IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Cricket Score Predictor using Machine Learning

Om Channe[1], Pranjali Nagose[2], Nomaan Sheikh[3], Pranshu Bhalekar[4], Pranit Channe[5], Prof. Sachin Dhawas[6]
Students, Department of Computer Science[1,2,3,4,5]
Guide, Department of Computer Science[6]
Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, Maharashtra, India

Abstract: *In cricket, particularly the twenty20 format is most watched and loved by the people, where no one can guess who will win the match until the last ball of the last over. In India, The Indian Premier League (IPL) started in 2008 and now it is the most popular T20 league in the world. So we decided to develop a machine learning model for predicting the outcome of its matches. Winning in a Cricket Match depends on many key factors like a home ground advantage, past performances on that ground, records at the same venue, the overall experience of the players, record with a particular opposition, and the overall current form of the team and also the individual player. This paper briefs about the key factors that affect the result of the cricket match and the regression model that best fits this data and gives the best predictions.*
In these the score prediction includes linear regression, lasso regression and ridge regression whereas in winning prediction SVC classifier, decision tree classifier and random forest classifier are used. The model used the supervised machine learning algorithm to predict the winning. Random Forest Classifier used for good accuracy and the stable accuracy so that desired predicted output is accurate.

Keywords: cricket, Indian premier league, prediction of match, supervised machine learning, Linear Regression, Ridge Regression, Naive Bayes, Random Forest Classifier, IPL Winning Prediction, IPL Score Prediction..

I. INTRODUCTION

After football, cricket is the game that is most adored. Beginning in the fifteenth century, the sport is played in England. With the number of fans growing globally, cricket may soon surpass football as the sport with the largest fan base. It is no longer simply a game in India today; it has become a religion. There are three main formats. 50 overs in a One Day International match are played in a single day. The second format is the test format, which was used in earlier iterations of the game. It is played over five days and includes of two innings from each team, with each innings containing 80 to 90 overs. For five days, the team must deliver regularly. In this extremely difficult game type, a player's endurance, strength, patience, and mental toughness are crucial.
As IPL viewers, people create their own predictions when watching a certain match. They base these predictions on the facts they have available and use various statistics and records to determine who will win. As a result, there is a significant market for algorithms that forecast the best score and winning team, which is more crucial. We shall make predictions for every IPL match that has already been played. Machine learning techniques are used in this process to anticipate the outcomes of the matches.

II. DATASET FEATURES

The approach over here we are using is ML based. So the basic requirements of an ML algorithm is dataset, training of that dataset using the algorithm and testing the model. So, we have imported dataset from Kaggle. Later on calculating the accuracy and improving the accuracy by using Random Forest classifier for winning prediction and Linear Regression for score prediction.

- Score Prediction:- For conducting our research, we collected data on all the IPL matches played in 2008. The dataset consists of 76015 numbers of rows. Dataset consists 15 columns over which we applied feature selection techniques and selected 8 features in which 7 are input feature and 1 is our target variable. The attributes selected were bat team,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

- bowl team, overs, runs, wickets, runs in previous 5, wickets in previous 5 for score prediction.
- Winner Prediction:- For conducting our research, we collected data on all the IPL matches played since 2008 till 2019. The dataset consists of 757 numbers of rows. Dataset consists 17 columns over which we applied feature selection techniques and selected 5 features in which 4 are input feature and 1 is our target variable. The attributes selected were team1, team2, winner, toss decision, toss winner for winning prediction. Each team was analyzed individually against every other team.

Table 1 : Dataset Attributes and their values

Attributes	Values
Batting Team	Batting Team Name among 8 teams in IPL
Bowling Team	Bowling Team Name among 8 teams in IPL
Overs	Value > 5 Over
Runs	0-300
Wickets	0-10
Run Scored in last 5 overs	0-300
Run Scored in last 5 overs	0-10
Total Runs	0-300

Table 2: Dataset Attributes and their values

Attributes	Values
Team1	Team1 Name among 8 teams in IPL
Team2	Team2 Name among 8 teams in IPL

PREPARING DATASET

Problem with the dataset

The Indian Premier League have completed only 12 seasons because of which only 580 matches data is only available for prediction purpose which is less compared to one-day international matches and test matches. Because of issues with some team franchises in some season, few teams removed and some teams added. Also, some teams change their names, and some teams get discontinued. If we remove the match data of the teams that are not active now then we will miss some important data of the teams which are still active.

Training Dataset

Till now only 12 seasons have happened of Indian Premier League, so we are taking the first 11 seasons dataset as a training dataset which consists of 580 matches.

Testing Dataset

Till now only 12 seasons have happened of Indian Premier League, so we are taking the last season i.e. 12th season as a testing dataset which consists of 60 matches.

Organizing Dataset

The machine learning model takes data only in numbers format so we have to convert the team names and other features into numbers as shown in the below figure.

MODULE 1

A. Features Selected

1. Team Name 1

This feature contains the 1st playing team name this will be converted to a specific number given to that team.

2. Team Name 2

This feature contains the 2nd playing team name this will be converted to a specific number given to that team.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

3. Venue

If a team playing at their home ground then that team has a better chance to win as they are known to pitch condition,c limate and also crowd support plays an important role to boost the confidence of players.

B. Machine Learning model used

1. Multiple Linear Regression

Multiple linear regression uses multiple variables to predict a label.

MODULE 2

A. Features Selected

1. Team 1 Batting Average

The player's batting average is calculated as dividing the total number of runs scored by the total number of times the batsman gets out. The Team 1 Batting Average is calculated by adding a batting average of all the players that are in playing XI and dividing it by 11.

2. Team 2 Batting Average

The player's batting average is calculated as dividing the total number of runs scored by the total number of times the batsman gets out. The Team 2 Batting Average is calculated by adding a batting average of all the players that are in playing XI and dividing it by 11.

Team1 Bowling Average

The bowler's bowling average is calculated as dividing the total number of runs conceded by the total number of wickets taken. The Team 1 Bowling Average is calculated by adding the bowling average of all the players that are in playing XI and dividing it by 11.

Team2 Bowling Average

The bowler's bowling average is calculated as dividing the total number of runs conceded by the total number of wickets taken. The Team 2 Bowling Average is calculated by adding the bowling average of all the players that are in playing XI and dividing it by 11.

Team1 Economy

The bowler's bowling economy is calculated as dividing the total number of runs conceded by the total number of overs bowled. The Team 1 Bowling Economy is calculated by adding the bowling economy of all the players that are in playing XI and dividing it by 11.

Team2 Economy

The bowler's bowling economy is calculated as dividing the total number of runs conceded by the total number of overs bowled. The Team 2 Bowling Economy is calculated by adding the bowling economy of all the players that are in playing XI and dividing it by 11.l

B. Machine Learning model used

1. Decision Tree Regression

Decision Tree Regression is used to predict a class by learning the decision rule from the features used to train the model and it is a supervised machine learning algorithm.

2. Random Forest Regression

Random Forest Regression is capable of both classification and regression tasks by using multiple decision trees and using a technique of bagging.

3. Naive Bayes

Naive Bayes classifier is a probabilistic machine learning model that is used for classification of the testing data given. Naive Bayes finds the probability of events and based on this it classifies the label.

4. Logistic Regression

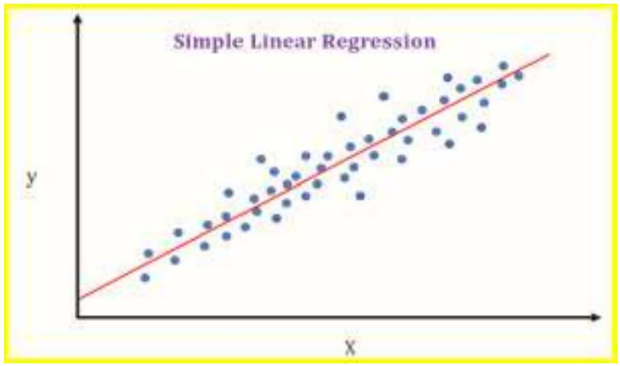
Logistic regression is used when the label contains only two values like happy or unhappy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

III. ALGORITHMS

We tried to use two machine learning techniques: regression and classification. Selected algorithms from each technique were trained then.

- Regression: Regression analysis uses various algorithm for the computation and based on that it predicts the continuous value. There are certain set of variables are used for the input and the continuous range value is the target variable. Based on the application different regression algorithms are used. There are different regression techniques. Out of which the linear, ridge and lasso algorithms are used for predicting the score.
- Linear regression: To predict the continuous values, Linear regression is used. Certain known parameters are given to the machine learning algorithms, it predicts the continuous values as output. It cannot used for the classification problems. The proposed model predicts the score using the Linear Regression.



IV. IMPLEMENTATION OF THE GUI

The Graphical User Interface is developed for the machine learning models using the Flask Framework. For the backend of the site Python is used. The site can be used to predict the IPL match score with the help of last 5 overs of the data.

All the input information necessary for the model for the prediction is provided to the model. The calculation is not stored in the system because all calculations computed at real time. We implemented it that way as we can add change more attribute to the system with minor changes to the program.

Score Prediction

The GUI required at least 5 overs of the data to predict the score as shown in the Fig 1. Model require the input data of Batting team, Bowling team, Over, Runs, Wickets, Run Scored in last 5 overs, Wickets fall in last 5 overs to predict the score of the match as shown in fig 2.

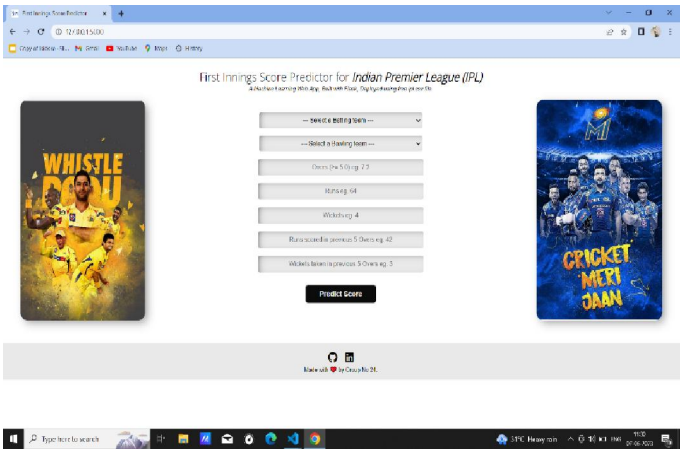


Fig.1- Score Prediction Model UI

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

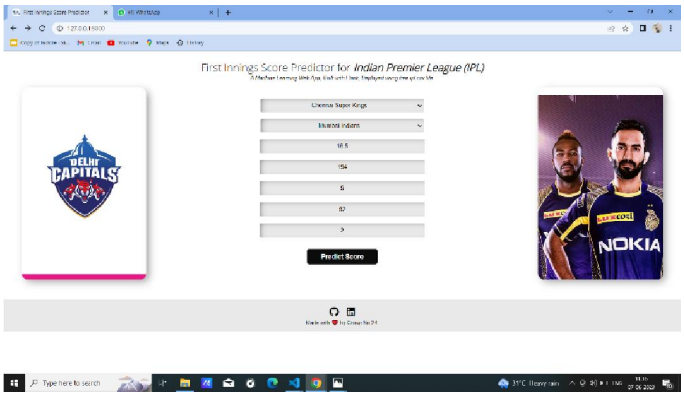


Fig.2- Input to Score Prediction Model

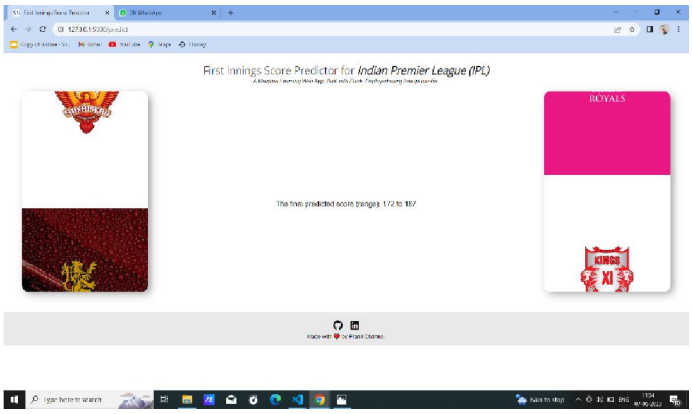


Fig.3- Score Prediction result

V. CONCLUSION

This paper will give the important information regarding IPL score prediction and winning prediction system, that which parameters are required also the classifiers and algorithms. it helps in mathematical operation. Using all the information we have developed a website. for that the important work we have to do for the model is comparative analysis of machine learning techniques that is for score prediction the regressions and for winning prediction the analysis of classifiers. In Score Prediction analysis accuracy of Linear Regression is more than Ridge and Lasso Regression and in winning prediction analysis among SVC, Decision tree classifier and Random forest classifier, we got Random forest classifier accuracy more than other 2, with all 90%, 80%, 75%, 70% training data.

ACKNOWLEDGEMENTS

We would like to thank Prof. Sachin Dhawas , professor, Rajiv Gandhi College of Engineering, Research & Technology, Department Of Computer Science & Engineering, Chandrapur for motivating and guiding us for this paper presentation.

REFERENCES

[1]. Pallavi Tekade, Kunal Markad, Aniket Amage, Bhagwat Natekar,"CRICKET MATCH OUTCOME PREDICTION USING MACHINE LEARNING"international journal of advance scientific research and engineering trends.

[2]. Preetham HK, Prajwal R, Prince Kumar, Naveen Kumar, "Cricket Score Prediction Using Machine Learning" Dayananda Sagar Academy of Technology and Management, Bangalore, India

[3]. Rohit Khade, Nikhil Bankar, Prashant Khedkar and Prof. Prashant Ahire (2019). "Cricket ScorePrediction using Machine Learning Algorithms"

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

[4]. Arjun Singhvi, AshishShenoy, Shruthi Racha and Srinivas Tunuguntla."Prediction of the outcome of a Twenty-20Cricket Match." (2015).

[5]. Swetha, Saravanan.KN,"AnalysisonAttributes Deciding Cricket Winning", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03 | (March 2017).