IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 7.301

Volume 3, Issue 1, April 2023

Cricket Match Prediction using Machine Learning

Ms. Shrunkhala Wankhede[1], Vaishnavi Desai[2], Vaibhavi Desai[3], Disha Tangade[4]

Assistant Professor, Department of Computer Science and Engineering[1]
Student, Department of Computer Science and Engineering[2,3,4]
Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, India

Abstract: *As cricket is the most popular game in the world. T20 and ODI are most loved by people. The IPL was launched in 2008. So we decided to develop a machine learning model that can predict both game scores and outcomes. In this paper, a model with three methods is proposed. The first is the score prediction, the second is his IPL match win percentage and the last is his ODI match win percentage. The model used supervised machine learning. We used previous statistics to build this model. The application is trained on historical data from previous games and uses these algorithms to make accurate predictions for upcoming games. Finally, several predictors are identified that can be used for data analysis. I used jypyter notebook. A label encoder was used for pre-processing. I created a predictive model using a machine learning algorithm. Algorithms used are decision tree classifier, random forest classifier, etc. for team winning prediction, Lasso regression, and ridge regression for score prediction. The web application is designed to provide real-time predictions for upcoming matches, allowing cricket enthusiasts to stay up to date with the latest predictions for upcoming matches. Overall, this project has the potential to revolutionize the world of cricket by providing accurate match predictions. Our web application enables cricket enthusiasts to make informed decisions and improve their overall viewing experience.*

Keywords: Decision Tree Classifier, Random Forest Classifier, Lasso Regression, Logistic Regression, Gaussian NB, Gradient Boosting Classifier

I. INTRODUCTION

Cricket is a sport that is widely played and watched across the world, and two of the most popular formats of cricket are IPL (Indian Premier League) and ODI (One Day International). These formats attract millions of fans globally, and with the growing popularity of machine learning and data science, it has become possible to predict the outcome of cricket matches using advanced statistical models. In this project, we aim to build a web application that uses machine learning algorithms to predict the results of IPL and ODI matches. Our web application will provide real-time predictions for upcoming matches based on various factors such as team composition, player statistics, pitch conditions, and weather. To build our web application, we will be using Python, a popular programming language for data science and machine learning. Python provides a wide range of libraries and tools for data analysis and machine learning, making it an ideal choice for building predictive models. We will be using popular machine learning algorithms such as regression analysis, decision trees, and support vector machines to predict the outcome of matches. Our application will be trained using historical data from previous matches and will use these algorithms to make accurate predictions for upcoming matches. One of the challenges of predicting cricket matches is the number of variables that can affect the outcome of a match. Factors such as the form of the players, the pitch conditions, and the weather can all have a significant impact on the outcome of a match. To address this challenge, we will be using a combination of statistical analysis and machine learning algorithms to build our predictive models. Our models will take into account multiple factors and will use a combination of historical data and real-time information to make accurate predictions.

II. LITERATURE SURVEY

We studied papers based on the area of Cricket prediction. Following papers are related to our area i.e. Cricket prediction.

1] Monoj Ishi explained "WINNER PREDICTION IN ONE DAY INTERNATIONAL CRICKET MATCHES USING MACHINE LEARNING". In this paper explained the focus is given to the prediction of victory in one day international game of cricket using machine learning. The concept of ensemble algorithm using voting and stacking classifier is used

DOI: 10.48175/568

for prediction with machine learning algorithms. 2] I.P. Wickramasinghe (2015) explained "Guessing the performance of batsmen in test cricket," in this paper explained how the performance of batsmen in a series of tests can be predicted using long-distance and long-range methods. In this paper they collect sample data from test cricket batsmen who played during the 2006-2010 season in nine overseas teams and show that these sample data shows long and high-level formations of three levels. 3] Jhanwar and Paudi explained about "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," It predicts the outcome of a cricket match by comparing the strengths of the two teams. For this, they measured the performances of individual players of each team. They developed algorithms to model the performances of batsmen and bowlers where they determine the potential of a player by examining his career performance and then his recent performances.4] The paper "Predicting Outcome of ODI Cricket Games" predict the outcome of One Day International games in cricket by proposing a numerical model. This paper is published by Kevin Desai, Siddhant Doshi, and Surekha Dholay. The conclusion of their project is in the form of win-loss percentage for both the participating teams at a given venue. 5] In the paper "Cricket Score Prediction Using Machine Learning" utilising machine learning algorithms such as SVM, Random Forest Classifier (RFC), Logistic Regression, and they have suggested a model for predicting the results of the IPL matches.6] The paper "Sport analytics for cricket game results using machine learning" published by Kumash Kapadia , Hussein Abdel-Jaber investigates machine learning technology to deal with the problem of predicting cricket match results based on historical match data of the IPL. Influential features of the dataset have been identified using filter-based methods including Correlation-based Feature Selection, Information Gain (IG), ReliefF and Wrapper.7] Prasad Thorat, Vighnesh Buddhivant and Yash Sahane predicted the cricket score. So, to get an accurate score prediction they should have a system that can predict the first innings score more effectively. Lots of people like watching cricket and they also like to predict the final score. This research paper focuses on an accurate prediction of cricket scores for live IPL matches considering the previous dataset available and also considers the various factors that play an important role in the score prediction.

III. DATASET FEATURE

The approach used here is based on ML. Therefore, the basic requirements of any ML algorithm are a dataset, training on that dataset using the algorithm, and testing the model. So we imported some of the records from Kaggle. After that, we compute the accuracy and use the random forest classifier for win probability prediction to improve the accuracy. Use linear regression for score prediction.

3.1 IPL Score Prediction

Dataset which we have used for this is ipl.csv. Dataset contains 76015 rows and 15 columns. This dataset has data from 2008 to 2017. We have selected 8 features from the dataset. Out of the eight features, one is target value and others are used to predict the target value.

Features are as follows

1. Batting team
2. Bowling team
3. Over (Greater than 5)
4. Runs
5. Wickets
6. Runs scored in previous 5 overs
7. Wickets taken in previous 5 overs
8. Total runs

3.2 IPL Winner Prediction

Datasets used for the winner prediction are matches.csv and deleveries.csv. To predict the winner of IPL match, we have used the dataset from Kaggle. First dataset contains 257 rows and 18 columns. Second dataset contains 179079 rows and 21 columns. These two dataset has been combined to form a one final dataset. From the dataset, we have selected some features which help to give better prediction. We have selected 8 features from the dataset. Out of the eight features, one is target value and others are used to predict the target value. Some attributes have been created using given attribute such as current run rate, required run rate, runs left, balls left.

Features are as follows

1. Batting team
2. Bowling team
3. City
4. Total runs
5. Wickets
6. Overs finished
7. Ball
8. Result

3.3 ODI Winner Prediction

Dataset which we have used is .csv. It contains 128212 rows and 16 columns. Some features have been selected which helps to give better prediction. We have also removed some columns from the dataset which are redundant. For the final prediction, we have selected 10 attributes. Among 10 attribute, one is target value and others are used to predict the outcome.

Features are as follows

1. Batting team
2. Bowling team
3. Batsman
4. Bowler
5. Current runs
6. Wickets
7. Overs finished
8. Striker runs
9. Non- striker runs
10. Result

IV. FEATURE SELECTION

Feature Selection is a process in which we select an optimal set of features from the input features set by using feature selection techniques. By removing redundant features, we reduce the dimension of data and we can improve the time and space complexity of data. Feature selection improves the performance of the model and saves time and space.

V. BLOCK DIAGRAM AND METHODOLOGY

The first step is to import the cricket dataset. After importing dataset, data has been pre-processed. We converted categorical features using one Hot Encoding method. We further pre-process data by checking null value and replace it with the mean, median values of respective column. As the dataset has some redundant columns, so we decided to do feature selection. Features which we have selected will be helpful to give better prediction. After that, Visualization of data has been done. So that we can give better prediction. Then, we have splitted data into train set and test set. For training set, 80% of data has been taken. For testing set, 20% of data from dataset has been taken. Then, prediction model is created. Machine Learning algorithms are applied. Then, we check accuracy of algorithms by using taking the ratio of the predicted testing data and the actual testing data. After that, we decided to go for one algorithm which has good accuracy.

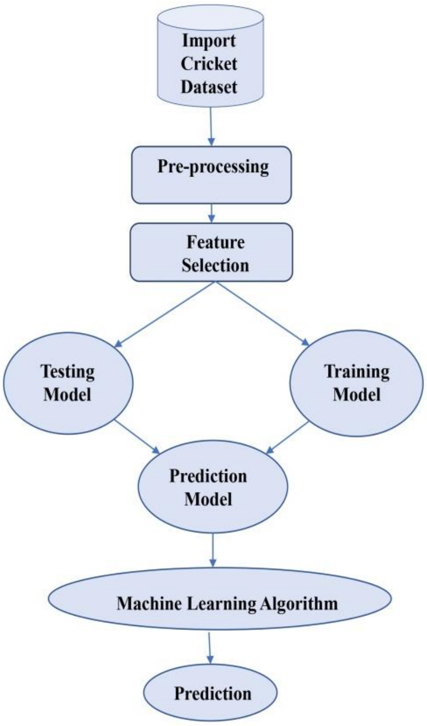


Fig. 1.Block Diagram and Methodology

VI. ALGORITHM

6.1 IPL Score Prediction

1. **Regression**: Regression is a method for understanding the relationship between independent variables or features and a dependent variable or outcome. Outcomes can then be predicted once the relationship between independent and dependent variables has been estimated. Regression is a field of study in statistics which forms a key part of forecast models in machine learning. It's used as an approach to predict continuous outcomes in predictive modelling, so has utility in forecasting and predicting outcomes from data. Machine learning regression generally involves plotting a line of best fit through the data points. The distance between each point and the line is minimised to achieve the best fit line.
2. **Ridge Regression:** Ridge regression is a method of estimating the coefficients of multiple regression models in scenarios where the independent variables are highly correlated.This method performs L2 regularization. Multicollinearity problems cause the least squares method to become unbiased, have high variance, and have predicted values that deviate significantly from the actual values.
3. **Lasso Regression:-**LASSO regression is a regularization technique. This is used via regression strategies for accurate predictions. This model uses shrinkage. Shrink shrinks the data values to a center point as the mean. The lasso method promotes simple, straight and thin models (that is, models with few parameters). This particular style of regression is suitable for models with a high level of multiple regressions and for automating certain aspects of the model like select variables/remove parameters.
4. **Linear Regression:-**Linear regression is used to predict continuous values. Certain known parameters are given to a machine learning algorithm to predict continuous values as output. The proposed model uses linear regression to predict scores.

6.2 IPL Win Prediction

1. **Logistic Regression:** The basic structure of logistic regression is: Get a dataset of N points. Each point i is a set of m input variables x1,i ... xm,i (also called independent variables, explanatory variables, predictor variables, characteristics or attributes) and a binary outcome variable Yi (also called dependent variable).).

Variable, response variable, output variable or class), i.e. can have only two possible values 0 (often 'no' or 'error') or 1 (often 'yes' or 'success') can. The goal of logistic regression is to use a data set to create a predictive model of the outcome variable.

2. **Random Forest:** Random Forest (Random Decision Forest) is an ensemble learning method for classification, regression, and other tasks that works by building different decision trees during training. For classification tasks, the output of a random forest is the classes chosen by most trees. For regression tasks, the mean or mean prediction for each tree is returned.

6.3 ODI Match Prediction

1. **Decision Tree Classifier:-**Decision tree learning is a supervised learning approach used in statistics, data mining, and machine learning. This format uses a classification or regression decision tree as a predictive model to draw conclusions about a set of observations.
2. **Logistic Regression:-**The basic structure of logistic regression is: Get a dataset of N points. Each point i is a set of m input variables x1,i ... xm,i (also called independent variables, explanatory variables, predictor variables, characteristics or attributes) and a binary outcome variable Yi (also called dependent variable).). Variable, response variable, output variable or class), i.e. can have only two possible values 0 (often 'no' or 'error') or 1 (often 'yes' or 'success') can. The goal of logistic regression is to use a data set to create a predictive model of the outcome variable.
3. **Gaussian NB:-**Gaussian Naive Bayes is an extension of Naive Bayes. Other functions are used to estimate the data distribution, but Gaussian or normal distributions are the easiest to implement, as they require computing the mean and standard deviation of the training data.
4. **Gradient Boosting Classifier:-**Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It produces predictive models in the form of an ensemble of weak predictive models (usually decision trees). If the decision tree is a weak learner, the resulting algorithm is called a gradient boosted tree generally better than random forest.

6.4 IPL Score Prediction Algorithm Analysis

Linear regression was found to be more accurate in predicting scores compared to ridge and LASSO regressions.

Table 1. Accuracy of the Score Prediction Models (IPL)

Algorithm	MAE	MSE	RMSE
Linear Regression	12.118699658080088	251.04770095081787	15.844484874896308
Ridge Regression	12.117294527004992	251.03172964112588	15.84398086470461
Lasso Regression	12.213583996827493	262.36538279606964	16.19769683615759

Linear regression was found to be more accurate in predicting scores compared to ridge and LASSO regressions.
y = Co + C1 * x (Linear regression equation)
Where y is the dependent variable, x is the independent variable,
C0 is the bias coefficient, and C1 is the coefficient of x.

6.5 IPL Win Prediction Algorithm Analysis

Logistic regression was found to be more accurate in predicting win probability compared to Random Forest Classifier. Although, Random Forest Classifier gives more accuracy, but we want the value in the form of percentage. So, we decided to go for Logistic Regression.

Table 2. Accuracy of IPL Win prediction

Algorithms	Accuracy
Logistic Regression	80.31
Random Forest Classifier	99.92

6.6 ODI Win Prediction Algorithm

As the Gradient Boosting Classifier has was found to be more accurate in predicting win probability compared to Decision Tree Classifier, Logistic Regression, and Gaussian Classifier.

Table 3. Accuracy of ODI Win prediction

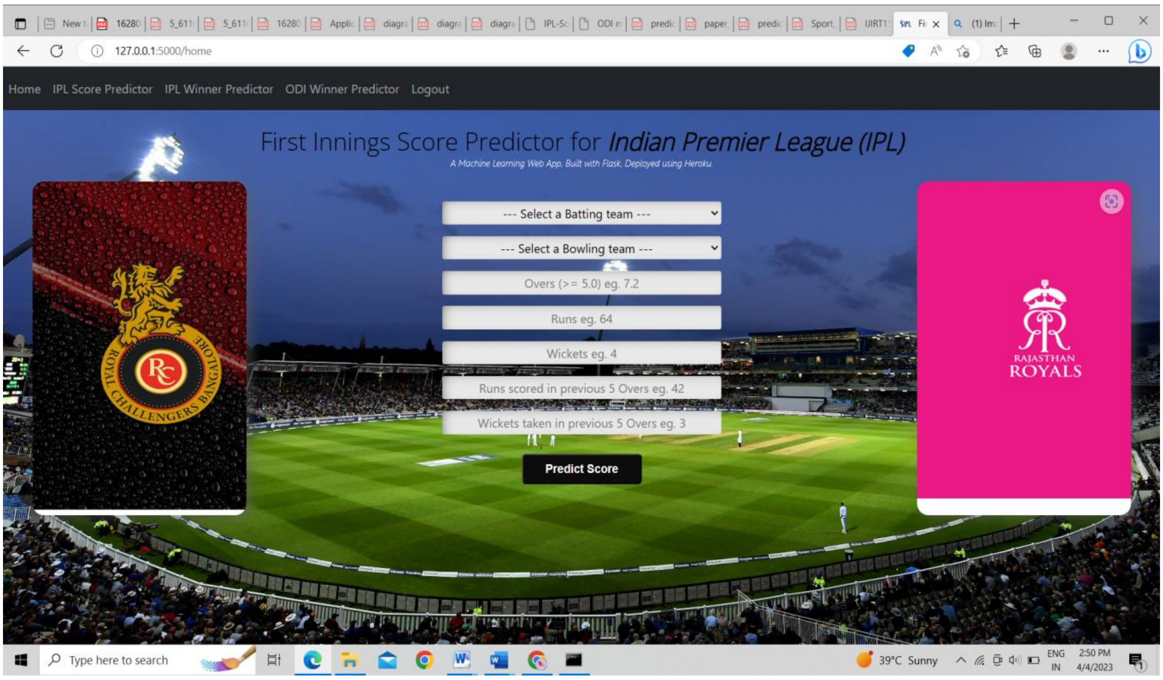
Algorithms	Accuracy on training data	Accuracy on testing data
Decision Tree Classifier	77.10	77.30
Logistic Regression	74.7	74.60
Gaussian Classifier	76.6	76.8
Gradient Boosting Classifier	100	99.9

VII. IMPLEMENTATION OF GUI

A graphical user interface is developed for machine learning models using the Flask Framework. The backend for pages uses Python. This site can be used to predict his IPL score using the last 5 over data. The model is provided with all the input information it needs to make a prediction. All calculations are performed in real time.

7.1 IPL Score Prediction: The GUI required at least 5 overs of the data to predict the score as shown in the Fig. 2.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 the Fig. 2

Score prediction model



127.0.0.1:5000/home

Home IPL Score Predictor IPL Winner Predictor ODI Winner Predictor Logout

First Innings Score Predictor for *Indian Premier League (IPL)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku.

--- Select a Batting team ---

--- Select a Bowling team ---

Overs (>= 5.0) eg. 7.2

Runs eg. 64

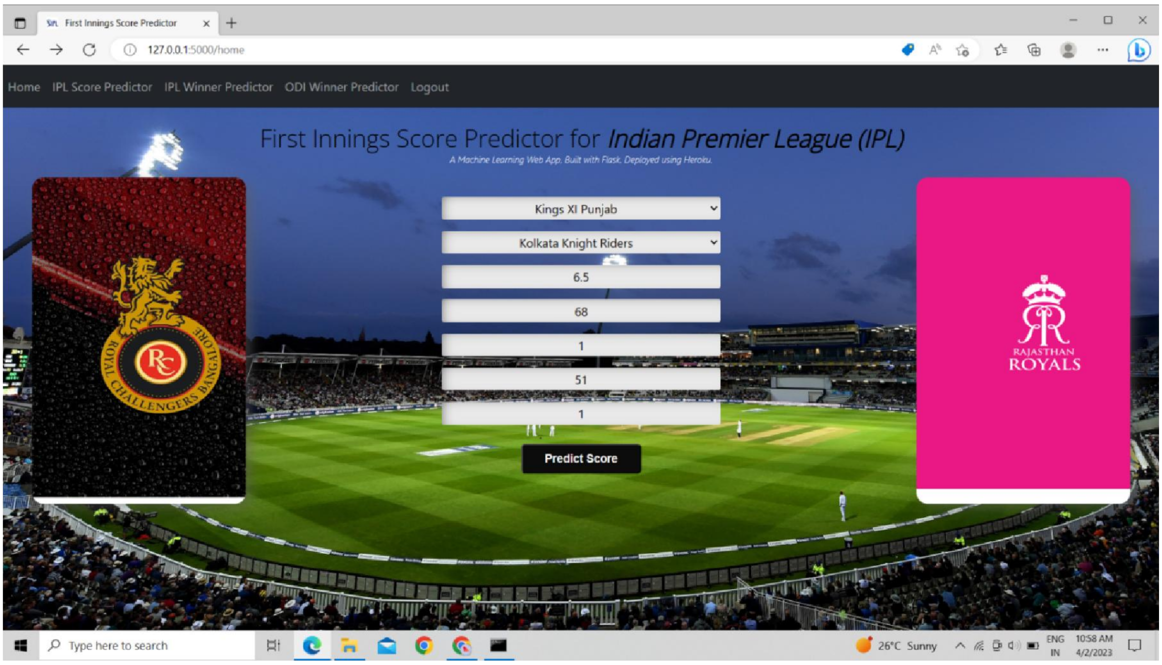
Wickets eg. 4

Runs scored in previous 5 Overs eg. 42

Wickets taken in previous 5 Overs eg. 3

Predict Score

Fig.2.Model of Score Prediction(IPL)



First Innings Score Predictor for *Indian Premier League (IPL)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku.

Kings XI Punjab

Kolkata Knight Riders

6.5

68

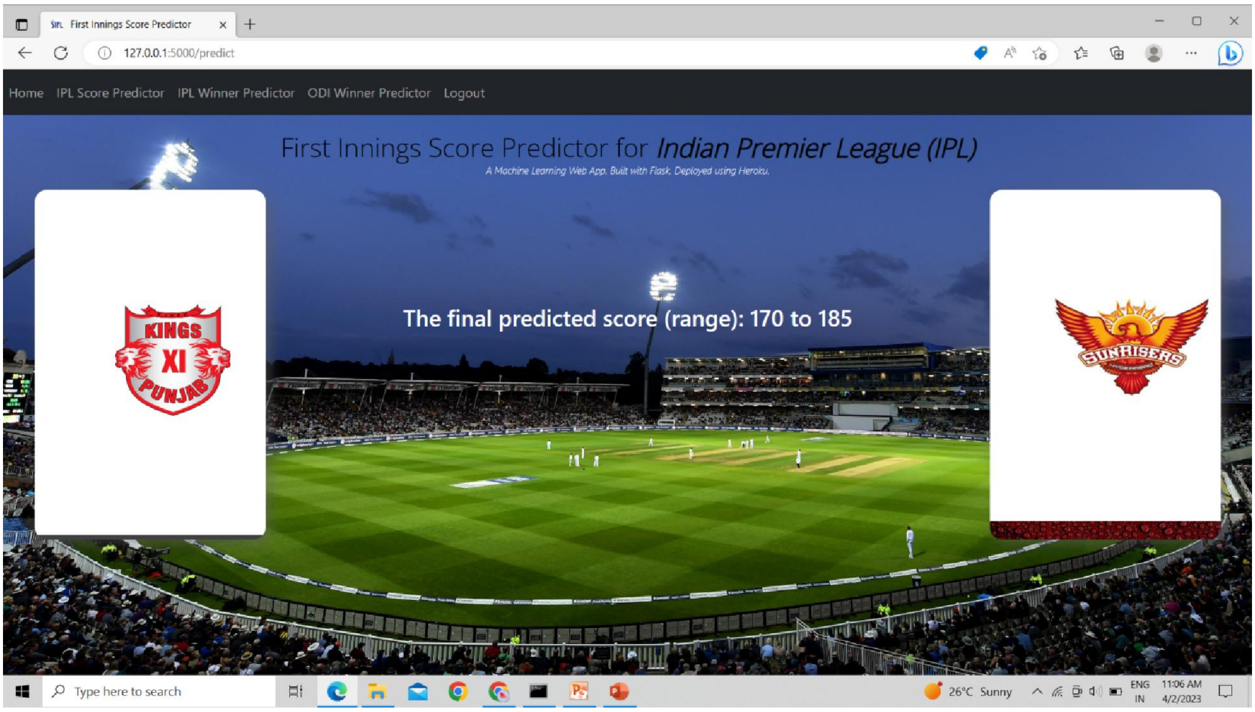
1

51

1

Fig. 3. Input to IPL Score prediction model

The output we get from the model is not exactly the predicted output. Therefore, to improve the accuracy of our model, we add and subtract 3 to get the score range shown in Fig 4. So our model works in most cases



First Innings Score Predictor for *Indian Premier League (IPL)*

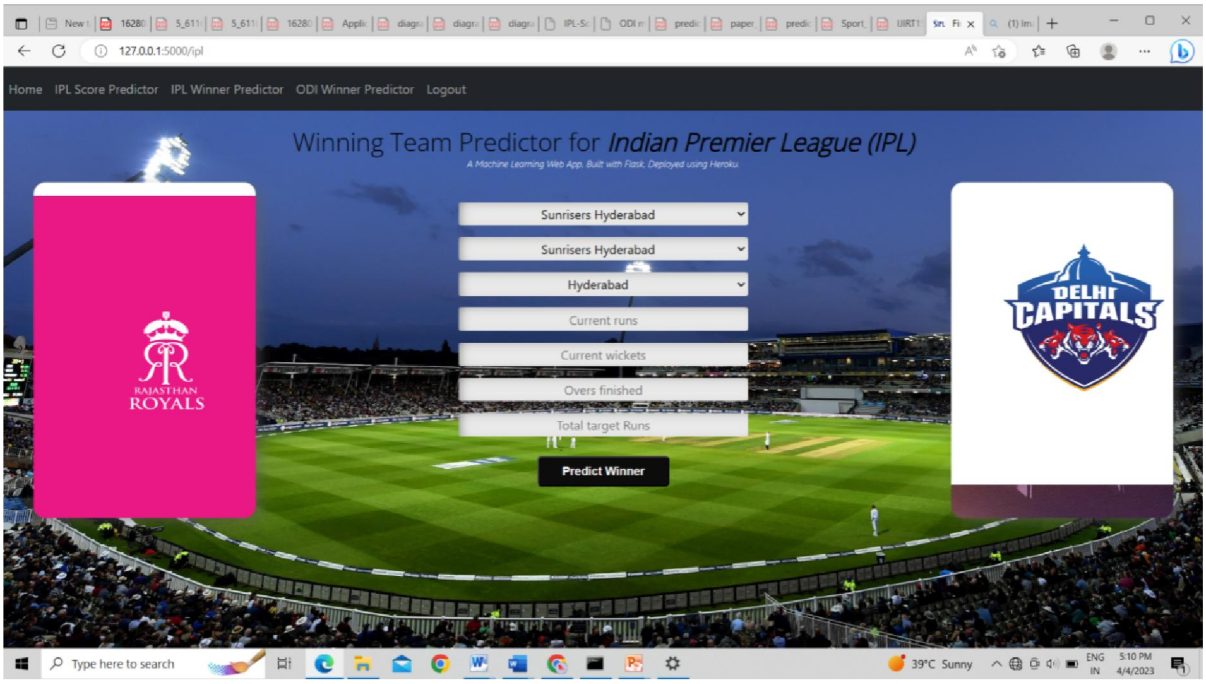
A Machine Learning Web App, Built with Flask, Deployed using Heroku.

The final predicted score (range): 170 to 185

Fig.4.IPL Score prediction result

7.2 IPL Win Prediction

The GUI required team1, team2, city , total runs, wickets, overs finished to predict the Winner of the IPL match as shown in the Fig 5. Model requires the input data as shown in the Fig 6.



Winning Team Predictor for *Indian Premier League (IPL)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku

Sunrisers Hyderabad

Sunrisers Hyderabad

Hyderabad

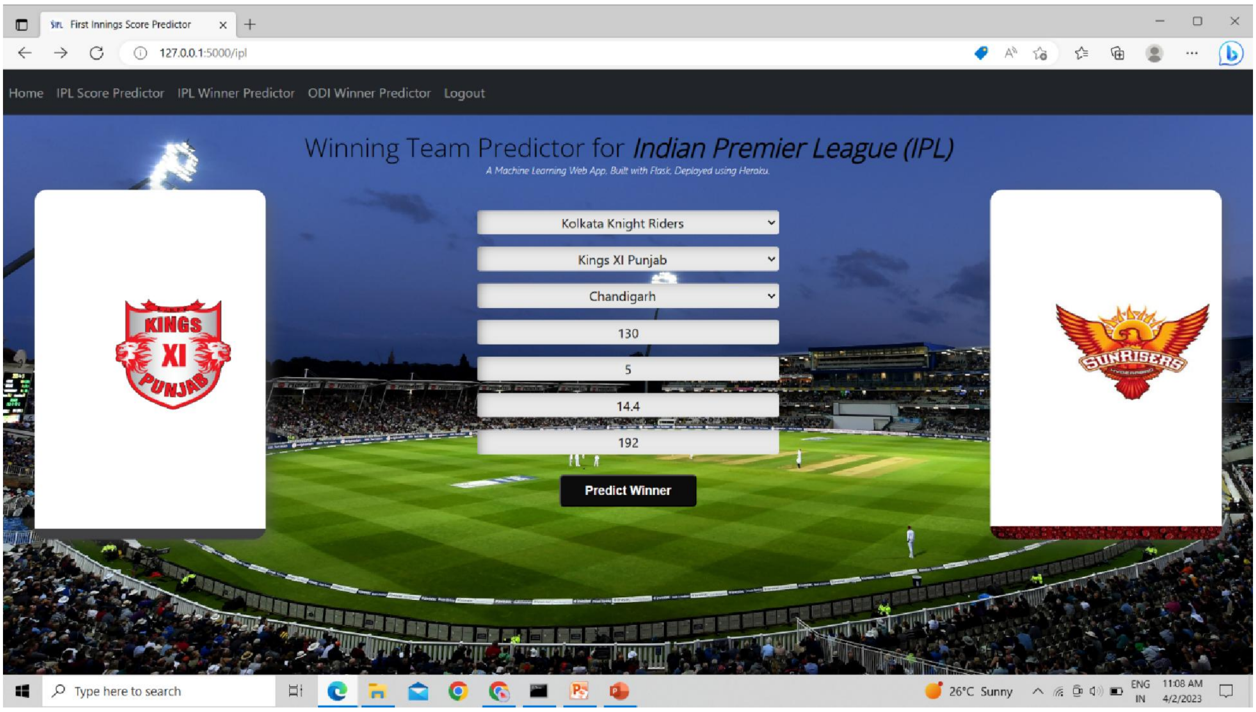
Current runs

Current wickets

Overs finished

Total target Runs

Fig.5.Winning prediction model of IPL



Winning Team Predictor for *Indian Premier League (IPL)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku

Kolkata Knight Riders

Kings XI Punjab

Chandigarh

130

5

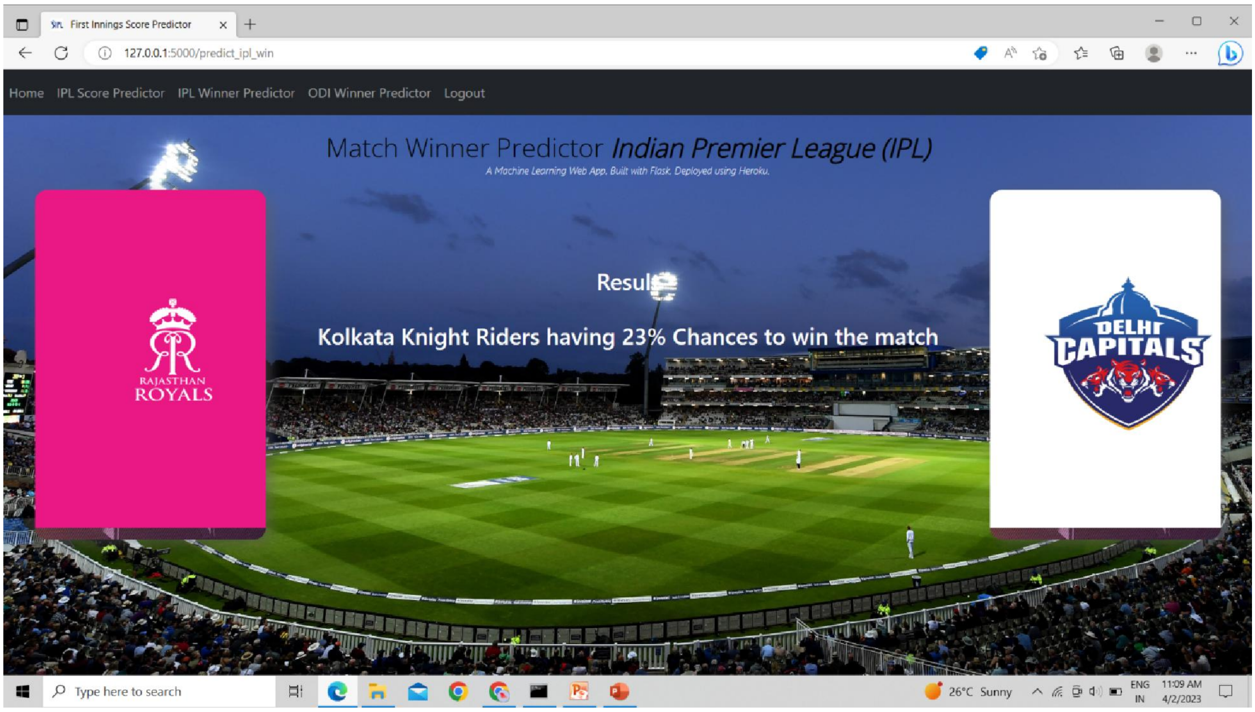
14.4

192

Fig.6.Input to Winning prediction model of IPL

DOI: 10.48175/568

The final output of this model is shown in Fig 7.



Match Winner Predictor *Indian Premier League (IPL)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku.

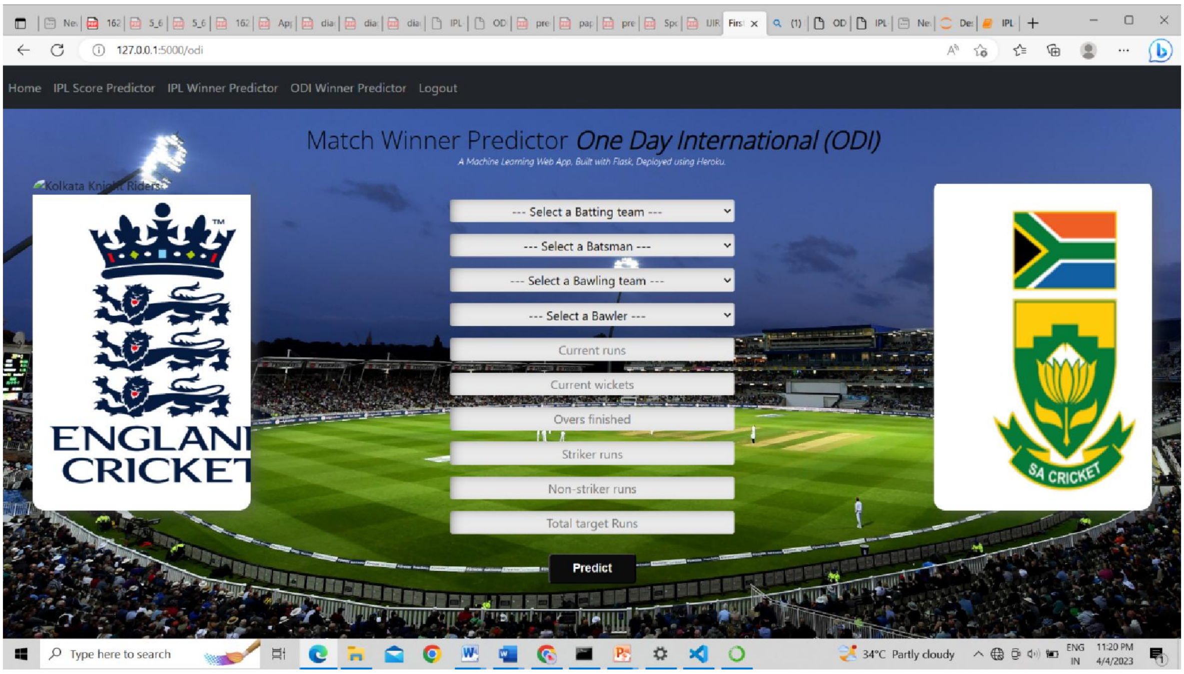
Result

Kolkata Knight Riders having 23% Chances to win the match

Fig. 7.IPL Winning prediction result

7.3 ODI Win Prediction

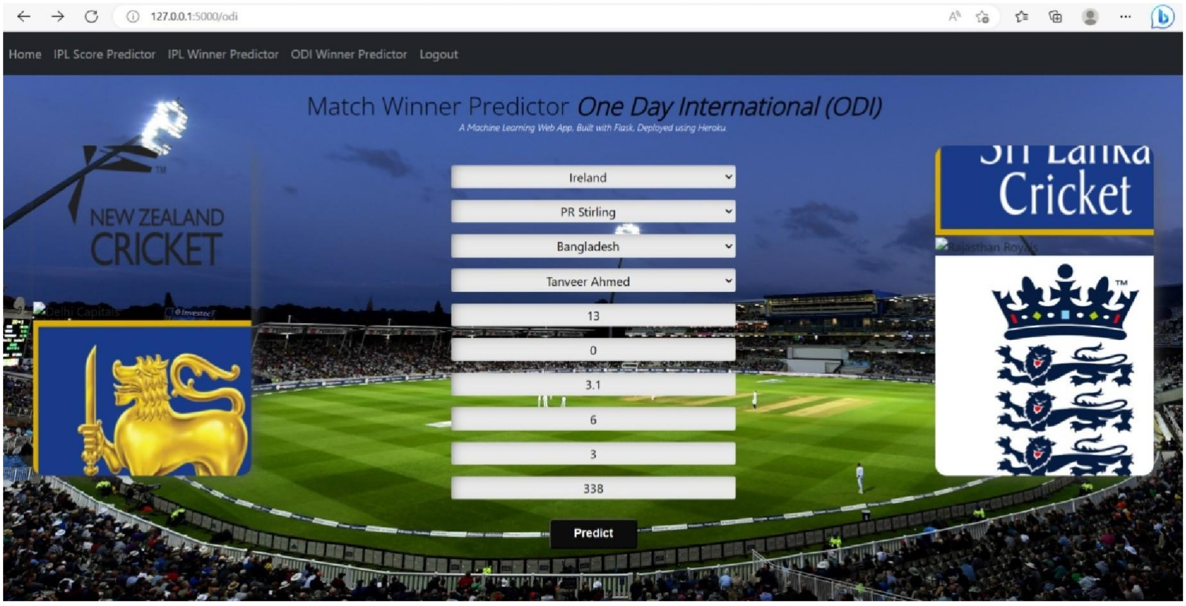
The GUI required team1, team2, Batting team, Bowling team, Batsman, Bowler, Current runs, Wickets, Overs finished, Striker runs, Non- striker runs to predict the Winner of the ODI match as shown in the Fig 8. Model requires the input data as shown in the Fig 9.



Match Winner Predictor *One Day International (ODI)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku.

Fig. 8.Winning prediction model of ODI



Match Winner Predictor *One Day International (ODI)*

A Machine Learning Web App, Built with Flask, Deployed using Heroku

Ireland

PR Stirling

Bangladesh

Tanveer Ahmed

13

0

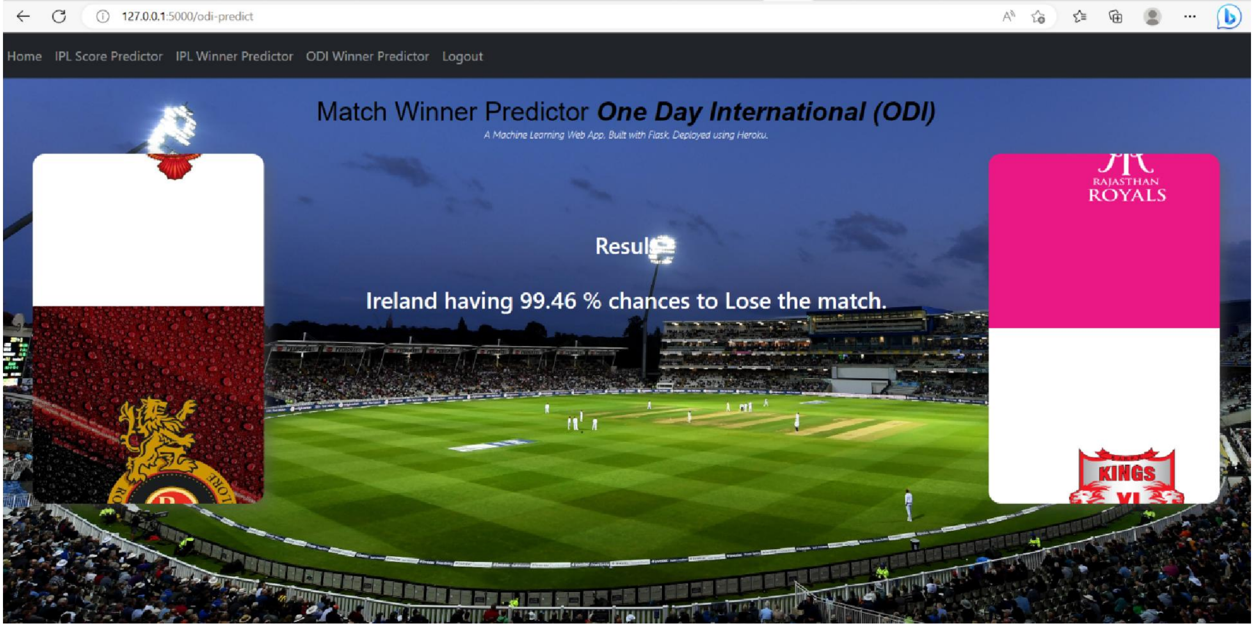
3.1

6

3

338

Fig. 9.Input to Winning model of ODI



Match Winner Predictor ***One Day International (ODI)***

A Machine Learning Web App, Built with Flask, Deployed using Heroku.

Result

Ireland having 99.46 % chances to Lose the match.

Fig. 10.ODI Winning prediction result

VIII. CONCLUSION

Predicting winners in sports, especially cricket, is a challenge and very complex. But you can make this much simpler and easier by incorporating machine learning. This paper gives information regarding IPL score prediction, IPL win prediction and ODI win prediction. By using all the information, we developed a website by using flask. In this study of his, various factors were identified that influence the outcome of Indian Premier League matches. Factors that significantly affect the outcome of an IPL match include the teams competing, the venue, and the city.We get better prediction by using linear regression for the IPL score prediction than other algorithms. We got 80% accuracy by using logistic regression for the IPL win prediction. We got 99% accuracy for the ODI win prediction by using Gradient Boosting Classifier.

REFERENCES

[1]. Monoj Ishi(2022) “WINNER PREDICTION IN ONE DAY INTERNATIONAL CRICKET MATCHES USING MACHINE LEARNING FRAMEWORK: AN ENSEMBLE APPROACH” ,Indian Journal of Computer Science and Engineering (IJCSE) Vol. 13 3 May-Jun 2022.

[2]. I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Excercise, vol. 9, no. 4, pp. 744-751, May 2014.

[3]. Madan Gopal Jhanwar and Vikram Pudi, Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach, Published in MLSA@PKDD/ECML 2016,Computer Science.

[4].]Kevin Desai, Siddhant Doshi, Surekha Dholay ,“Predicting Outcome of ODI Cricket Games”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Volume 3, Issue 01, Special Issue – 2015.

[5]. Preetham HK , Prajwal R , Prince Kumar, Naveen Kumar, "Cricket Score Prediction Using Machine Learning", INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY, Volume 9 Issue 8

[6]. Kumash Kapadia, Hussein Abdel-Jaber, Fadi Thabtah, Wael Hadi,"Sport analytics for cricket game results using machine learning: An experimental study", Emerald Publishing Limited, Published in Applied Computing and Informatics

[7]. Prasad Thorat, Vighnesh Buddhivant, Yash Sahane, "CRICKET SCORE PREDICTION", International Journal of Creative Research Thoughts (IJCRT),Volume 9, Issue 5 May 2021