International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 17th Sept 2021. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-09

DOI: 10.48047/IJIEMR/V10/I09/04

Title **INDIAN PREMIER LEAGUE MATCH SCORE PREDICTION USING MACHINE LEARNING ALGORITHMS**

Volume 10, Issue 09, Pages: 12-20

Paper Authors
V.Sowjanya,A.Pramod,G.Sandhya,K.Uma,K.Hema Vinay Kumar

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

INDIAN PREMIER LEAGUE MATCH SCORE PREDICTION USING MACHINE LEARNING ALGORITHMS

[1]V.Sowjanya,[2]A.Pramod,[3]G.Sandhya,[4]K.Uma,[5]K.Hema Vinay Kumar
[1]Associate Professor,Department of CSE, PSCMRCET, Vijayawada.sowjanyav@pscmr.ac.in
[2]IV Year CSE Student,PSCMRCET, Vijayawada.pramodarikati@gmail.com
[3]IV Year CSE Student,, PSCMRCET, Vijayawada.sandhyachinnu400@gmail.com
[4]IV Year CSE Student, PSCMRCET, Vijayawada.umakoyya9398@gmail.com
[5]IV Year CSE Student,PSCMRCET, Vijayawada.hemavinaykumar@gmail.com

Abstract:

In cricket, especially the 20-20 strategy is most watched and valued by individuals, where it's illogical for anyone to examine who will oversee the game until the last wad of the last wrapped up. In India, The IPL began in 2008 and eventually this is the prevalent prominent Twenty20 class inside the planet . So we chose to energize an AI model for foreseeing the deferred results of its matches. Overpowering during a game relies on many key factors sort of a trademark natural components advantage, past introductions consequently ground, records at an equivalent scene, the general information on the players, record with a picked obstruction, and likewise the general current kind of the get-together and moreover the individual player. This assignment gives about the key factors that influence the aftereffects of the match and accordingly the discretionary woods model that the greater part vivaciously fits this information and gives the single measures. Cricket, the standard and thoroughly played game across India which has the prevalent critical fan base. IPL follows a 20-20 game-plan which is unimaginably whimsical. IPL match marker could even be a ML based guess approach where the informational indexes and past nuances are prepared out and out measurements covering terrifically significant factors, for example, Toss, environment , Captains, Favorite Players, Opposition Battle, Previous Stats and so forth.

Key Words: The Indian Premier League; Machine learning, rectilinear relapse , Runs scored, Prediction, Winning likelihood, Random Forest Classification, Prediction System, Score Prediction, Cricket, Decision Tree, Gradient Boosting, Linear Regression.

Introduction

IPL is an expert cricket class dependent on Twenty20 arrangement and is represented by the BCCI. The class happens each year with taking part groups' names addressing different urban communities of India. There are numerous nations dynamic in getting sorted out Twenty20 cricket leagues. The IPL (Indian Premier League) is a 20-20 cricket class in India where eight groups (addressing eight urban areas in India) play against one another.

In India, The IPL started in 2008 and presently it's the chief notable T-20 pack inside the world. So we chose to encourage an AI model for anticipating the deferred outcome of its matches. Overpowering during a game relies on many key factors kind of a living space advantage, past introductions thusly ground, records at a vague scene, the general information on the players, record with a particular obstruction, and the overall current kind of the social gathering besides the individual player. i.e., eventually in International, T20 and Test Matches.

IPL is a 20-20 cricket competition association set up with the goal of advancing cricket in India and along these lines supporting youthful and skilled players.A method of foreseeing the result of matches between different groups can help with the group determination measure. The made models can help supervisors during the IPL matches to guage the strength of a social occasion against another.

II. LITERATURE WORK

The related work for this project is taken from and implemented is “Foreseeing Cricket Score By Using Machine Learning Concepts" H.V Ramachandra Prof., Jspm's, RSCOE. R.R.Kamble Assistant Prof., Jspm's, RSCOE from this reference able to get the minimum knowledge of this project. Daniel etal.have examined many features to anticipate the match champ before the start of the

match by training the chosen features on machine learning models. To achieve this they have applied many machine learning algorithms on test and preparing dataset of various sizes which are Random Forest, SVM, Naive Bayes, Logistic Regression and Decision Tree.

Sandesh et al. in this paper performance of the players in terms of the strength and weaknesses is used to predict the match outcome.This will help the mentors, skippers of the group, selectors and directors by thinking about different factual proportions of the players.

RameshwariLokhande and P.M.Chawan during this paper, match outcome prediction is completed while the match is ongoing i.e. live match prediction. Thus, features like number of wickets fallen, match venue, team ranking, pitch report, habitat advantage, etc, are considered, instead of pre match features like past player performance, past team experience, etc.

III. EXISTING SYSTEM AND DISADVANTAGES

Nowadays, the cricket records are stayed aware of and expected genuinely, no valid data used for cricket. and their present methods essentially perceive cricket score expect subject to the dataset and projected score.

Disadvantages:

- We can't guarantee the score expected is correct.
- Time burning-through measure.
- Not simple to utilize system.

There has been a ton of work done related to cricket. Huge proportion of datasets have been investigated,
and information, for instance, and the kind of cricket have been isolated to help people with noticing law execution. On the other hand, the past related work

IV. PROPOSED APPROACH

The fundamental point is to anticipate the match

result, execution of each player upheld the recorded information. To understand a solid exactness, we'd prefer to break down an outsized measure of information . In this manner, the underlying advance of the execution was to gather information for all conceivable matches.

After studying the research papers, it's been observed that the prediction is typically avoided considering the dynamic nature of the cricket game. We are proposing a system Cric First Predictor which will consider the dynamic nature of the sport . The accuracy of the models used by authors are slightly on the lower side. We are trying to improve the accuracy of the algorithm by considering various parameters.

V. System Architecture

The below figure of our task shows that we removed the vital highlights from the crude information and afterward split it into a preparation dataset and testing dataset. The AI model then, at that point prepared utilizing the preparation dataset and tried utilizing the testing dataset. The model then, at that point predicts the match result.

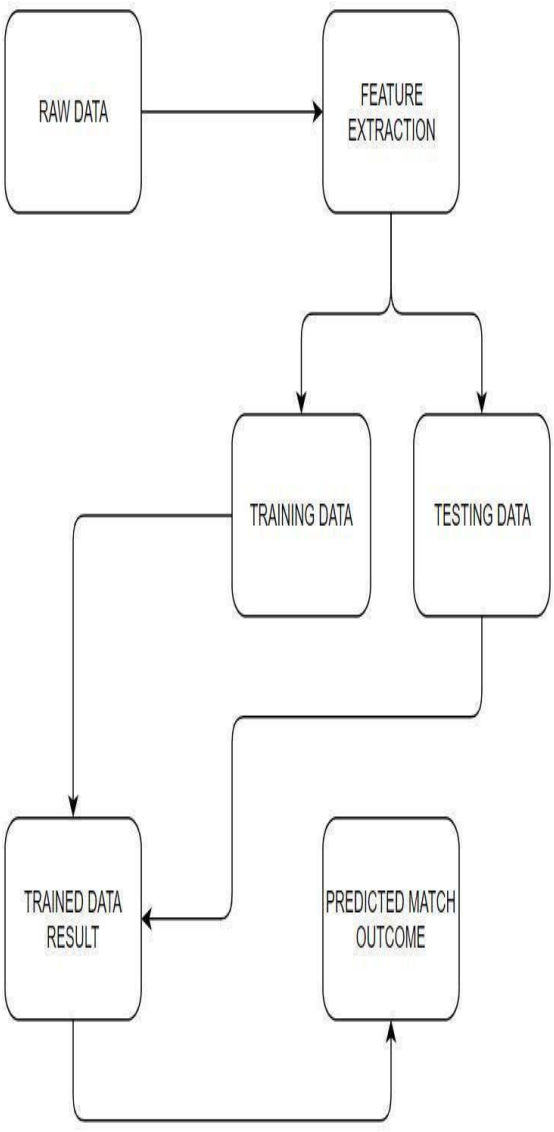

Fig.1.SystemArchitectureofIPL

VI. METHODOLOGIES

The approach comprises of 4 fundamental stages- Data Preprocessing, Data Cleansing, Data Preparation, Encoding the information . At first, the seven IPL seasons continuous dataset is taken in CSV design

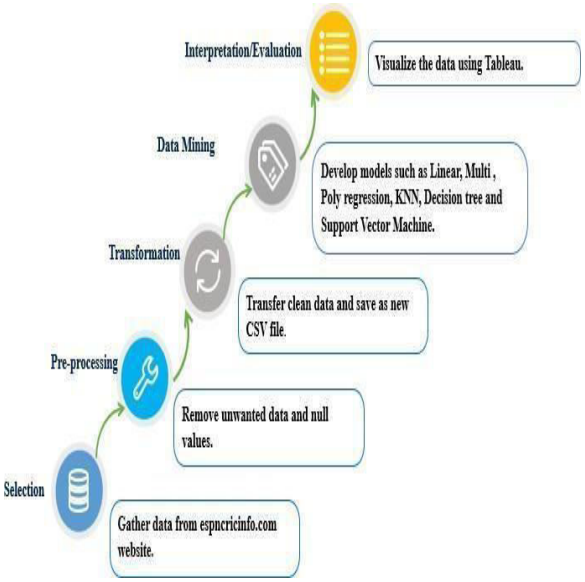

Fig: 2. Selection Methodology

VIII. DESIGNING THE DATASET

The main source of past matches data is the official website of Indian PremierLeague. The data is scrapped from this official site and stored in a Comma Separated Values (CSV).

Attribute	Description
Match_id	Distinct ID assigned to a Match (referred from Matches Table)
Innings	1 – First Innings, 2 – Second Innings
Batting Team	Team Batting currently
Bowling Team	Team Bowling currently
Over	Current Over of the Innings
Ball	Current Ball of the Over
Batsman	Name of the Batsman on Strike
Non-Striker	Name of the Batsman at the non-striker end
Bowler	Name of the Bowler for the Over in progress
Is_Super_Over	Indicates whether the over is super over or not
Wide Runs	Runs granted as wide
Bye Runs	Runs granted as bye
Leg Bye Runs	Runs granted as Leg-Bye
No Ball Runs	Runs granted for No Ball
Penalty Runs	Runs granted for Penalty
Batsman Runs	Runs scored by the Batsman
Extra Runs	Combined Extra Runs
Total Runs	Extra Runs + Runs scored by the Batsman
Player Dismissed	Name of the dismissed player (if any)
Dismissal Kind	Reason for the dismissal of the player(if any)
Fielder	Name of the Fielder responsible for dismissal (if any)

Table- 1: Attributes of Matches Table

In the information preprocessing stage, the information is fragmented, loud and conflicting.

VII. MODELS

a) Logistic Regression: Logistic backslide uses a condition because the depiction, broadly like rectilinear regression. Information respects (x) are joined straightly using burdens or coefficient regards (implied considering the way that the Athens letter Beta) to anticipate a yield regard (y).A key contrast from rectilinear relapse is that the yield esteem being demonstrated might be a parallel qualities (0 or 1) rather than a numeric worth.

b) Random Forest Classifier: Random Forest is an ensemble technique that is used to perform regression and classification tasks. Ensemble techniques combine results of various machine learning models and give the best accurate prediction of any individual model.

c) Gradient boosting Regression: Figures the distinction between the current expectation and subsequently the realized right objective worth. This distinction is called lingering.

Attribute	Description
Id	Distinct ID assigned to a Match
Season	Year in which the IPL Match was conducted
City	The city in which the match was held
Date	Date of the Match in dd/mm/yyyyformat
Team 1	First Contesting Team
Team 2	Second Contesting Team
Toss Winner	Winner of the Toss
Toss Decision	Decision taken on whether to bat orfield first by the winner of Toss
Result	Normal/Draw
Winner	Winning team for the Match
Win by Runs	Number of runs by which the team won
Win by Wickets	Number of wickets by which the team won
Player of theMatch	The best performer of the Match
Venue	Location of the Match
Umpire 1	First Umpire for the match
Umpire 2	Second Umpire for the match
Umpire 3	Third Umpire for the match

Table- 2: Attributes of Deliveries Table

IX. RESULTS

We used different machine learning algorithms to find the best combination that gives the most accuracy. We used five ML calculations: Decision Trees, Random Forest, Linear Regression, Gradient boosting and KNN in our investigations.

	id	city	date	player_of_match	venue	neutral_venue	team1	team2	toss_winner	toss_decision	winner	result	result_margin
476	829741	Ahmedabad	21-04-2015	SE Marsh	Sardar Patel Stadium, Motera	0	Rajasthan Royals	Kings XI Punjab	Kings XI Punjab	field	Kings XI Punjab	tie	NaN
535	980937	Rajkot	24-04-2016	V Kohli	Saurashtra Cricket Association Stadium	0	Gujarat Lions	Royal Challengers Bangalore	Royal Challengers Bangalore	bat	Gujarat Lions	wickets	6.0
717	1178397	Chandigarh	08-04-2019	KL Rahul	Punjab Cricket Association IS Bindra Stadium, ...	0	Kings XI Punjab	Sunrisers Hyderabad	Kings XI Punjab	field	Kings XI Punjab	wickets	6.0
639	1136564	Hyderabad	09-04-2018	S Dhawan	Rajiv Gandhi International Stadium, Uppal	0	Sunrisers Hyderabad	Rajasthan Royals	Sunrisers Hyderabad	field	Sunrisers Hyderabad	wickets	9.0
745	1178425	Chennai	01-05-2019	MS Dhoni	MA Chidambaram Stadium, Chepauk	0	Chennai Super Kings	Delhi Capitals	Delhi Capitals	field	Chennai Super Kings	runs	80.0

Fig 3: Loading the CSV files into data frame

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

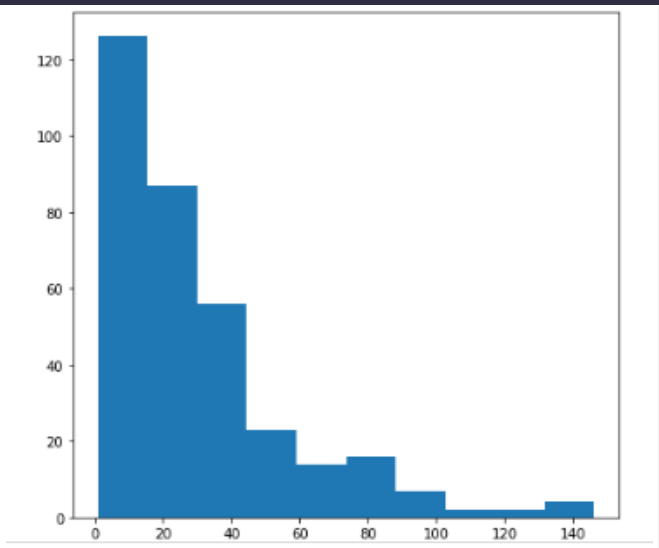

Fig 4: Distribution of Runs

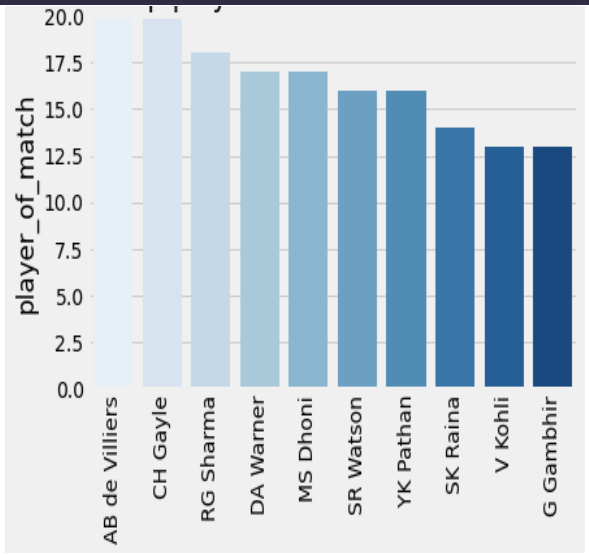

Fig 5:Top Players of the Match Winners

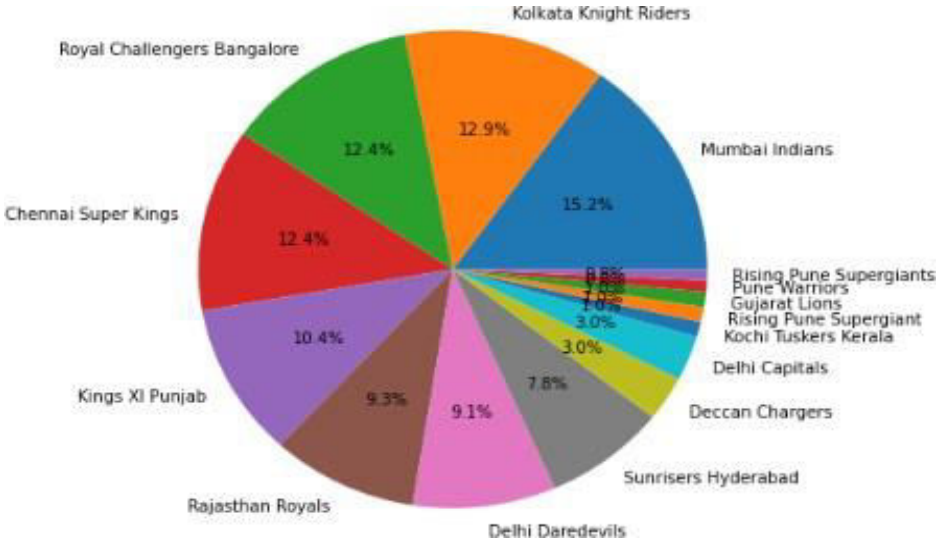

Fig 6: Pie chart of who won the Maximum IPL Matches

Random Forest Accuracy

The accuracy of the Random Forest classifier is 88.24 %, while that of SVM is 87.38%The accuracy of the Random Forest classifier is 88.24%, while that of SVM is 87.38 %, thus we chose Random Forest classifier as our preferred classifier. The dataset contains 20 percent testing set and 80 percent training set.

```
model = RandomForestClassifier(n_estimators=100)
target_var = ['winner']
outcome_var=['winner']
predictor_var = ['team1', 'team2', 'venue', 'toss_winner','city','toss_decision']
print_model_scores(model, match_df,predictor_var, target_var)
```

```
[ 3  5  7  1  2  6  7  1  6  9  6  5  4  9  5  1  3  9  6  5  3  2  1  6
  9  4  1  5  2  6  1  5  2  6  9  2  1  9  7  1  9  3  1  5  7  6  9  5
  9  1  5  4  2  6  5  6  5  6  1  3  7  4  5  2  4  7  6  3  4  7  9  4
  1  6  3  9  7  5  1  3  6  5  9  3  5  6  7  4  6  5  7  9  5  1  7  4
  3  1  7  3  6  9  5  4  9  7  2  3  2  5  7  3  4  3  4  2  1  7  2  4
  7  3  5  1  3  5  4  6  3  4  9  1  3  6  1  6  2  7  6  1  2  1  5  7
  2  3  5  1  9  7  6  1  6  2  4  9  4  3  9  6  4  1  5  6  5  4  1  2
  5  4  1  1  5  3  5  5  6  3  7 13  2  6  1  9 13  4  2 12  5  9  7  2
 12  4  1 12  9  1  3  9  1  6  5  3  5  4  2  6  3  7  2  6  5  1 12  2
  5  1 12  7  9  5  1  3 13  6 13  9  3  5  9  3  4  9 12  4  9  5  2  6
  4 15  3  1  5  1  3  5  1  2 13  6  7  5  2 13  1  3  7  1  5  9  2  4
 13  2  6  1  6  3  2  7  5  9  5 13  9  2  6 13  1  4  7  9  2  7  1  5
  4  7  9  1  5  2  6  1  3  2  6  9  1  5  3  1  5  6  9  1  5  7  1  9
  3  4  9  2  4  6  2  5  5  2  2  1 10  7  1 13 10  2  1  5  3 13 10  1
  5  2  6 13  9  7 10  6  5 10  5  3  7  9  5  3  9  1  5  2  6  1  5  7
  6  9 13 10  7  5  3  2 10  9  5  6  3  6  1  5  9  2  3  1 10  2  6  9
  1 13  5  1  9 10  9 13  3 10  5  6  1  1  2  3  9  6  3  7  9  5  9  5
  2 10  5  6  9  7  5  9  6 10  5  1  6  3  6  5  1  2  9 10  9 10  1  2
  6  1  5  7 10  2  7  3  2  1  9 10  2  9  3 10  1  9  3  2  9  1  2  5
  9  2  1  5  6  5  3  6  9  3  1  2  7  6  5  7  2  6  1  7  9 10  3  7
  3  1  5  3 10  2  6  5  7  1  3 10  9  6  5  2  1  9 10  1  2 10  3  5
 10  7  9  1  3  5  6  7  1  1  3  5  1 11  2  8  3  1  8  7  2  8  9  7
 10  2  1 10  3  7 10  8  2  9 11  7  1  8  7 10  9  1  2  7  2 11 10  3
  9 10  8  3 10  1  7  9  3  2  9  1  3 11  9  8  7 11  8  2  3  3 10 10
  3 10 11  2  9  3 10  1  9  7  1  2  1  8  2  9  1  3  2 10  3 10  1  8
  1 11  9  3 11  2  3  2  9 11  1  9 10  3 11  7 11  7  9 11  1  2  8 10
  2  7  9  7 10  1 11  3 11  2  1  1  1  9  2 10  5  6 10  3  1  2  6  9
  2  1  2  9  5  2  3  5  6  7  1  5 10  7  1  6  2  5  3  6  2  1  5 10
  1  9 10  6  1 10  6  2  3  5  1  3  2  1  3  5  6  2  7  5  5  6  2  5
  5  2 14  6  5  2  1 10  9 14 10  5  9  6  1 10  2  5 10 14  2  9  5  9
  5 14  1  3  5 14  1  9 10  1  2  6 14 10  5 14  5  3  6  1  6 14  1 10
 15  5  1  2 14  3  9  1  1 14  5  1  5 14  3 10  6  2  9  1 14  2 10  1
  2 14  5  6  1 14  9  6  2  5  3 14 10  9 10 14  9  1  3  2 10  5  1  6
  5  1  2  9 10  6  3  1  5  6  1 14  3  6 10 14  5  2  9  3  1 10 14  1]
Accuracy : 88.24%
Cross-Validation Score :[3.1834174  3.23705539 3.38709206 3.87377529 3.76209705]
Average RMSE: 3.4886874384949484
```

Algorithms	Accuracy
Logistic Regression	31.29 %
Random Forest Classifier	88.24 %
KNN	64.58 %
SVM	87.38 %

Table 3:Accuracy Comparison Table

Fig 8: Data Visualization

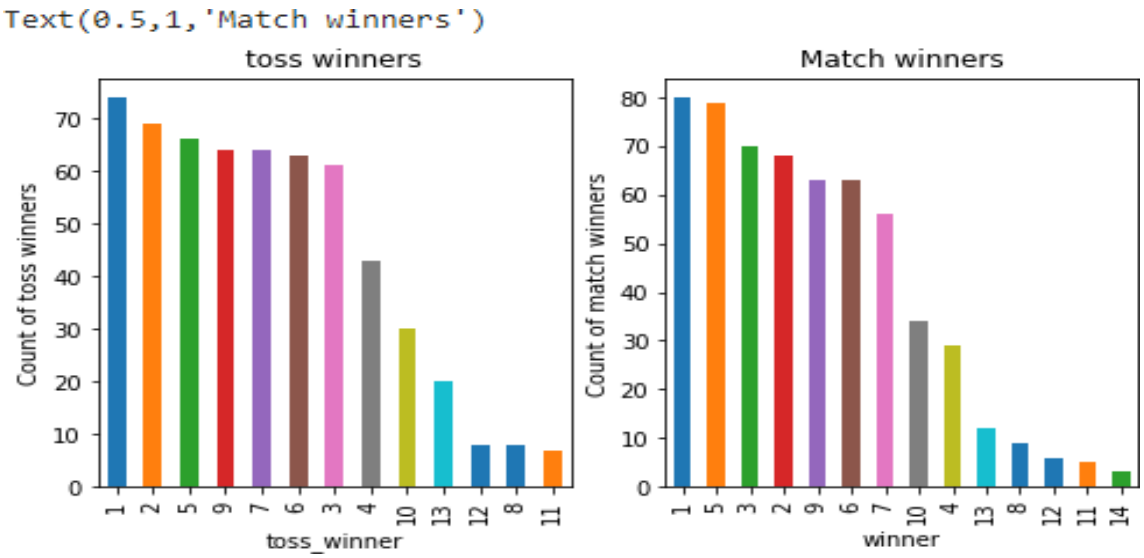

Fig9: Toss winners and count of toss winners

X. DEPLOY THE MODEL

For deployment we'll use the flask framework and heroku app platform. Further we likewise made an expectation website page for showing the outcome. When fixing a web webpage you'll generally have your live site, which is named the live environment or creation environment.

Here is link for my deployed project:http://ipl-match-score-prediction-ml.herokuapp.com/

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

RESULTS OF SCORE PREDICTION

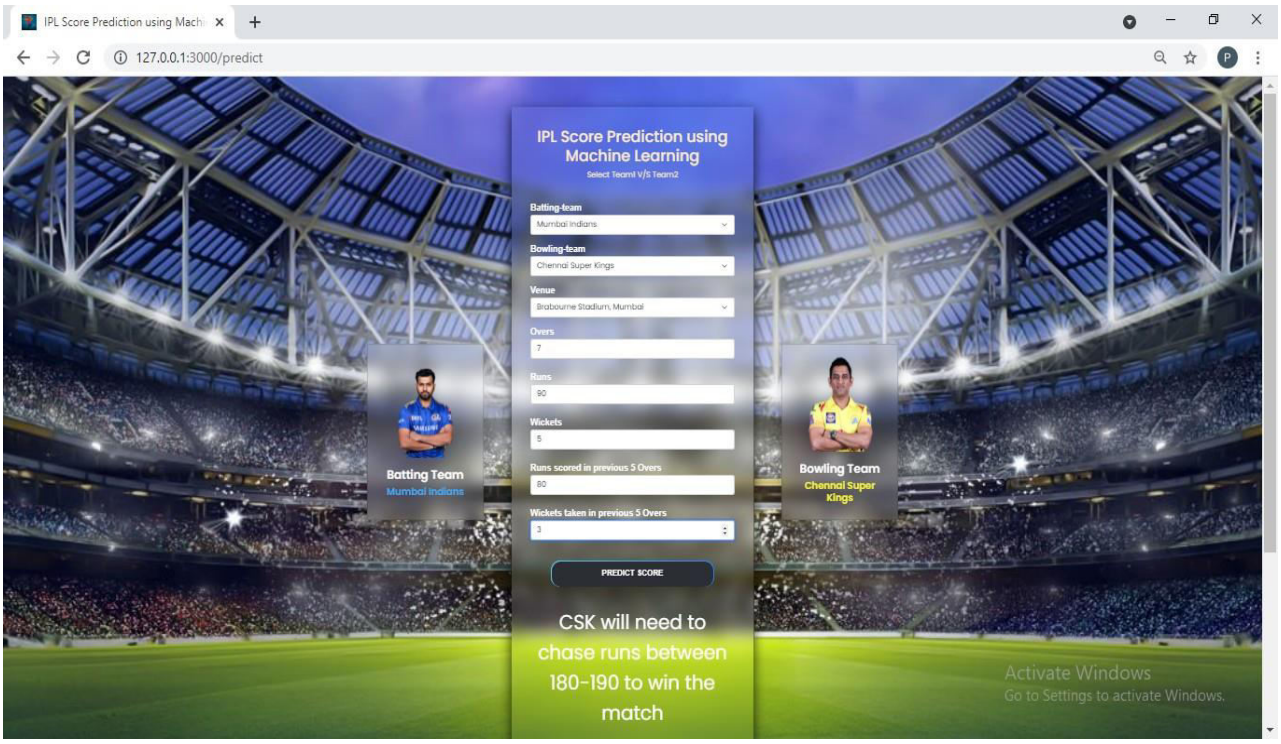

Fig 10: Selecting Two Teams Manually

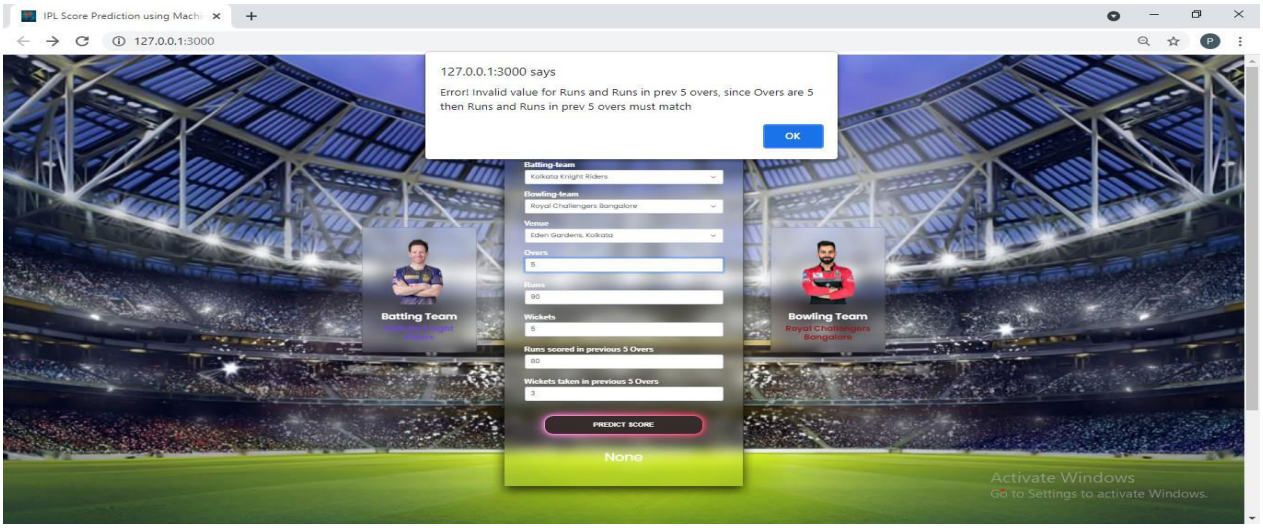

Fig 11: Giving Invalid values

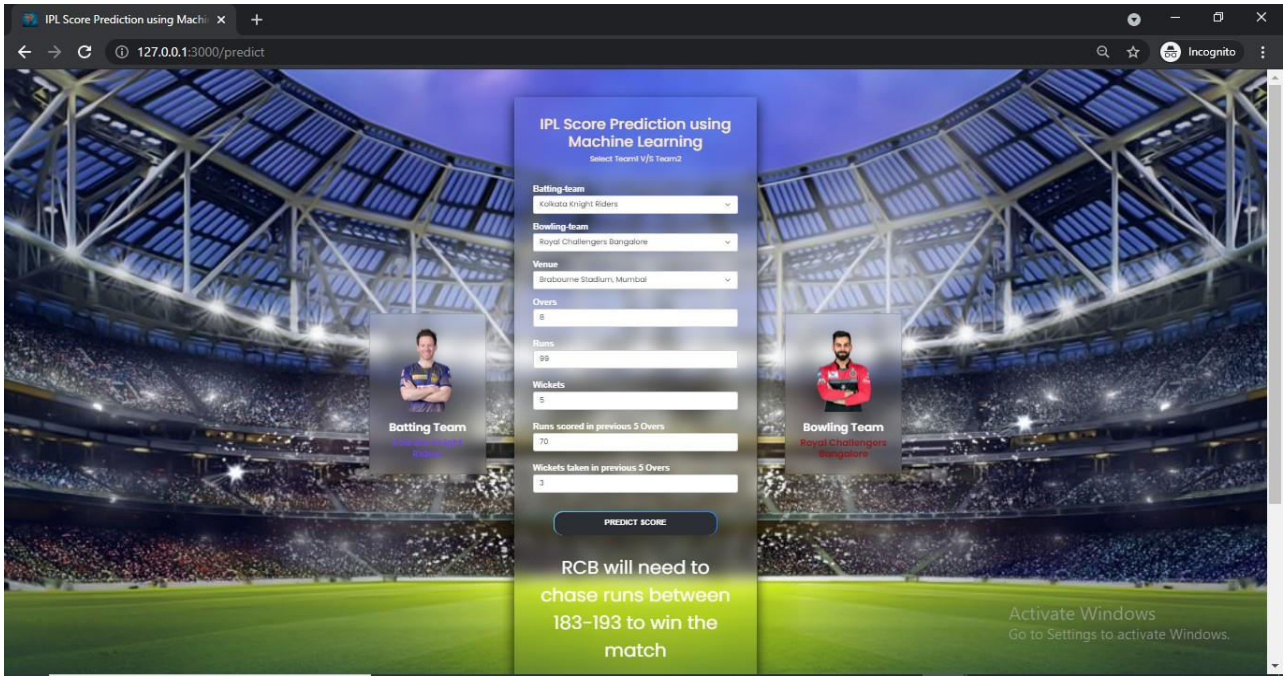

Fig 12: Selecting Another Two teams

XI. CONCLUSION

Selection of the simplest team for a match plays a big role for the team's victory. The principle objective of this project is to dissect the IPL cricket information and anticipate the players' performance. Here, three arrangement calculations are utilized and contrasted with track down the best precise calculation. This knowledge are going to be utilized in future to predict the winning teams for subsequent series of IPL matches. Hence using this prediction, the simplest team are often formed.

XII. REFERENCES

"Anticipating Cricket Score By victimisation Machine Learning ideas," H.V Ramachandra academician.,Jspm's, RSCOE. R.R. Kamble Assistant academician.,Jspm's, RSCOE. Worldwide Journal of Grid and Distributed Computing Vol. 13, No. 1, (2020), pp. 2396–2401.

http://sersc.org/journals/index.php/IJGDC/article/view/30487/16884

"The cricket victor forecast with use of AI and information examination," Vistro, Daniel, Leo Gertrude David Global Diary of Scientific and Technology Research, Volume 8, Issue 09, 2019.

https://www.kclas.ac.in/wp-content/uploads/2021/01/

The-Cricket-Winner-Prediction-With-Application-Of-Machine-Learning-And-Data-Analytics-.pdf

"Gauge of Live Cricket Score and Winning," Lokhande, Chawan, Pramila", Overall Journal of Trend in Research and Development, Volume 5(1), (2018).https://www.researchgate.net/publication/326416741_Prediction_of_Live_Cricket_Score_and_Winning