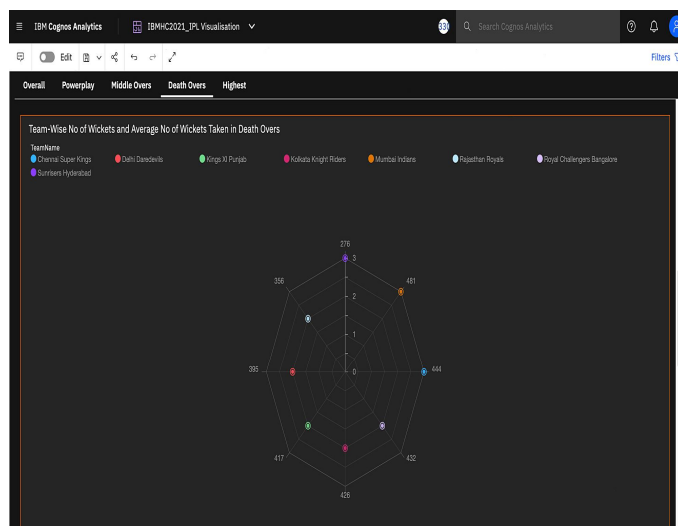
4) Death Over Statistics

- CSK had the highest average death over score of 25.03 while SRH had the highest run rate in death overs of 9.79.
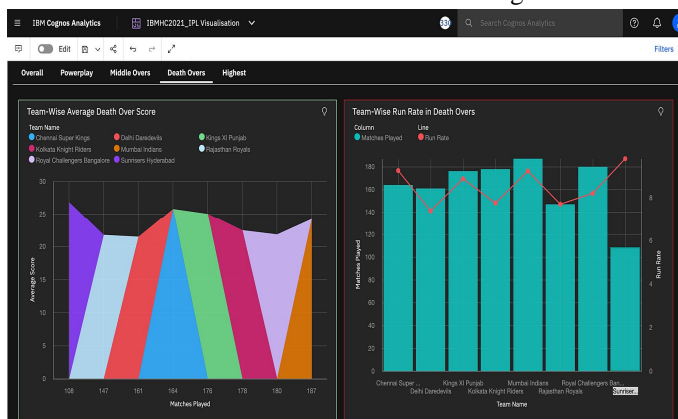
- SRH and KXIP had the highest average number of dismissals in death overs, while KXIP had the highest number of dismissals.
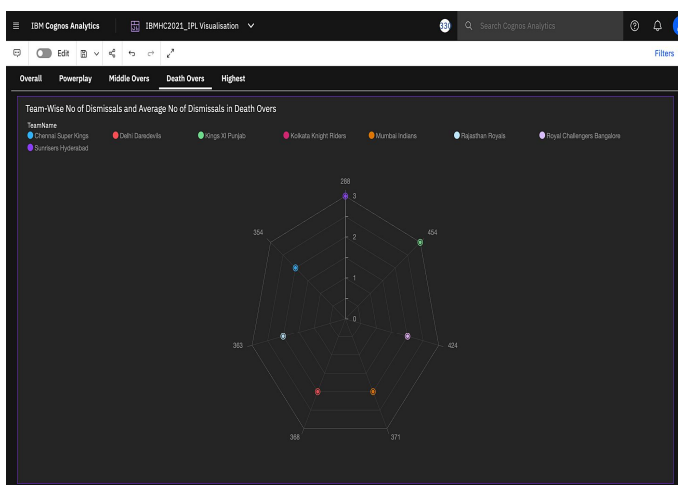
- MI and CSK had highest average number of wickets taken in death overs, while MI had the highest number of wickets taken.

- Furtherly, two graphs with the ability to fitler team-wise, showed the number of times a player was dimissed and number of wickets taken by a player in death overs.

5) *Highest Run Scorers and Wicket Takers*

- V Kohli had scored the most number of runs (5434) in the history of IPL.

- SL Malinga had taken the most number of wickets (187) in the history of IPL.

B. Predictive Analysis

1) Sequential Model

```
[29] from tensorflow.keras.models import Sequential
     from tensorflow.keras.layers import Dense,Dropout
     from tensorflow.keras.callbacks import EarlyStopping

     model = Sequential()

     model.add(Dense(43, activation='relu'))
     model.add(Dropout(0.5))

     model.add(Dense(22, activation='relu'))
     model.add(Dropout(0.5))

     model.add(Dense(11, activation='relu'))
     model.add(Dropout(0.5))

     model.add(Dense(1))

     model.compile(optimizer='adam', loss='mse')
```

```
[30] early_stop = EarlyStopping(monitor='val_loss',patience=25,verbose=1, mode='min')
     model.fit(x=X_train, y=y_train, epochs=400, validation_data=(X_test,y_test),callbacks=[early_stop] )

Epoch 127/400
1784/1784 [==============================] - 5s 3ms/step - loss: 142.8805 - val_loss: 139.8870
Epoch 128/400
1784/1784 [==============================] - 5s 3ms/step - loss: 142.6740 - val_loss: 139.0326
Epoch 129/400
1784/1784 [==============================] - 5s 3ms/step - loss: 142.4787 - val_loss: 138.3337
Epoch 130/400
1784/1784 [==============================] - 5s 3ms/step - loss: 142.9588 - val_loss: 139.2448
Epoch 131/400
1784/1784 [==============================] - 5s 3ms/step - loss: 143.1478 - val_loss: 139.4276
Epoch 132/400
1784/1784 [==============================] - 5s 3ms/step - loss: 143.5061 - val_loss: 139.7390
```
1s   completed at 12:06 PM
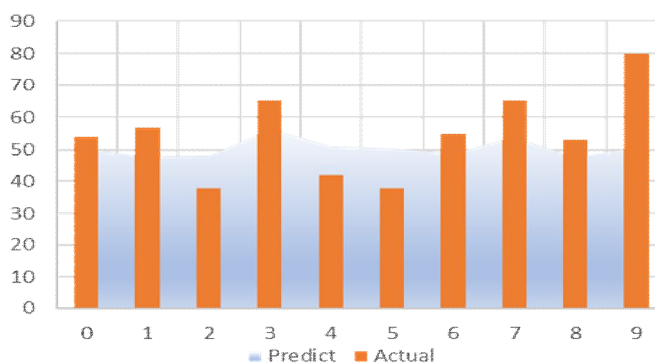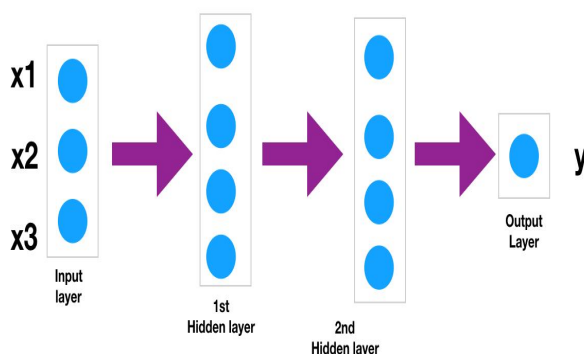
2) Model's Losses
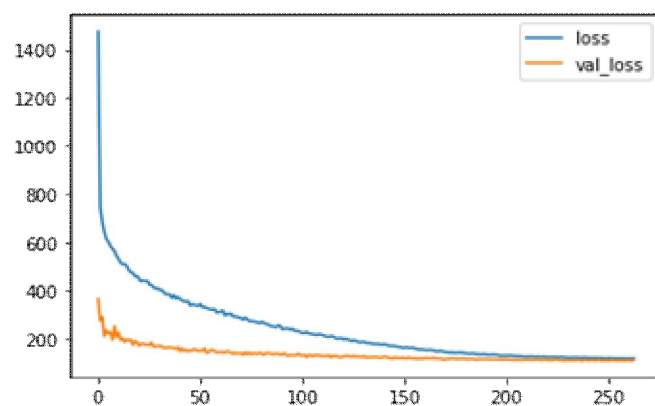
3)   Performance Metrics

```
[29] from sklearn.metrics import mean_absolute_error,mean_squared_error

     mean_absolute_error(y_test,predictions)

     8.72793745640346

     np.sqrt(mean_squared_error(y_test,predictions))

     10.609787723929372
```
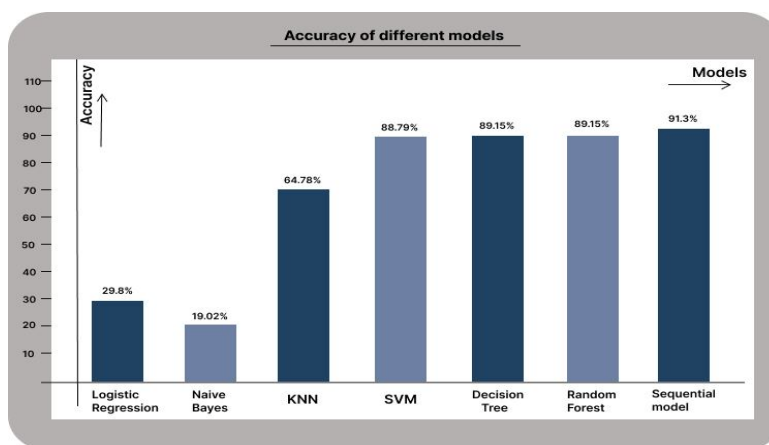
4)   Accuracy of Different Models

X. RESULT AND ANALYSIS

1)   From the analysis most IPL matches tool place in Eden Garden and Wankhade stadium and it is also important factor for winning the match.

2)   Mumbai Indians are the most successful groups who has won 109 games out of 11 times (2008-2020). Chennai Super kings is the second most successful group in the 100 wins, followed by the Kolkata Night Riders with 92 wins. Mumbai Indians won the most cups.

3)   Most of the time the man of the match of IPL game goes to Chris Gayle who owns the top number followed by AB de Villers and David Warner and Indians MS Dhoni is top on the list.

4)   In this it is clear that toss win or loss has least effect to win a match. After winning the toss, team wins the match is higher than when team win the toss but did not manage to win the match but however the decision to balling or batting first play a major role.

5)   In a different models of machine learning, we find that Sequential model has 91.3% accuracy.

XI. BUSINESS / SOCIAL IMPACT

As I am forecasting the predictive analysis of IPL, it will help the Sponsors to know whom they can sponsor and which players they can buy in auction.

XII.   CONCLUSION

Analytics can be used for Cricket match Prediction and its analysis in very easy way. For IPL game, Teams, Venue, Winning Toss, Venue of the Match and Decision after winning the toss are important influencers to win a match. Different Machine Learning helps to predict outcome of a match. Right selection of Machine Learning Model helps to increase Accuracy of Prediction. From Different machine learning Models, Sequential Model, Linear Regression and Random Forest are best to predict outcome of an IPL games. Sequential Model gives 91.3% accuracy. With this we can predict the IPL match score through machine learning models.

XIII. SCOPE OF FUTURE STUDY

I have taken only few factors as a predictor. but in cricket there are many factors which could have impact on a Match result. Player is one of those influencers. Every year IPL teams use to change their players; lots of new cricketer gets chances to play for the teams. So, in future studies Player can be used as a Predictor. Another important factor is pitch, in future pitch can be use as predictor. In this study I used only simple popular machine learning models, more complex models can be used to increase accuracy in future. Thus, using modelling effective decision-making can be accomplished through using tools and techniques in sports analytics.

REFERENCES

Extensive study about the topic was performed and various methodologies used in this domain were found. Predominantly there were two types of analysis: exploratory and predictive. Exploratory analysis visualizes events that have occurred in the past and provides meaningful insights that can be used for decision making

[1] Application of Machine Learning for Predictive Analytics: Indian Premier League (IPL) T-20 Cricket Matches 02 Subhashish Mahata, Neetu Kamra, Naina Kumari Agarwal

[2] INSIGHTS ON IPL TEAM PERFORMANCE USING VISUAL ANALYTICS S. Sharuka, R. Vani [IJESRT][Sharuka, et al., 8(11): November, 2019]

[3] T. H. Davenport, "What businesses can learn from sports analytics," MIT Sloan Management Review, vol. 55, no. 4, p. 10, 2014.

[4] Get started with Dashboards and Stories

[5] Score Prediction and Player Classification Model in the Game of Cricket Using Machine Learning Sonu Kumar, Sneha Roy

[6] Watson Assistant: Intelligent virtual agent

[7] Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning Rabindra Lamsal and Ayesha Choudhary [arXiv:1809.09813 [stat. AP]]

[8] PREDICTION ON IPL DATA USING MACHINE LEARNING TECHNIQUES IN R PACKAGE G. Sudhamathy and G. Raja Meenakshi (OCTOBER 2020, VOLUME: 11, ISSUE: 01)

[9] Analysis and Predicting Results of IPL T20 Matches Praveen Banasode, Minal Patil, Supriya-Verma

[10] C.Deep Prakash, C.Patvardhan and Sushobhit Singh," A new Category based Deep Performance Index using Machine Learning for ranking IPL Cricketers", Int. Jl. of Electronics, Electrical and Computational System IJEECS ISSN 2348-117X Volume 5, Issue 2 February 2016

[11] K Rushikesh Reddy, Chandrakanth V, and Premsai A. "Super Predictor of Indian Premier League (IPL) using Various ML techniques with help of IBM Cloud" International Journal Of Advance Research And Innovative Ideas In Education Volume 8 Issue 2 2022 Page 1156-1162

AUTHOR'S BIOGRAPHY

Author Name K RUSHIKESH REDDY

Second Author Name CHANDRAKANTH V

Third Author Name PREMSAI A

Fourth Author Name SHIVA SUMANTH REDDY

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)