

IPL First Innings Score Prediction Using Machine Learning Techniques

Mayank Agarwal, Prof. Dr. Archana Kumar

Department of Computer Science & Engineering,
ADGITM (Affiliated to GGSIPU),
Shastri Park, Delhi, India

Abstract- In India, Cricket is one of the most watched and most played sports. India Cricket team calendar is action packed throughout the year and they don't even get rest for even a single month like other countries. So, this huge popularity of cricket has resulted in introduction of Indian Premier League (IPL) by BCCI, India. Now it is conducted among 10 teams. It was started by having 8 teams in the tournament. Since the start of the tournament, it has become the largest and biggest event of cricket in the whole world. People really enjoy this tournament and different players from different playing countries are part of the IPL as well. In this paper we made the model for score prediction using different machine learning regression techniques. In this the different score prediction includes linear regression, lasso regression and ridge regression and then we have calculated the accuracy of each algorithm and chosen the best one. The model used the supervised machine learning algorithm to predict the IPL first Innings Score. In our model the linear regression gave the best result in comparison of the other algorithms so we are using it.

Keywords- Indian Premier League (IPL), Machine learning, Linear Regression, Ridge Regression, Lasso Regression, Prediction, First Innings, Tournament, IPL Score Prediction.

I. INTRODUCTION

Cricket was brought in to North America by way of the English communities as early as the 17th centennial. Cricket is most popular game. Most of the nations complicated in it. In 2008 Indian Premier League, BCCI was settled so many gambling's were taking advantage of it very well. So, there is monstrous demand for the algorithms that thinks highest in rank result of score and triumphant group that is to say more influential. Machine learning is the best habit for prophecy. All algorithms maybe top-secret as supported, alone, supervised knowledge. These algorithms secondhand established the use and the result completed.

As a cricket fan I have always wondered how the first innings score predictor works during a match and when we searched it, I was in the first year of my college year and the result they showed implied that they used different machine learning algorithms and data analytics. So, I decided that time when we will complete a course on machine learning we are going to make a project on this particular problem statement IPL (Indian Premier League) score prediction using machine learning is a process of using algorithms to analyze historical data and make predictions on future IPL cricket match results. This involves collecting data on past IPL matches, such as teams, players, scores, and various performance metrics, and training a machine learning model on this data to predict future match outcomes. The model can then be used to make predictions on future IPL matches, providing

insights into potential winners and helping fans and analysts make informed decisions.

The algorithms used to predict the IPL first Inning Match Score are linear, lasso and ridge regression. In the Linear Regression, labelled data is given to the machine learning model and the labelled data is already known. Linear regression used for the continuous values prediction than classification of the object. Multicollinearity in the data can be analyzed with the help of ridge regression and give the results with more accuracy. It can be used as both classification and regression

II. LITERATURE SURVEY

The research paper of G. Sudhamathy helps to understand the different machine learning algorithms working principal and their implementation. It creates the Model and Training dataset and helps to predict with the help of the model created. The model classifies the data and compares the results and gets accuracy which is the important one [4].

As in the dataset there are many parameters are present. Out of them which parameters are helpful in the project? The factor's affecting concept was taken by Maheshwari in their prediction of live cricket score paper from that we get to know the main factors which required for the prediction of score and the prediction of winning team [2]. [1] The role of classification is clarifying in the paper of Tejinder Singh it gives proper information or use of naive

bias and linear regression. They give the proper knowledge of data collection and preparation also how to train the and test the data is given by them which is more helpful. The support vector machines brief idea is been taken from Aminul Islam Anik paper which is about players performance in this paper the idea about SVM system is given in detailed where the player performance prediction is given by collecting the old information or data [3]. From the literature survey it is concluded that the machine learning is need of prediction.

So, we have concluded the following from the above the literature Survey and we can summaries the following from this. Using the IPL dataset, predict the score of any ipl team. Here we need to look at number of factors while predicting the score. Factors like Ground History, Team Balance, Current Situation, run rate, overs remaining and so on. We will predict the First Innings Score with it of the team batting first. In sports, most of the prediction job is done using regression or classification tasks, both of which come under supervised learning.



Fig 1. Score prediction in 2018 using data analytics. [11]



Fig 2. Ipl First Innings Score Predictor in match (2020) [9]

III. DATASET FEATURES

The approach over here we are using is ML based. So, the basic requirements of an ML algorithm are dataset, training of that dataset using the algorithm and testing the model. So, we have imported dataset from CricSheet. Later on, calculating the accuracy and improving the accuracy by Linear Regression for score prediction of First Innings in the IPL.

When dealing with cricket data, CricSheet is considered as an appropriate platform for gathering the data and thus we took the data from <https://cricsheet.org/downloads/ipl.zip>. It contains data from the year 2007 to 2019. For better accuracy of our model, we have used IPL players' stats to analyze their performance from here. This dataset contains details of every IPL player that has played in the IPL from the year 2007 – 2019.

1. Score Prediction:

For conducting our research, we collected data on all the IPL matches played in 2008 to 2019. The dataset consists of 96015 numbers of rows. Dataset consists 16 columns over which we applied feature selection techniques and selected 9 features in which 8 are input feature and 1 is our target variable. The attributes selected were bat team, bowl team, Cricket Ground, overs, runs, wickets, runs in previous 5, wickets in previous 5 for score prediction.

Table 1. Dataset Attributes and their values.

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

2. Feature Selection:

Feature Selection is process in which we select an optimal set of features from input features set by using feature selection techniques. By removing redundant features, we reduce dimension of data and remove the unwanted columns from our respective dataset and hence resulting in systematic approach of calculating first innings score. Like we don't need the column of date from our data set, we only want consistent teams that are there in the Ipl etc.



Fig 4. Feature Selection Process [12]

IV. BLOCK DIAGRAM AND METHODOLOGY

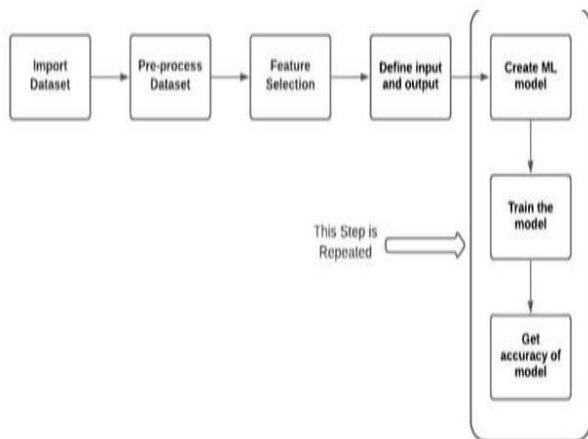


Fig 5. Block diagram of Architecture. [1]

The first step is to import the dataset using the panda's library and then further preprocess the dataset by checking the null values and replace it with the mean or median values of the respective column. The categorical data in the columns is mapped into numerical values. After that the feature selection techniques are applied to the dataset and select only the optimal set of features. The set of input features (X) and the output (Y) are defined in the dataset. The input features are independent of each other and the output feature depends on the input features.

Then the library is imported, and the ML Model is created, and the train-split-test method is used to separate the data into training data and testing data and then train the ML Model with training data and predict using test data. Then the accuracy of the model is calculated by simply taking the ratio of the predicted testing data and the actual testing data. This method is repeated with each ML Algorithm and the accuracy of each algorithm is calculated. Finally, the accuracy of the algorithms is compared and then which of the algorithms is the best for this dataset is determined.

V. ALGORITHMS

We tried to use three machine learning techniques of regression. Selected algorithms from each technique were trained then.

1. Regression:

Regression reasoning uses miscellaneous invention for the computation and established that it thinks the constant advantage. There are certain set of variables are secondhand for the recommendation and the constant range advantage is the target changeable. Based on the request various reversion algorithms are secondhand. There are different reversion methods.

2. 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.

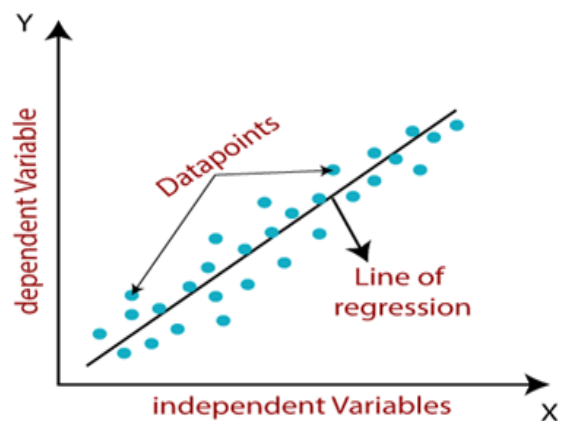


Fig 6. Linear Regression. [13]

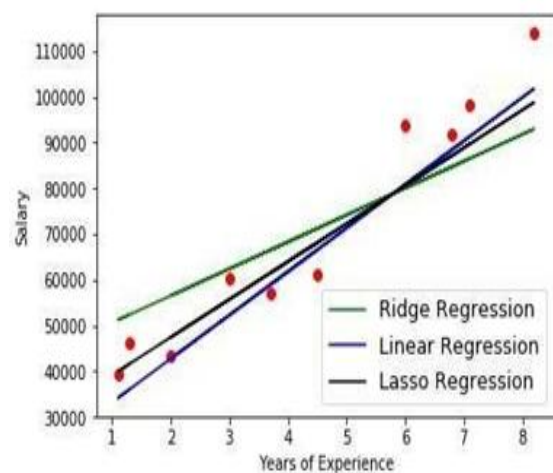


Fig 6. Comparison of different regressions techniques. [3]

3. Ridge Regression:

Ridge regression is also used to predict the continuous values. When the variables used for the prediction greater

than the observations of when multicollinearity present in the data, ridge regression is used set has multicollinearity (correlations between predictor variables).

4. Lasso Regression:

Lasso regression is a type of linear regression that used for predicting the continuous values. Shrinkage is used in the lasso regression. When data values focus towards central point shrinkage occurs. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models.

VI. COMPARATIVE ANALYSIS OF ALGORITHMS

1. Score Prediction Algorithms:

It is found that the linear regression is giving the more accuracy as compared to Ridge regression and Lasso regression. In comparison of the last paper, we have increased the accuracy of our model by 8-10 % and increased the number of parameters as well.

Table 2. Accuracy of the Score Prediction Models.

Algorithm	Accuracy
Linear regression	80.23
Ridge regression	74.00
Lasso regression	75.00

For the score prediction the linear regression gives the highest accuracy result as we see. So, the formula of linear regression for getting theoretical result is as follows.

$$y = B_0 + B_1 * x \dots \text{(linear regression equation)}$$

So here, y is the dependent variable x is independent variable

B₀ is bias coefficient & B₁ is coefficient of x.

Thus, we are using Cost function which helps to get the most accurate possible values for B₀ and B₁. So, as we need the best values for B₀ and B₁ we have converted it into minimization problem where it minimizes the errors between the predicted score and actual score.

VII. 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. This is the first Screen we will see when first load the web application. Here we have to enter and select different things like the Batting Team, Bowling Team, Cricket Ground, Runs scored, overs bowled (has to be

greater than 5), runs scored in the last 5 overs, wickets taken in the last 5 overs.

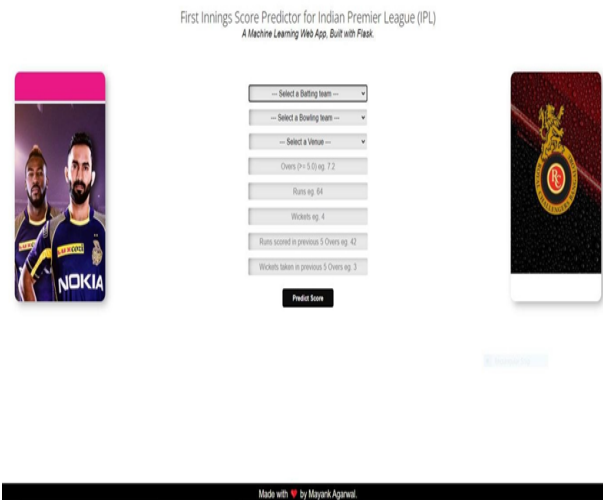


Fig 7. Basic GUI first loaded.

Let's enter some random data that fits inside the scope of the project. First let's select the Batting team as Mumbai Indians, and then select the Bowling Team as Chennai Super Kings, then we will select the cricket ground as Eden Gardens Kolkata. Then enter Runs scored as 78, overs completed as 9, wickets down as 2, runs scored in the last five over as 40 and wickets gone in the last 5 overs as 2 wickets. And see the prediction of score.

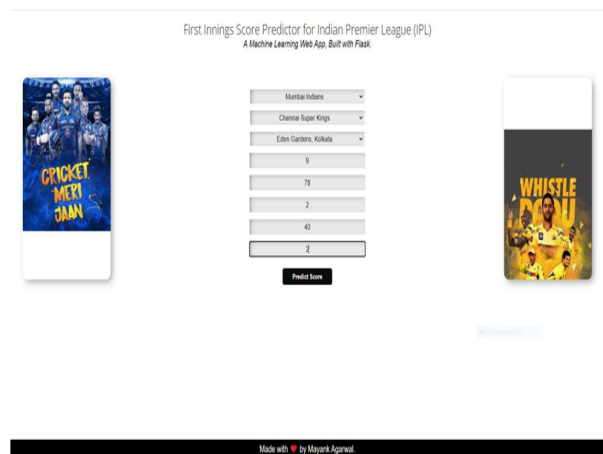


Fig 8. Implementation of IPL Score prediction.

VII. CONCLUSION AND FUTURE SCOPE

Predicting the first innings score in a sport such as cricket is especially challenging and involves very complex processes. But with the introduction of machine learning, this can be made much easier. This project has helped me a lot in understanding the basics of prediction systems in machine learning and related concepts. This IPL prediction mini-project has also helped in learning a new data science operation in Python and its in-built libraries.

The algorithm will be prominent for sports news analysis since it was chosen after being tested on sample data and has been trained on a large collection of historical data in comparison to previously implemented machine learning models, the project shows that the machine learning model can predict the match-winner with greater accuracy. This knowledge will be used in future to predict the winning teams for the next series IPL matches. Hence using this prediction, the best team can be formed. This project opens scope for future work in the field of cricket and predicting other important things like best team of players, best venue, best city, and best fielding decision to win a match. There can be a chance to predict the man of the match for the two teams.

This report will give the important information regarding IPL score prediction, that which parameters are required 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. In Score Prediction analysis accuracy of Linear Regression is more than Ridge and Lasso Regression.

REFERENCES

- [1] "Ipl Cricket Score and Winning Prediction Using Machine Learning Techniques" Nikhil Dhonge Shraddha Dhole, Nikita Wavre, Mandar Pardakhe, Department of Electrical Engineering, MIT Academy of Engineering (Affiliated to SPPU), Pune, Maharashtra, India Issue: 5 May, 2021
- [2] "Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org),ISSN:2349-5162, Vol.8, Issue 6, page no.c437-c444, June-2021
- [3] I. Anik, S. Yeaser, A. G. M. I. Hossain and A. Chakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT).
- [4] "Prediction on Ipl Data Using Machine Learning Techniques in R Package" G. Sudhamathy and G. Raja Meenakshi Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, India CTACT JOURNAL ON SOFT COMPUTING.
- [5] "Prediction of IPL Match Score and Winner Using Machine Learning Algorithms", International Journal of Emerging Technologies and Innovative Research (www.jetir.org),
- [6] Lokhande, Chawan, Pramila, "Prediction of Live Cricket Score and Winning," International Journal of Trend in Research and Development, Volume 5(1), (2018).
- [7] https://en.wikipedia.org/wiki/List_of_Indian_Premier_League_seasons_and_results
- [8] <http://statisticstimes.com/sports/indian-premier-leaguewinners.php>
- [9] https://www.sas.com/en_in/insights/analytics/machine-learning.html
- [10] https://en.wikipedia.org/wiki/Machine_learning
- [11] <https://www.geeksforgeeks.org/ipl-score-prediction-using-deep-learning/>
- [12] <https://medium.com/@mehulved1503/feature-selection-in-machine-learning-variable-ranking-and-feature-subset-selection-methods-89b2896f2220>
- [13] <https://www.javatpoint.com/linear-regression-in-machine-learning>