



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** 1 **Month of publication:** January 2022

DOI: <https://doi.org/10.22214/ijraset.2022.40176>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Performance Analysis of a Cricketer by Data Visualization

Harshitha G¹, Sreehari G², Mahesh Kumar S³, Chinmai L⁴, Dr. Geetha D⁵

^{1, 2, 3, 4, 5}Department of Information Science and Engineering, BNM Institute of Technology

Abstract: Indian Premier League is a very competitive tournament where team selection is a very tricky and tedious procedure. Analysis of sports data and Prediction of each player's performance helps in filtering the best players. A novel method employing the techniques of Data Analytics and Data Visualization is used in this research paper to extract individual player performance from huge statistics and datasets. An application is created to bridge the space between selecting team, coaches, and team management and to give a better interpretation on player steadiness, scoring and further capabilities. In this paper, pandas library is used for data analysis and manipulation tool, Microsoft azure is used for performance prediction and HTML, CSS, flask for the front-end application. Additionally, various machine learning algorithms are applied on the same data to find the best fit. The proposed application can be beneficial for team managements and decision making

Keywords: Indian Premier League, Data analytics, Data Visualization, Prediction of player's Performance, Microsoft Azure.

I. INTRODUCTION

Sports analytics and Data Visualization has given a great platform for Player's selection, team managers, and to boost their on-field performance. Decision making and analysis, is the process of applying different algorithms on data to gain insights into prediction of the future. This data is made to undergo several algorithms, tools, and visualization techniques to make way for suggestion of the players to create the team. To build predictive models various machine learning techniques are applied.

Indian Premier League (IPL) was established in 2008. The league is based on a round-robin group and knockout format, has teams in major Indian cities. Each team management bids for almost up to 25 players and there can be only 4 foreign players in current playing 11 and at most 8 foreign players in total. It is difficult to find best squad for the upcoming seasons. In this paper the application is introduced to evaluate the performance of players. This tool provides a visualization of players' performance and helps in predicting scores. The developed model can help decision makers during the IPL matches to evaluate the strength of a team against another.

The work bestowed are as follows:

- 1) To analyse and interpret the raw data in an easily accessible format.
- 2) Selecting the most efficient machine learning algorithms by analysing their accuracy.
- 3) To predict each individual's performance.
- 4) To extract individual player performance from the huge dataset and represent visually in the form of graphs for better analysis.

II. LITERATURE SURVEY

Player data analysis is used in most sports. Sports analysis is full of statistics. This is the present and future of any professional in the field of sports. Both the opening of the stadium aids in player and team analysis and predicts relevant results. The task at [1] is facing the challenge of predicting the outcome of an IPL cricket match. A total of 644 game statistics were used in this study. Factors such as player strength and luck are used as important factors in prediction. The problem with this study is the dynamic and use of the relevant non-relationship database, the HBase application firmness. The authors at [2] analysed the performance of IPL players in terms of runs, the most successful team with wickets, Team performance in general, Man of the match with runs and wickets, throwing winners with runs, Toss winner with wickets, Analysis of Duckworth law winners. The full analysis is based on the presentation using a tableau. The results are predicted from different IPL teams and predicted in extreme analysis so that the winner of the match is predicted in almost any game situation. The accuracy of the selected number of adjectives for each group using feature selection was rated at [3]. In forecasting analytics, Put and Data are used in [4]. The batting and bowling datasets are modelled according to the players' statistics and features. Four multiclass phase algorithms were used and compared. The most accurate classification of both data sets was Random Forest and at least the most accurate was SVM. [5] highlights the performance of athletes especially batters and oversees analysis of Man of the Matches, Maximum Centuries Strikes by Batsmen, Top Batsmen,

Batsmen with Top Strike Rate, Top 10 Players with Maximum Runs. Refining and refining of data is done by modification, consolidation. The authors in [6] discussed how to analyse things to study the performance of cricket players and the findings of his study say that the force of battering dominates more than bowling. Studies show that the performance of throwers is one of the most important factors in changing the status quo. [7] described the player rating model at the IPL auction. Their model considered factors such as previous player bid price, player information, strike rate etc. Prakash, Patvardhan and Lakshmi [8] described the batting and bowling index to measure the performance of players in their models to predict the results of IPL matches. The mathematical method of proposing correct strike orders for ODI games is shown in [9]. In paper [10] the authors proposed a two-way model using the Naive Bayes and the Linear Regression Classifier. The first way is to predict the points of the first innings based on the current running rate, etc. The second method predicts the outcome of a given goal by a batting team. The authors in [11] predict the performance of the fourth-season IPL batsmen using the first three seasons. A Multi-Layer perceptron (MLP) neural network is used to predict previous activity. The outcome of a match by comparing the strength of two teams is predicted by the performance of each player [12] measured. They used algorithms to predict the performance of batsmen and bowlers from past and recent activity data. The so-called Combined Bowling Rate is a combination of three traditional bowling algorithms: bowling rate, strike rate and economy used to analyse bowlers in [13].

III.IMPLEMENTATION

A. Tools and Methodology

Indian Premier League has millions of fans across seas. It is one of the largest leagues played worldwide. Around 816 matches have been played from 2008 to 2020. We can find large amount of data on the internet which consists of all the stats of every match. Jupyter Notebook, an open-source application and python language is used for data exploration, data extraction, feature selection. Packages like Pandas, NumPy is used as a data analysis and manipulation tool. The analysed data is visualized using Am charts. Player performance prediction is done using Microsoft Azure. And the front end is developed using flask, a python web framework and is designed using html and CSS.

B. Data Collection

This section describes the datasets selected for the project. The datasets were collected from www.kaggle.com. They provide information on all the teams played from 2008-2020. There are two datasets used, namely Matches.csv and Ball-by-Ball.csv. In Matches.csv data set, information such as match ID, city in which the match was played, date, venue, player of the match, the two teams that took part, winner and decision of the toss, winner of the match, results, and names of the umpires of the matches are listed. Ball-by-Ball dataset provides details that include match id, innings, in which over which particular bowler bowled, who was at strike and non-strike, runs scored by the batsman, total runs scored, wickets that were taken, and the names of batting team and bowling team.

C. Pre-Processing of Data

Data pre-processing is the most essential part of a data science project. It consumes a major time dedicated to the project. Pre-processing of data includes getting rid of erroneous data, inconsistent data, formatting the data present and to fill the missing values. The unwanted data is removed including duplicate observations. It mainly deals with correction, standardization, and transformation of data. This is done to make sure outcomes are reliable.

```
In [3]: df=pd.read_csv("C:/Users/sreeh/OneDrive/Desktop/final_year_project/IPL_Ball-by-Ball_2008-2020.csv")
In [4]: df.head()
Out[4]:
   riker  bowler  batsman_runs  extra_runs  total_runs  non_boundary  is_wicket  dismissal_kind  player_dismissed  fielder  extras_type  batting_team  bowling_team
0  BB      AA              1           0           1           0           0           NaN              NaN              NaN      NaN      Kolkata Knight Riders  Royal Challengers Bangalore
1  nting   AA              1           0           1           0           0           NaN              NaN              NaN      NaN      Kolkata Knight Riders  Royal Challengers Bangalore
2  nting   Z              0           0           0           0           0           NaN              NaN              NaN      NaN      Kolkata Knight Riders  Royal Challengers Bangalore
3  nting   Z              1           0           1           0           0           NaN              NaN              NaN      NaN      Kolkata Knight Riders  Royal Challengers Bangalore
4  BB      Z              1           0           1           0           0           NaN              NaN              NaN      NaN      Kolkata Knight Riders  Royal Challengers Bangalore
```

Fig 1. Dataset before cleaning

df	id	inning	over	ball	batsman	non_striker	bowler	batsman_runs	extra_runs	total_runs	non_boundary	is_wicket	dismissal_kind	player_dis
0	335982	1	6	5	RT Ponting	BB McCullum	AA Noffke	1	0	1	0	0	0	0
1	335982	1	6	6	BB McCullum	RT Ponting	AA Noffke	1	0	1	0	0	0	0
2	335982	1	7	1	BB McCullum	RT Ponting	Z Khan	0	0	0	0	0	0	0
3	335982	1	7	2	BB McCullum	RT Ponting	Z Khan	1	0	1	0	0	0	0
4	335982	1	7	3	RT Ponting	BB McCullum	Z Khan	1	0	1	0	0	0	0

Fig 2. Dataset after cleaning

Fig 1. depicts the dataset and its cleaning in Jupyter notebook using python. And Fig2. depicts the first five entries in the dataset after cleaning.

D. Feature Selection and Extraction

Feature Selection is an essential phase where the parameters to analyze cricketer’s performance are to be decided. Parameters such as venue, opponent team, type of bowler to which the batsman got out, runs scored in powerplay vs runs scored in death overs, runs scored in first innings vs runs scored in second innings are considered for a batsman. Parameters such as venue, opponent team are considered for bowlers. These features are extracted from the cleaned but huge dataset. Two new datasets were made with respect to the project requirements. Batsman.csv and Bowler.csv datasets are the combination of Matches.csv and Ball-by-Ball.csv datasets. Batsman.csv consists of match id, number of runs scored in each match, number of runs scored in powerplays, number of runs in death overs venue of the match and bowling team. Number of wickets, match ID, venue and batting team are provided in Bowler.csv.

A	B	C	D	E
id	batsman	batsman_runs	bowling_team	Venue
335982	RT Ponting	20	Royal Challengers Bangalore	M Chinnaswamy Stadium
335982	BB McCullum	158	Royal Challengers Bangalore	M Chinnaswamy Stadium
335982	DJ Hussey	12	Royal Challengers Bangalore	M Chinnaswamy Stadium
335982	Mohammad Hafeez	5	Royal Challengers Bangalore	M Chinnaswamy Stadium
335982	SC Ganguly	10	Royal Challengers Bangalore	M Chinnaswamy Stadium
335982	CL White	6	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	MV Boucher	7	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	B Akhil	0	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	AA Noffke	9	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	P Kumar	18	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	Z Khan	3	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	SB Joshi	3	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	W Jaffer	6	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	JH Kallis	8	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	R Dravid	2	Kolkata Knight Riders	M Chinnaswamy Stadium
335982	V Kohli	1	Kolkata Knight Riders	M Chinnaswamy Stadium
335983	K Goel	24	Chennai Super Kings	Punjab Cricket Association Stadium, Mohali
335983	MEK Hussey	116	Kings XI Punjab	Punjab Cricket Association Stadium, Mohali
335983	JR Hopes	71	Chennai Super Kings	Punjab Cricket Association Stadium, Mohali
335983	SK Raina	32	Kings XI Punjab	Punjab Cricket Association Stadium, Mohali
335983	JDP Oram	13	Kings XI Punjab	Punjab Cricket Association Stadium, Mohali

Fig 3. Batsman.csv (Extracted dataset)

A	B	C	D	E
id	bowler	wickets	venue	batting_team
335982	P Kumar	0	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	Z Khan	1	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	AA Noffke	1	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	JH Kallis	1	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	SB Joshi	0	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	CL White	0	M Chinnaswamy Stadium	Kolkata Knight Riders
335982	AB Dinda	2	M Chinnaswamy Stadium	Royal Challengers Bangalore
335982	I Sharma	1	M Chinnaswamy Stadium	Royal Challengers Bangalore
335982	AB Agarkar	3	M Chinnaswamy Stadium	Royal Challengers Bangalore
335982	SC Ganguly	3	M Chinnaswamy Stadium	Royal Challengers Bangalore
335982	LR Shukla	1	M Chinnaswamy Stadium	Royal Challengers Bangalore
335983	B Lee	1	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	S Sreesanth	0	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	JR Hopes	1	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	IK Pathan	2	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	K Goel	0	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	PP Chawla	0	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	WA Mota	1	Punjab Cricket Association Stadium, Mohali	Chennai Super Kings
335983	JDP Oram	0	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab
335983	MS Gony	1	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab
335983	M Muralitharan	1	Punjab Cricket Association Stadium, Mohali	Kings XI Punjab

Fig 4. Batsman.csv (Extracted dataset)

E. Analysis and Interpretation

An important factor of data analysis project is to visualize the data in a visually appealing format, for the users to draw insights from the data. Graphically represented data are easy to be interpreted. The graphs contain all the details of the features that were extracted from the huge datasets. We have done this using amCharts, A JavaScript library for data visualization. We have further used different packages to get the exact analysis and visualization for teams and their players. The processing of these data for each individual happens in the backend program which is written with the help of flask. Pandas, an open source for data analysis is used as the data processing tool and to provide input output for both csv files. Dataframes are framed from the dataset and the computations are performed on the dataframes in order to obtain the data in the desired format to visualize graphs which is JSON. NumPy is used for numerical computing on the datasets

1) *Analysing Player vs Teams:* The figure 1 shows analysis of a player against all the teams. It exhibits each player performance with respect to runs scored irrespective of the venue, innings, or who the bowler is. This helps in selecting the best batting player for a particular opponent team.

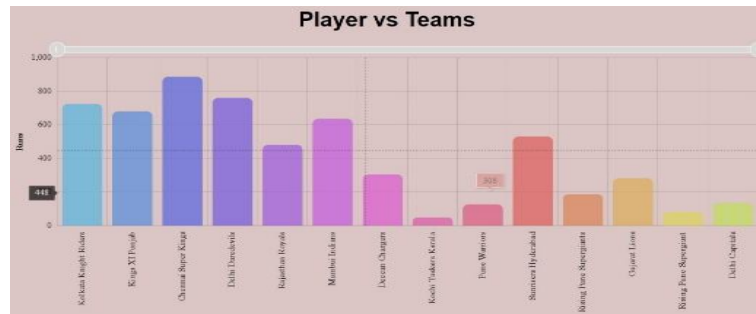


Fig 5. Analysing Player vs Teams

2) *Analysing Powerplay vs Death Overs:* The figure 2 shows player’s performance during powerplay and death overs. It gives a brief study of how a batsman scores during powerplay and in death overs. This gives an insight on which batsman must be selected for such overs.

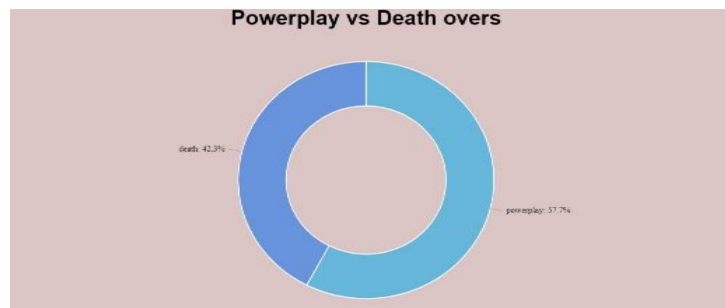


Fig 6. Powerplay vs Death Overs

3) *Analysing Player vs Venues:* The figure 3 shows analysis of a player in different venues. It exhibits each player performance with respect to all the stadiums he has played in. This helps in selecting an individual depending on where the match is being conducting.



Fig 7. Analysing Player vs Venue

- 4) *Analysing Player vs Innings*: The pie chart in figure 4 depicts analysis of player in different innings. It shows how well a player scores depending on the innings. This will help a captain to choose to bat or bowl after the toss.

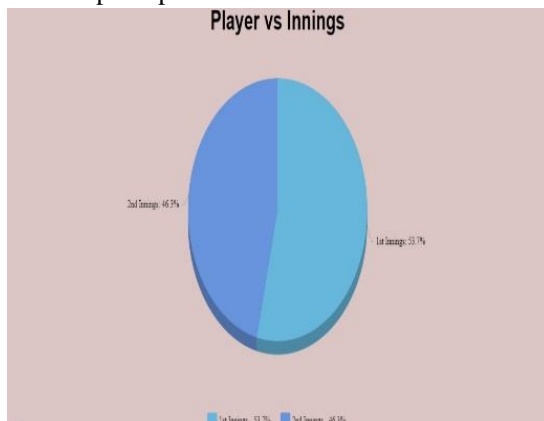


Fig 8. Analysing Player vs Innings

- 5) *Analysing Player vs Bowlers*: The Figure 5 depicts analysis of player against different Bowlers. It shows scores of a single player against all bowling types, such as right arm medium pace, left arm medium pace, right arm leg off, left arm leg off, leg spinner etc.

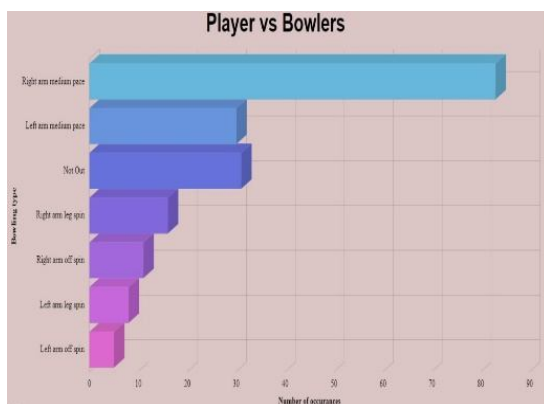


Fig 9. Analysing Player vs Bowlers

- 6) *Analysing Bowler against Venues*: The Figure 6 shows analysis of Bowler against venues. It describes how many wickets a bowler takes in different venues irrespective of the batsman and opponent teams.

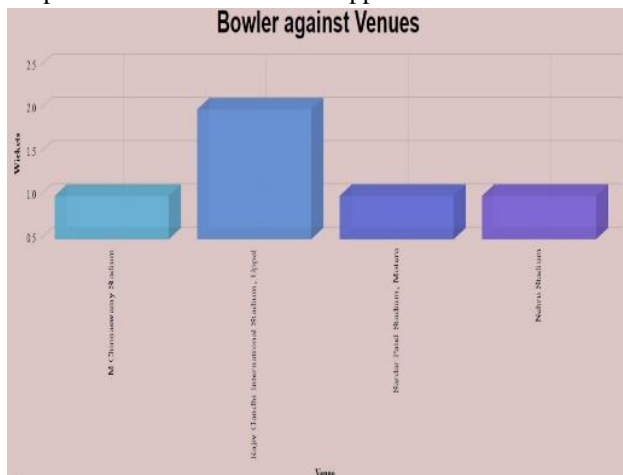


Fig 10. Analysing Bowlers against Venues

7) *Analysing Bowler against Team:* The Figure 7 depicts analysis of Bowler against teams. It shows how many wickets a bowler takes against different teams irrespective of the venue.

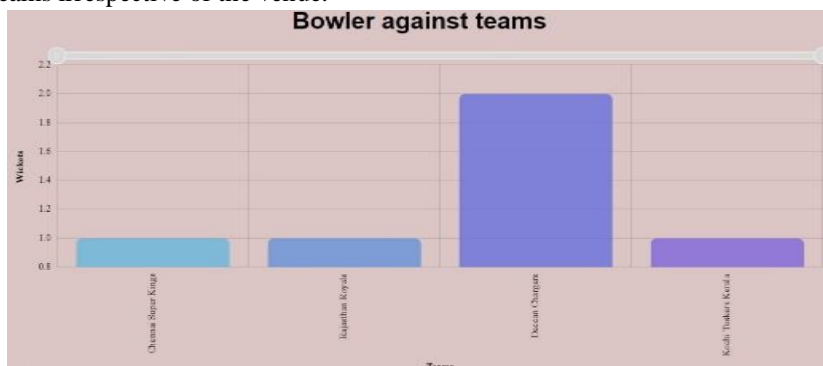


Fig 11. Analysing Bowlers against Team

F. Algorithm Analysis and Prediction

The data extracted from the cleaned dataset is used to build machine learning model using Microsoft Azure platform. The dataset is being applied to different regression algorithms like Linear Regression, Poisson Regression, Bayesian Linear Regression, Neural network Regression, Boosted Decision tree and Decision Forest Regression and the performance of each algorithm is analysed as shown in the figure 7.

ALGORITHM	COEFFICIENT OF DETERMINATION
Linear Regression	0.120875
Poisson Regression	0.081029
Bayesian Linear Regression	0.125567
Neural Network Regression	-0.002445
Boosted Decision Tree	0.017992
Decision Forest Regression	0.099266

Fig 12. Algorithm Analysis

The algorithm analysis as depicted in the figure provides the coefficient of determination for each model against which the data is tested. In a regression model coefficient of determination is a statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable. Using this algorithm analysis, the Boosted Decision Tree is used as the prediction model as it provides the highest accuracy having the value of coefficient of determination very close to zero, and it was built in the Machine Learning studio in Azure platform which is shown in figure 8. The dataset is divided into training data and testing data and fed into the model as depicted in the figure.

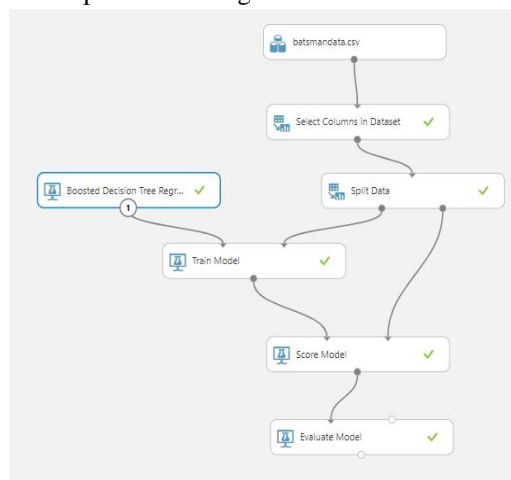


Fig 13. Prediction Model

IV. RESULTS AND DISCUSSION

The user interface of web application is shown in Figures below. It consists of three webpages. In the first page as shown in figure 10, all the eight teams are visible. The user must select one among them to continue to the next web page. As shown in figure 11, the webpage contains all the players of the respective team. Upon choosing a player the analysis and visualization webpage opens. In this page there are seven different charts initiated depending on the previous choices.



Fig 14. Teams Page



Fig 15. Players Page

The web application also provides an option for predicting the player's performance. This button is provided in the analysis page. On clicking this button, the name of the player is taken implicitly. Bowling team and Venue are to be chosen from the list provided and these three parameters are taken as inputs for prediction. The result page appears once the inputs are submitted which outputs the predicted score of the particular batsman. The Prediction form and result is shown in separate windows which are depicted in figures 11 and 12.

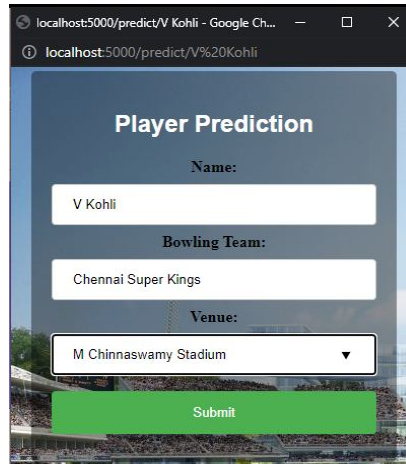


Fig 16. Player Prediction user form

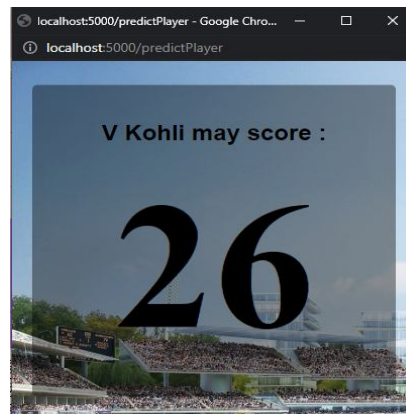


Fig 17. Score Prediction

V. CONCLUSIONS

In this proposed work, the performance analysis of cricketers in IPL from season 2008-2020 has been visualized. The project highlights the player performance with respect to venue, innings, death overs, powerplay overs, and type of bowlers. For selecting best player for particular match against team and venue, an accurate prediction of batsman runs prior to the commencement will help the team management in selecting the best players for each match. Depending on the stats and characteristics we have modelled batting and bowling datasets. The best fit algorithm is found out for the dataset and the performance of the player is predicted using Microsoft Azure

VI. FUTURE ENHANCEMENT

Future scope of the project is to include other features like position of the batsman. The project can be extended to predict the bowler's performance by predicting the number of wickets a bowler might take. We also want to extend the project's dataset of other cricket matches like international Cricket, Big Bash Leagues etc.

REFERENCES

- [1] Vidit Kanungo, Tulasi B., "Data visualization and toss related analysis of IPL teams and batsmen performances", International Journal of Electrical and Computer Engineering, Vol. 9, No. 5, October 2019.
- [2] Shubhra Singh, Parmeet Kaur, "IPL Visualization and Prediction Using Hbase", Information Technology and Quantitative Management, 2017.
- [3] S. Sharuka, R.Vani, "Insights on IPL Team Performance using Visual Analytics", International Journal of Engineering Sciences & Research Technology, November, 2019.
- [4] Kasukruti Raviteja, Ganesh Kumar Macha, Dr. GR Anantharaman, "Predicting and Analyzing the Performance of the IPL Cricket Using Regression Models", Complexity International Journal, Volume 23, Issue 03, Dec 2019.
- [5] Kalpdrum Passi and Niravkumar Pandey, "Predicting players performance in one day international cricket matches using Machine Learning", 8th International Conference on Computer Science, Engineering and Applications, February 2018.
- [6] Sricharan Shah, et al., "A Study on Performance of Cricket Players using Factor Analysis Approach," International Journal of Advanced Re-search in Computer Science, vol. 8, no. 3, 2017.
- [7] D. Parker, P. Burns and H. Natarajan, "Player valuations in the Indian Premier League", Frontier Economics, vol. 116, October 2008.
- [8] C. D. Prakash, C. Patvardhan and C. V. Lakshmi, "Data Analytics based Deep Mayo Predictor for IPL-9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10, October 2016.
- [9] M. Ovens and B. Bukiet, "A Mathematical Modelling Approach to One-Day Cricket Batting Orders," Journal of Sports Science and Medicine, vol. 5, pp. 49 5-502, 15 December 2006.
- [10] Tijender Singh, et al., "Score and Winning Prediction in Cricket through Data Mining," International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Oct 2015
- [11] Hemanta Saikia and Dibyojyoti Bhattacharjee, "An application of multilayer perceptron neural network to predict the performance of batsmen in Indian premier league," International Journal of Research in Science and Technology, vol. 1, no. 1, 2014.
- [12] M. G. Jhanwar and V. Pudi, "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD), 2016.
- [13] H. H. Lemmer, "The combined bowling rate as a measure of bowling performance in cricket," South African Journal for Research in Sport, Physical Education and Recreation, vol. 24, no. 2, pp 37-44, Jan 2002.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)