

Cricket Data Analytics and Score Prediction

Puneet Choudhary

*Department of Computer Science & Engineering
Apex Institute of Technology
Chandigarh University
Mohali, India
Puneetmahala1101@gmail.com*

Sukhmeet Kour

*Department of Computer Science & Engineering
Apex Institute of Technology
Chandigarh University
Mohali, India
Sukhmeet.e13420@cumail.in*

Abstract – Cricket is the most popular game in India and is played in all regions of India in different formats such as T20, ODI and Test. India's regional teams, national teams and international teams all participate in the domestic football competition, the Indian Premier League (IPL). The league is well known among cricket lovers due to its television, radio and television broadcasts. For online businesses and promoters IPL match results prediction is important. In addition to different factors such as batting, fielding and day-night, we can predict the game between two teams based on various features such as team composition, batting and bowling. most played in India and all parts of India Matches Many formats like T20, ODI and Test. The Prime Minister of India calculates the probability of winning before a match against a particular team at a particular venue. In this study, we prepare a model to predict the outcome of IPL matches using machine learning algorithms such as SVM, Random Forest Classifier (RFC), logistic regression. According to the test results, the accuracy of the random forest algorithm is 88.10% compared to other algorithms.

Contents – Cricket Prediction, Cricket Analysis, Lasso Regression, Naive Bayes, Logistic Regression, Random Forest Classifier.

I. INTRODUCTION

After football, cricket is the most popular sport. Starting in the 15th century, the sport became popular in England. As the number of fans around the world continues to grow, cricket will soon surpass football as the sport with the largest fan

base. In India today, this is not just a game; It has become a religion. There are three main types of text. One Day Internationals are played more than 50 times in a day. The second mode is the test mode used early in the game. The game is played over five days and each team has two matches of 80 to 90 overs each. Within five days, the team must send regularly.

This game type is quite difficult, the player's strength, patience and courage are important. Cricket's Twenty20 format is the third and most accurate format. This style was created in 2006, India won its first world championship in 2007. This short game consists of 20 shots and is finished in three hours. Each team has 20 games to play and only 2 teams participate. t20 format is very popular India due to IPL.

The popularity of t20 format in India is the result of this event. According to IPL viewers, people make their own predictions while watching a particular match. They base their predictions on reality and use various statistics and information to determine who will win. Therefore, it is very important that there is a large market for algorithms that predict the best results and the winning team. We will provide predictions for every IPL match played.

II. LITERATURE SURVEY

1) CRICKET MATCH OUTCOME PREDICTION USING MACHINE LEARNING

Summary - In cricket, especially in the twenty-twenty format, is the most popular and popular format, br > No one can predict who will win till the last ball of the last over game. In India, the Indian Premier League (IPL) started in 2008 and is now the most popular T20 league in the world. That's why we decided to build a learning machine to predict competitive outcomes. Winning in football matches depends on many factors such as home advantage, past performance on the field, record on the same field, complete information of the player, special

information of the opponents and the current situation of your team, including the player. This article briefly outlines the key factors that influence the outcome of a tennis match and the regression model that best fits the data and provides the best prediction.

2) CRICKET SCORE PREDICTION USING MACHINE LEARNING

Summary- There is now a system that can calculate the current running pace and from that calculate the team's final score. The number of wickets and the fact of where the match is played are not taken into account. The problem with the current system is that it cannot predict the second team's score or winning percentage. There will be 2 models of this structure. The first model predicts one group. Points to be obtained after playing 50 innings under current conditions. The second method predicts the percentage for both teams by selecting players before the game starts. The average classifier regression error is smaller than the error of a novice mathematician 2-15. We see that it predicts the outcome of the match from rounds to round 91 starting from zero to round 68 42nd of the session.

3) CRICKET SCORE PREDICTION USING MACHINE LEARNING ALGORITHMS

Summary - Cricket is a team sport with 11 players on the field. Cricket has a huge fan following in India. With such support, many people try to predict the outcome of the game based on their personal opinions. The game has some rules and credits. Factors such as stadium and player performance have a positive impact on the outcome of the game. These variables are interdependent, making it difficult to predict games accurately. In this project, we will create a prediction system that uses data from past matches to predict future matches, including final results, wins and losses. Our system uses various machine learning algorithms to analyze match data and then stores it to predict the match outcome. It takes into account a variety of factors including sugar, weather, shooting results, and player performance in the game. Our final system provides multiple results with the highest accuracy found through optimal methods. Additionally, our algorithm has proven to be effective in predicting results, which is one of the most important factors in competition results.

4) CRICKET_SCORE_AND_WINNING PREDICTION

Summary - As we all know cricket is the most played game. There are so many series in cricket which are played in our country, one of them is the Indian Premier League (IPL). Now it is conducted among 8 teams. Our proposed system consists a model that has two parts the first one is prediction of score and the second one is team winning prediction. In this the score prediction is done with the help of Lasso Regression algorithm whereas in winning prediction.

SVM classifier, decision tree classifier and random forest classifier are used. The model uses the supervised machine learning algorithm to predict the winning. Random Forest Classifier is used for good accuracy and stable accuracy so that desired predicted output is accurate.

5) IPL CRICKET SCORE AND WINNING PREDICTION USING MACHINE LEARNING TECHNIQUES

Summary - Since cricket is the most popular sport. There are many sports in the country and one of them is the Indian Premier League (IPL). He currently plays in 8 teams. The model mentioned in this document has two methods; The first is a bet and the second is a group bet. Among them, the prediction score includes linear regression, lasso regression, and ridge regression, while the prediction method uses SVC classifier, decision tree classification, and random forest classifier. The model uses a machine learning algorithm to predict earnings. The use of random forest classifier has good accuracy and stable accuracy, resulting in the need for accurate output prediction.

6) SPORT ANALYTICS FOR CRICKET GAME RESULTS USING MACHINE LEARNING: AN EXPERIMENTAL STUDY

Summary - Indian Premier League (IPL) is one of the most popular football leagues in the world and its revenue is increasing every season, viewership is increasing and the IPL betting market is increasing. As cricket is an extremely dynamic game, both punters and punters place bets on the outcome of the match as it is a game that changes ball by ball. This article examines machine learning techniques to solve the problem of predicting cricket matches based on historical IPL match data. Interference characteristics of the data were identified using filtering techniques including relationship-based selection, Information Gain (IG), ReliefF, and Wrapper. More importantly, machine learning techniques including Naive Bayes, Random Forests, K-Nearest Neighbors (KNN) and Trees (regression).

7) THE CRICKET WINNER PREDICTION WITH APPLICATION OF MACHINE LEARNING AND DATA ANALYTICS

Summary - As data science grows, all companies are adopting new technologies to improve their business. There is competition in the market to provide better management, better analysis and better service. The only way to meet all these advantages is through clear and accurate data analysis. Machine learning is an emerging field that uses existing data to predict future outcomes and make better decisions based on those predictions. Cricket is a famous game played and watched in 104 countries around the world. Many cricket fans want their team to perform well and achieve victory. In order for the team to win, its strength and team performance must increase. Predicting the winner of a cricket match depends on many factors such as batsman performance, team strength, pitch and weather conditions. In this study, many features were analyzed to predict the winner of the competition. This research paper is about predicting the IPL opponent before the match starts. The winner of IPL is predicted by training a machine learning model based on selected features. For modeling purposes, different machine learning algorithms are used to test and train data on different parameters, including random forest, SVM, naive Bayes, logistic regression, and pruning trees. These prediction models will provide useful results for cricket teams such as evaluation of team strength and cricket analysis. For gambling apps and advertising campaigns, this model will be a blessing in disguise.

8) CRICKET SCORE PREDICTION

Summary - Currently the final score of the first match of every sports match is predicted using the CRR (Current Run Rate) method. Average the score of each match across all matches to get the final score. This type of system is not useful when considering T20 matches because in T20 cricket the game can change its state very quickly even if the current run is fast. Matches can be changed in 1 or 2 rounds. Therefore, in order to get accurate scores, we need to have a system that can better predict the first match. Many people like to watch football matches and also predict the final score. The aim of this research paper is to predict the cricket scores of the IPL match by taking into account the historical data and incorporating various factors that play an important role in the

prediction.

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

Summary - This article is about a model that can predict the winner and expected score of the first match of an IPL cricket tournament. The performance of the model depends on various factors such as last 5 wickets, last 5 wickets, number of overs, total runs, wicket with current ball, etc. The proposed model includes IPL match data from 2008 to 2019. This article will help you understand step by step how to predict the expected score for the first half of the action. A linear regression algorithm was used to estimate the scores. The model explained approximately 75.226% of the data. This model provides a unique opportunity to use data from the previous five matches to predict the expected outcome of matches not included in the current model. Using this formula can give you a good idea of how many runs your current baseball team will score per game.

10) CRICKET SCORE AND WINNING PREDICTION USING DATA MINING

Summary: Data mining and machine learning in sports research is a new field in computer science and poses great challenges. The aim of this study is to generate probability prediction for T20 cricket matches, especially IPL matches during regular matches. Different machine learning and statistical techniques are used to find the best results. The findings were compared using a very popular mathematical technique called horizontal multivariate analysis. This model is very popular in forecasting models. Now, in twenty-twenty (T20) cricket, the first innings score is estimated based on the current run rate, which can be

calculated as the number of points per run. Number of wickets lost, field of play, number of innings etc. It does not include many factors such as in addition, it is impossible to predict the outcome of the game in the second game. This article introduces a model that uses multivariate linear regression and logistic regression to predict the score of each game and finally predicts the winner of the game using valley test random forest.

III. METHODOLOGY

A. System Architecture:



Fig. - Flow Diagram for Search System

First, the system receives login data including player details, player scores, and tournament positions. The data will be further processed and divided into training and testing data. Training materials are now divided into supervised learning and unsupervised learning. Here some suitable algorithms will be used to track learning data, these algorithms are Lasso Regression, Naive Bayes, Logistic Regression, Support Vector Machine and Random Forest algorithms. An appropriate algorithm will be selected to predict results, compare them with test data, and create benchmarks.

B. Input Pre-processing:

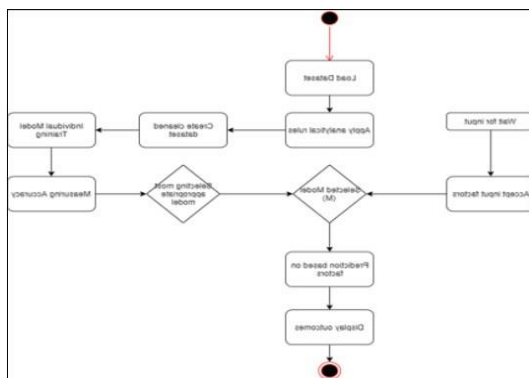


Fig. - Flow Diagram of Input Pre-processing
The steps in input pre-processing include:

First load the configuration file and then use the analysis code. Since the materials are dirty and filthy, cleaning will be carried out to remove the material. The next step is to train a separate model that evaluates accuracy and therefore generates a prediction score. For this, we need to choose the most appropriate model and choose the appropriate model according to the data set.

Now after training and building the model the user needs to give feedback. The system will take the input and match it to the model. The output will then be estimated and the required score will be displayed.

C. Algorithm

1) Lasso Regression:

Lasso regression is a common regression method. Used in regression strategies for accurate predictions. This model uses contraction. Compression compresses data values toward the center. The lasso technique promotes simple, fair, and economical models (i.e., fewer models). This particular regression style is ideal for models that exhibit high-level multiple regression, or when you want to automate certain aspects of the model (such as variable selection/removal).

2) Random Forest Classifier:

Random forest is a classifier that has many decision trees on different subsets of a given dataset and averages them to improve the prediction accuracy of the dataset. Instead of investing money in a decision tree, a random forest makes a prediction from each tree and predicts the final outcome based on the vote of the majority of the prediction. The more trees in the forest, the more accurate it will be and the problem of overfitting will be avoided.

3) Naive Bayes Algorithm:

Naive Bayes is a simple method for generating classes: a model that assigns class labels to a problem, represented as a vector of positive values, where class labels are taken from a problem. There is no single algorithm for training such operators, but there is a family of algorithms based on the same principle: all Naive Bayes classifiers assume a variable whose value is specifically independent of the value of other features. For example, if the fruit is red, round and about 10 centimeters in diameter, it is considered an apple. The Naive

Bayes classifier considers each of these features to be independently associated with the chances of the fruit being an apple, regardless of the relationship between color, roundness, and shape and diameter.

4) Logistic Regression Algorithm:

Logistic regression is a classification method borrowed from machine learning statistics. Logistic regression is a technique used to analyze data using one or more outcome variables. The purpose of using logistic regression is to find the best model that explains the relationship between neighborhood and independence.

IV. RESULTS

It is good to use machine learning to analyze sports matches by taking into account historical match data, players' performance, negative, pre-match and other features for many participants. In a dynamic format like T20, where the match changes with every ball, it is difficult to predict the outcome of the match. We looked at machine learning to see if it could improve the accuracy of predicting match outcomes in T20 cricket. To better understand the problem, we divide the problem into two cases: the home team environment and the winner-winner decision.

A model based on relevant features and enabling better results: Using four different machine learning models based on 10 years of T20 matches (actual, true, return, FP, FN, etc.) Evaluation of the results obtained using the evaluation method Lasso Regression, Random Forest Classifier, Naive Bayes and Logistic Regression algorithms outperform other algorithms when processing the Toss Winner feature set because they produce better predictive models than decision trees, inference and more analytical models. Additionally, the above algorithms (FP and FN) have a small number of misclassifications, thus increasing precision and recall. The above process successfully classified 134 cases as the "Failure" category and 105 cases (about 35%) incorrectly as the "Victory" category.

However, the assumption of class independence of the process makes the results of Naive Bayes negative in terms of decision impact. However, the home team achieved better results using Nave Bayes. Team management and courses with interest in analyzing polished data will help people

analyze and bet better.

V. FUTURE SCOPE

This survey helps to propose a model that helps in the prediction and analyzes of the team and player performance. The useful applications are online fantasy games, used by team analyst, which provides stats to cricket lovers and they can also use to access an opponent's strengths and weakness. The IPL prediction helps people who are willing to play online fantasy games, such as dream11, MPL, and other online platforms.

Future research should analyze different supervised and unsupervised machine learning techniques and feature selection techniques with additional performance metrics for better IPL prediction. Hence, the scope of the project is to build predictive model that works with maximum accuracy and includes all the important factors that influences the results is taken, which will work with maximum accuracy and it should consider all important factors that could influence the result.

CONCLUSION

The aim of this study is to predict the final score and winner of the match using historical data. Data visualization, data visualization, data planning, data selection, and machine learning are some of the data science techniques that come together to perform research and predict scores for the competition. Various types of machine learning will be applied to the data set to complete the game's bets and achieve the desired results.

REFERENCES

- [1] Pallavi Teja, Kunal Markad, Aniket Amage and Bhagwat Natekar (2020). "CRICKETMATCH 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, Nomesh Katore, Salil Hoskeri, Santosh Takle, Prof. Supriya. B. Jadhav (2021).
 - [10] Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju (2018). "Cricket score and winning prediction using data mining.
 - [11] T. B. Swartz, P. S. Gill, D. Beaudoin, and B. M. Desilva, —Optimal batting orders in one-day cricket, *Computers & Operations Research*, vol.33, no. 7, pp. 1939–1950, 2006
 - [12] Tamanna Siddiqui, Mohammad Alkadri, Najeeb Ahmad Khan, —Review of Programming Languages and Tools for Big Data Analytics, *International Journal of Advanced Research in Computer Science*, vol.8,no.5,May-June 2017.
 - [13] Tamanna Siddiqui, Mohammad Alkadri, Najeeb Ahmad Khan, —Review of Programming Languages and Tools for Big Data Analytics, *International Journal of Advanced Research in Computer Science*, vol.8,no.5,May-June 2017.
 - [14] Predicting The Match Outcome in One Day - jssm.org. [Online]. Available: <http://www.jssm.org/volume05/iss4/cap/jssm-05-480.pdf&p=DevEx.LB.1,5063.1>. [Accessed: 11-Aug2018].
 - [15] C. Deep, C. Patvardhan, and C. Vasantha, —Data Analytics based Deep Mayo Predictor for IPL-9, *International Journal of Computer Applications*, vol. 152, no. 6, pp. 6–11, 2016.
 - [16] F. C. Duckworth and A. J. Lewis, —A Fair Method for Resetting the Target in Interrupted One-Day Cricket Matches, *Operational Research Applied to Sports*, pp. 128–143, 2015.
 - [17] Vijay Ramakrishnan, Sethuraman K, and Parameswaran R, —Target Score Prediction in the game of Cricket. [Online]. Available: https://people.ucsc.edu/~praman1/static/pub/ML_Proje ct_CS7641_report.pdf. [Accessed: 11-Aug-2018].
 - [18] —Analysis on Attributes Deciding Cricket Winning, *Scribd*. [Online]. Available: <https://www.scribd.com/document/357690109/Analysis-on-Attributes-Deciding-CricketWinning>. [Accessed: 11-Aug-2018].
 - [19] S. Muthuswamy and S. S. Lam, "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in *Industrial Engineering Research Conference*, 2008.
 - [20] G. D. I. Barr and B. S. Kantor, "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," *Operational Research Society*, vol. 55, no. 12, pp. 1266-1274, December 2004.
 - [21] S.. R. Iyer and R. Sharda, "Prediction of athletes performance using neural networks: An application in cricket team selection," *Expert Systems with Applications*, vol. 36, pp. 5510-5522, April 2009.
 - [22] Lemmer, H. H. (2008). Analysis of players' performances in the first Cricket Twenty-World Cup Series. *South African Journal for Research in Sport*, 30(2), pp.71- 77.

