

# Cricket Score Prediction Using Machine Learning

Utkarsh Sinha<sup>1</sup>, Vyom Tiwari<sup>2</sup>, Er. Ayush Pratap Singh<sup>3</sup>

<sup>1</sup>UG-Students, Department of Computer Science Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India

<sup>3</sup>Assistant Professor, Department of Computer Science Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India

**Abstract** – Cricket is the most popular sport in India, and it is played there in all of its regions in various formats such T20, ODI, and Test. Indian regional teams, the national squad, and international teams all field players in the Indian Premier League (IPL), a national cricket competition. This league became wellknown among cricket enthusiasts due to numerous factors, including live streaming, radio, and TV broadcasts. For internet traders and sponsors, the outcome predictions of IPL matches are crucial. In addition to more conventional variables like the toss, venue, and day-night, we may anticipate a match between two teams based on a variety of characteristics like the team's composition, the batting and bowling averages of each player on the team, and the team's success in prior matches, Cricket is the most popular sport in India, and it is played there in all of its regions in various formats such T20, ODI, and Test. The Indian Premier calculates the likelihood of winning by batting first against a certain team at a specific match location. In this research, utilising machine learning algorithms such as SVM, Random Forest Classifier (RFC), Logistic Regression, and we have suggested a model for predicting the results of the IPL matches. The accuracy of the Random Forest algorithm, which is 88.10%, exceeds other algorithms, according to experimental results.

**Keyword** – Cricket Prediction, Cricket analysis, Lasso Regression, Naïve Bayes, Logistic Regression, Random Forest Classifier.

## 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. The Twenty20 format of cricket is the third and most recent one. This style was developed in 2006, and India won the inaugural world championship in 2007. The short game, which has 20 over, is finished in under three hours. Each team has 20 over to play, and there are only 2 teams involved. Due to the IPL, the t20 format is extremely popular in India. The popularity of the t20 format in India is a result of this event.

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.LITERATURE SURVEY

### 1) CRICKET MATCH OUTCOME PREDICTION USING MACHINE LEARNING

**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.

#### 2) CRICKET SCORE PREDICTION USING MACHINE LEARNING

Abstract - Currently, there is a system which can calculate the current run rate and from it calculates the final score of the team. It doesn't consider the fact about the no of wickets and also where the game is being played. The problem with the current system is that it is unable to predict the score of the 2nd team and also unable to predict the win percentage This system which is developed will have 2 model in it the 1st model predict the score a team will get after playing 50 over from the current situation. The second method predicts the win percentage of both teams

even before the match has started this done by player selection. We found that error in regression toward the mean classifier could be a smaller quantity than Naïve mathematician in predicting match outcome has been sixty-eight ab initio from 2-15 overs to ninety-one until the top of 42th over.

#### 3) CRICKET SCORE PREDICTION USING MACHINE LEARNING ALGORITHMS

Abstract - Cricket is 11 player team sport played on ground. Cricket has huge fan base in India. With such great spectator support and many people try to predict the outcome of matches based on their individual cricket sense. The game has certain rules and scoring system. Factors viz, match venue and individual player performance have great impact on outcome of the match. Such various parameters are highly interdependent on each other which makes it difficult to make precise prediction of the match. In this project, we are going to build a prediction system that takes in data of matches played in past and makes a prediction of future match events such as final score and results in a victory or loss. Our system will predict match outcome by analysing prestored match data using various machine learning algorithms. We intend to use more features such as pitch condition, weather condition, outcome of toss, individual player performance with respect to match venue. Our system finally present quantitative results displayed by best suited algorithm having highest accuracy. Also, demonstrating the performance of our algorithms in predicting the number of runs scored which is one of the most important parameter of match outcome.

#### 4) IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES

Abstract - As cricket is the mostly played game. There are so many series are played in country one of them is Indian Premier League (IPL). Now it is conducted among 8

teams. In these papers the model has been proposed that has two methods the first one is prediction of score and the second one is team winning prediction. 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.

#### 5) Sport analytics for cricket game results using machine learning: An experimental study

Abstract - Indian Premier League (IPL) is one of the more popular cricket world tournaments, and its financial is increasing each season, its viewership has increased markedly and the betting market for IPL is growing significantly every year. With cricket being a very dynamic game, bettors and bookies are incentivised to bet on the match results because it is a game that changes ball-by-ball. This paper 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), Relief and Wrapper. More importantly, machine learning techniques including Naïve Bayes, Random Forest, K-Nearest Neighbour (KNN) and Model Trees

(classification via regression) have been adopted to generate predictive models from distinctive feature sets derived by the filter-based methods. Two featured subsets were formulated, one based on home team advantage and other based on Toss decision. Selected machine learning techniques were applied on both feature sets to determine a predictive model. Experimental tests show that tree-based models particularly Random Forest performed better in terms of accuracy, precision and recall metrics when compared to © Kumash Kapadia, Hussein Abdel-Jaber, Fadi Thabtah and Wael Hadi. Published in Applied Computing and Informatics.

#### 7) The Cricket Winner Prediction with Application Of Machine Learning And Data Analytics

Abstract - With the evolution in the field of Data Sciences, every business firm is adapting latest technologies to grow their business. There are competitions in delivering better management, better quality of evaluations and better services in the market. The only possible way to meet all these qualities is to conduct analysis of data with purity and more accurately. Machine learning is the emerging field to predict future outcomes with existing data and based on these predictions better decisions can be made. Cricket is a well-known game that played and watched around a globe

in 104 countries. Many of these cricket fans want their team to perform good and declare as a winner. To make sure their team's win, team should work on their strengths and team performances. Predicting winner of a cricket match depends on many factors like batsman's performances, team strengths, venues and weather conditions etc. In this research various features have been analyzed to predict the match winner of the game. This research paper is about prediction of an IPL match winner before the match started. The winner of IPL is predicted by training machine learning models on the selected features. For this purpose of model building, different machine learning algorithms has been applied on test and training datasets of different sizes which are

Random Forest, SVM, Naive Bayes, Logistic Regression and Decision Tree. The prediction model will have benefits for cricketing boards like evaluating the team's strength and cricket analysis. For gambling applications and match reporting media this model will be a blessing of disguise.

8) CRICKET SCORE PREDICTION

Abstract - Nowadays the final score of the first innings of any cricket match is predicted using CRR (Current Run Rate) method. The number of average runs scored in an over is multiplied by the total number of overs to get the final score. These kinds of systems are not useful when the T20 matches are considered because in T20 cricket the match can change its state very quickly irrespective of current run rate. The match can change within 1 or 2 overs. So, to get an accurate score prediction we 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.

9) Cricket Analysis and Prediction of projected Score and Winner using Machine Learning Abstract - This paper is about a model that can predict the projected score of 1st inning as well as the winner in a IPL cricket match. The performance of model depends on various features like wickets taken in last 5 overs, runs scored in last 5 overs, overs, overall score and wickets at current ball. The proposed model contains data from IPL matches played between years 2008 and 2019. This paper will give us step by step insights on how one can predict projected score of 1st inning while the match is still in progress. Linear Regression algorithm is used to predict the score. This model explains about 75.226% of data. The model specifically emphasize on using the data from past 5 overs to predict what might be the projected score of the match which has not been considered in any existing model. Using this model, we

can get good insight during the match on how much score will the current batting team obtain.

10) Cricket score and winning prediction using data mining

Abstract - Data Mining and Machine Learning in sports analytics, is a brand-new research field in computer science with a lot of challenge. In this research the goal is to design a result prediction system for a T20 cricket match, in particular for an IPL match while the match is in progress. Different Machine Learning and statistical approach were taken to find out the best possible outcome. A very popular mathematical technique named Multiple Linear Regression is used in order to make comparison of results found. This model is very much popular in predictive modelling. Currently, in Twenty-Twenty (T20) cricket matches first innings score is predicted on the basis of current run rate which can be calculated as the amount of runs scored per the number of overs bowled. It does not include factors like number of wickets fallen, venue of the match, toss. Furthermore, in second innings there is no method to predict the outcome of the match. In this paper a model has been proposed that predicts the score in each of the innings using Multiple Variable Linear Regression along with Logistic regression and finally the winner of the match using Random Forest algorithm.

III.METHOLODOGY

A. System Architecture:

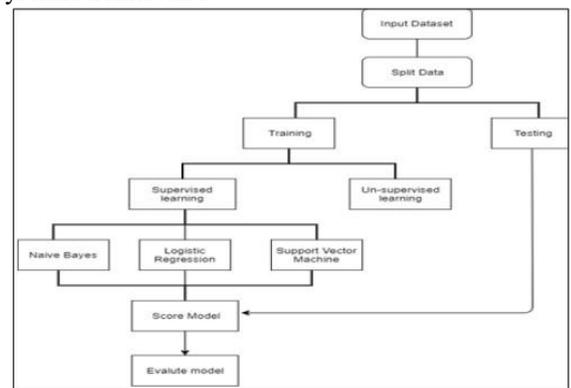


Fig. - Flow Diagram for Search System

First the system will be fed with the input dataset where it comprises data like player details, player score and place where match is played etc. Then the data will be processed further and split into training and testing datasets. Now the training dataset is further split into supervised and unsupervised learning.

Here some suitable algorithms will be applied for the supervised learning datasets and those algorithms are Lasso Regression, Naïve bayes, Logistic Regression, Support vector Machine and Random Forest Algorithms. The suitable algorithm will be picked to predict the

outcome hence it will be matched with testing datasets and score model will be generated.

#### B. Input Pre-processing:

The steps in input pre-processing include:

First the dataset will be loaded and later the analytical rules will be applied. As the data is not pure and cleaned the cleaning process will be performed which clears the outliers. The next step is to train the individual model which is used to measure the accuracy which in return helps the score prediction. For this to happen we need to select the most appropriate model and select the suitable model based on the dataset.

Now after the training and forming the model user will have to give the input. The system will accept the input and will be matched to the model. Later the output will be predicted and the desired score is displayed.

#### C. Algorithm 1) Lasso Regression:

Lasso regression is a regularization technique. It's used over regression strategies for accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunken to a central point as the mean. The lasso procedure encourages simple, straightforward, thin models (i.e. models with fewer parameters). This specific style of regression is well-suited for models showing high levels of multiple regression or when you want to automatise certain elements of model, like variable selection/parameter elimination.

#### 2) Random Forest Classifier:

Random Forest is a classifier that contains variety of decision trees on varied subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of hoping on one decision tree, the random forest takes the prediction from each and every tree and based on the majority votes of predictions, and it predicts the final output. The more the number of trees within the forest results in higher accuracy and prevents the matter of overfitting.

#### 3) Naive Bayes Algorithm:

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to

contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

#### 4) Logistic Regression Algorithm:

Logistic regression is a classification technique borrowed by machine learning from the field of statistics. Logistic Regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The intention behind using logistic regression is to find the best fitting model to describe the relationship between the dependent and the independent variable.

### IV.RESULTS

It is advantageous for many stakeholders to use machine learning to analyse cricket games by taking previous game data, player performance, natural parameters, pre-game conditions, and other features into account. Predicting the result of a game in a dynamic format like T20, when the scenario in a game changes with every ball, is difficult. We have looked into machine learning technologies to see if it can increase the accuracy with which results of matches are predicted for T20 cricket matches. In order to better understand the issue, we have divided it into two scenarios: the Home Team features set and the Toss Winner determination features set. The model constructed on Toss related features yields marginally better outcomes than Home Advantage in terms of the assessment measures utilised, according to an analysis of the results obtained using four different machine learning approaches on 10 years' worth of T20 matches (Accuracy, Precision, Recall, FPs, FNs, etc). When processing the Toss Winner feature set, Lasso

Regression, Random Forest Classifier, Naive Bayes Algorithm, and Logistic Regression Algorithm performed better than the other algorithms because they produced more accurate predictive models than Decision Trees, Probabilistic, and Statistical models. Furthermore, the number of occurrences that the aforementioned algorithms, both FPs and FNs, wrongly classify is low, leading to improved Precision and Recall rates. The aforementioned approach successfully identified 134 instances that were incorrectly categorised as belonging to the "Lose" class and 105 instances—or around 35%—were incorrectly labelled as "Wins." The class independence assumption of the procedure, however, makes the outcomes of Nave Bayes on the Toss Decision subset unpromising. However, the Home Team subset yielded higher results using Nave Bayes. Team management and academics interested in cricket data analytics will help people to analyse and bet well.

## I. CONCLUSION

The goal of this research is to use past data to forecast the final score and match winner. Data Preprocessing, Data Visualizations, Data Preparation, Data Selection, and Machine Learning Model Implementation are some of the fields of Data Science that will come together to conduct the study and forecast the match's score. To accurately forecast the score of innings and obtain the desired outcome, a number of machine learning models will be applied to specified data

## REFERENCES

- [1] Pallavi Tekade, Kunal Markad, Aniket Aamage and Bhagwat Natekar (2020). "CRICKET MATCH OUTCOME PREDICTION USING MACHINE LEARNING"
- [2] Prof. R. R. Kamble, Nidhi Koul, Kaustubh Adhav, Akshay Dixit and Rutuja Pakhare (2021). "Cricket Score Prediction Using Machine Learning"
- [3] Rohit Khade, Nikhil Bankar, Prashant Khedkar and Prof. Prashant Ahire (2019). "Cricket Score Prediction using Machine Learning Algorithms"
- [4] Omkar Mozar, Soham More, Shubham Nagare and Prof. Nileema Pathak (2022). "Cricket Score and Winning Prediction"
- [5] Dhonge, N., Dhole, S., Wavre, N., Pardakhe, M., & Nagarale, A (2021). "IPL Cricket Score and Winning Prediction Using Machine Learning Techniques"
- [6] Kumash Kapadia, Hussein Abdel-Jaber, Fadi Thabtah, Wael Hadi (2019). "Sport analytics for cricket game results using machine learning: An experimental study"
- [7] Daniel Mago Vistro, Faizan Rasheed, Leo Gertrude David (2019). "The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics"
- [8] Prasad Thorat, Vighnesh Buddhivant, Yash Sahane (2021). "CRICKET SCORE PREDICTION"
- [9] Apurva Lawate, Nomesk Katore, Salil Hoskeri, Santosh Takle, Prof. Supriya. B. Jadhav (2021). "Cricket Analysis and Prediction of projected Score and Winner using Machine Learning"
- [10] Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju (2018). "Cricket score and winning prediction using data mining"