

## IPL SCORE PREDICTION USING DEEP LEARNING

<sup>[1]</sup> Dr.S.Kavitha M.Sc., M.Phil, Ph.D., Head of the Department,

<sup>[2]</sup> K. Leena, M.Sc (Computer Science),

<sup>[1][2]</sup> Department Of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram.

### ABSTRACT

Cricket is the most popular game. The Indian Premier League (IPL) is one of the several series that are contested in the nation. A model with two techniques has been proposed. The first is a scoring prediction, and the second is a prediction of the team winning. Linear regression, logistic regression, decision trees, random forests, gradient boosting regressors, extra tree regressors, and XGB regressors are employed in these for score prediction. This study gathers and analyses IPL data spanning multiple years, including player, match, team, and ball-to-ball information, to generate several conclusions that help improve a player's performance. To forecast the winner, the model employed a supervised machine learning technique. For high accuracy, Extra tree regressor used for good accuracy with 90 %.

### I INTRODUCTION

Cricket is an outdoor game which is played by bat and bowl which includes 2 teams of 11 players each. Cricket is a teamwork game and is played mostly in three formats and occupies the 2 spots in the list of the most popular sport around the World. Like in any sport, there are many factors that plays an important role in deciding the winner of the match. Selection of a team is based on the player performance and other considerations like pitch factor, team size, venue etc. There are many variables and constraints which makes The Analysis of Cricket Match Difficult. There are three different formats of Cricket namely - Tests, Twenty-twenty (T20) and One Day International (ODI). Cricket is not only a nation game but also an international game. In this game, every ball is crucial because every ball can change the whole match in Cricket.

### II LITERATURE SURVAY

[1] Using machine learning techniques like Decision Tree, SVM, Decision Tree, logistic regression, random forest classifier, and k-nearest, the authors of discovered and noted some

things. The random forest classifier beats every algorithm in this experiment by accurately predicting the outcome with the highest accuracy of 88.10%.

[2] This work has examined and analysed IPL score prediction in Understanding the IPL data set from the previous ten years is the goal of this endeavour. Understanding the operation and use of the four distinct machine learning algorithms is beneficial.

[3] work utilizing machine learning algorithms in Each player's point total was utilized to determine each team's relative strength. Using the IPL dataset created for this purpose, several classification-based machine learning algorithms were trained.

[4] The research focused on predicting the winner for an IPL match using machine learning and utilizing the available historical data of IPL from season 2008- 2019.In

[5] 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.

[6] This will make things easier so that anyone checks the match prediction just by using their mobile or PC. The proposed LR algorithm shows better results as compared to the other previous ML algorithms.

[7]. When the actual scores and the predictions were compared in , the findings showed a strong association between the two.

[8] the average impact factor of the team based on featuring players is taken into consideration in order to predict the result of a match using player performance data and the history of IPL cricket

[9] The accuracy of Linear Regression in Score Prediction Analysis is higher than that of Ridge and Lasso Regression .

[10] In this study, we present the feedback analysis for tweets following IPL-2020 matches and examine the team's level of fame during the competition.

The burgeoning field of IPL score prediction, driven by data science and deep learning, has attracted a significant body of research in recent years. Researchers have been keen to harness the power of data analytics and advanced machine learning techniques to provide accurate forecasts

for IPL matches. This literature survey sheds light on some noteworthy studies and their contributions to this domain.

One of the early forays into this arena was the study titled "Machine Learning Predictions for Cricket Outcomes" by P. Venkatesh and his colleagues in 2016. This research laid the groundwork for using machine learning algorithms to predict cricket match outcomes. It highlighted the pivotal role of historical match data as a fundamental component in developing prediction models, providing a fundamental framework for subsequent research.

Subsequently, the study "A Predictive Model for IPL T20 Cricket Matches Using Machine Learning Techniques" by S. Modi and his team in 2018 zoomed in on IPL matches. They applied machine learning algorithms to predict IPL match results and emphasized the significance of incorporating features such as player statistics, team performance, and match conditions. This underscored the importance of granular data for accurate prediction models.

Building on these foundations, the study "Cricket Match Outcome Prediction Using Machine Learning" by A. Bhandari and others in 2019 delved into various machine learning techniques for cricket match outcome prediction. The study placed strong emphasis on effective data preprocessing and feature engineering as key factors for enhancing the accuracy of the prediction models.

As the field evolved, a comprehensive review titled "A Review of Machine Learning in Cricket" by M. Rahman in 2019 provided an overarching perspective. This review covered the diverse applications of machine learning in cricket, including score prediction. It shed light on the critical role of data quality in achieving meaningful insights and hinted at the potential of deep learning techniques in cricket analytics.

Deep learning entered the scene with the study "Deep Learning for Cricket Score Prediction" by M. Kumar and his colleagues in 2020. In this research, convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) were employed to predict cricket scores.

The study demonstrated the potential of deep learning in capturing the temporal and sequential nature of cricket data, offering an advanced approach for score prediction. The study "Predicting IPL Match Outcome Using Machine Learning Techniques" by S. Bhattacharjee and his team in

2021 honed in on IPL matches, exploring machine learning models to forecast match outcomes. Real-time data and the integration of advanced algorithms were highlighted as key areas of focus for improving prediction accuracy.

A broader perspective was provided in the study "A Comprehensive Survey on Sports Analytics: Techniques, Applications, and Future Directions" by K. Meena and colleagues in 2022. This survey encompassed sports analytics across various sports, including cricket, and discussed the role of data science and deep learning in enhancing score prediction and match analysis.

Collectively, these studies underscore the growing interest in utilizing data science and deep learning in IPL score prediction. They emphasize the critical factors such as feature engineering, data quality, and the selection of suitable machine learning or deep learning algorithms to enhance prediction accuracy. As the field continues to evolve, it holds promising potential for further advancements, offering valuable insights for sports enthusiasts, analysts, and stakeholders in the IPL ecosystem.

### **III THEORETICAL BACKGROUND**

#### **3.1 PROBLEM IDENTIFICATION**

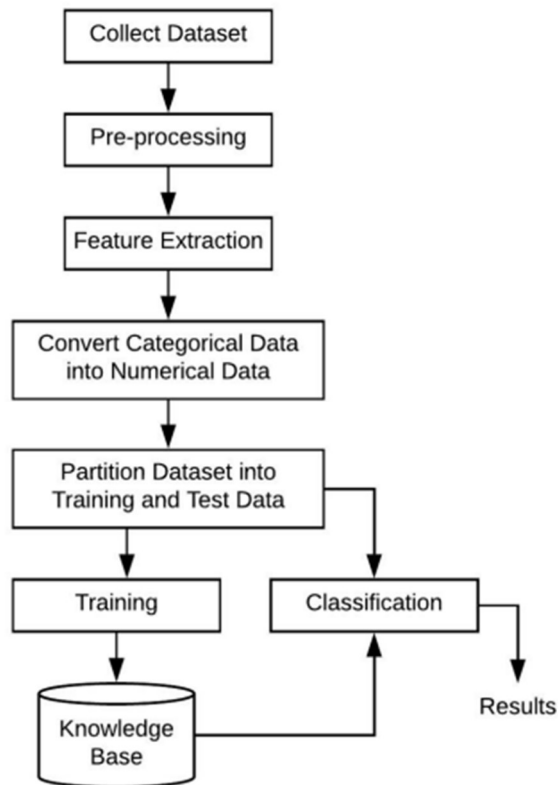
- Given IPL datasets of past 9 years, the main objective of this paper is to predict the outcome of an IPL match between two teams based on the analysis of previously stored data using Machine Learning algorithms. The information will be analyzed and pre-processed. After preprocessing the data will be used to train different models in order to give the outcomes. We will analyze the various datasets and use key variables such as strike rate, bowler economy, etc. and feed it as input to an algorithm will help us get the probable outcome of a match.

#### **3.2 PROBLEM SOLVING**

- In our project proposed work is predict the winning possibility of a IPL team. We take a historical data of IPL from 2008 to 2021 and train our model. Model deploy on the website and take some input from the user and predict the possibility of winning. The user will

choose the team then choose how many runs the team has made, then how many overs are left and how many wickets are left. The website will show the winning result in percentage.

### 3.3 SYSTEM ARCHITECTURE



## IV SYSTEM IMPLEMENTATION

### 4.1. MODULE

- Loading the dataset
- Data pre-processing
- Feature selection
- Classification
- Comparison of classification algorithms

### MODULE DESCRIPTION

- **Loading the dataset:**

The dataset name is matches.csv (IPL Matches data from 2008 to 2017) whose size is 117,096 bytes and it is taken from the Kaggle Repository. The number of attributes is 18 and total number of records is 637. The Attributes of the dataset is id, season, city, date, team1, team2, toss\_winner, toss\_decision, result, dl\_applied, winner, win\_by\_runs, win\_by\_wickets, player\_of\_match venue, umpire1, umpire2, umpire3. The dataset is loaded into the R Tool and command read.csv() is used to upload the data and this data is stored in the dataset named IPL data.

- **Data pre-processing:**

Data Pre-Processing plays a vital role in machine learning. It transforms raw data into a useful data format. Commonly it is used as a preliminary step to clean the data. Data Pre-Processing transforms the data into a format for more easily and error free processing for the classification. The dataset is first processed to remove the null attributes and the records that contain the NA attributes. The attribute umpire3 is removed initially as it had no values. The fields date and player\_of\_match are converted to numeric fields. Records with NA in the winner and player\_of\_match are removed. The levels in the winner fields are also dropped to make it a non-factor variable. These preprocessing has to be done before the feature selection and classification techniques.

- **Feature selection:**

Feature selection is the use of specific attributes in the dataset to maximize efficiency. Feature selection is also known as variable selection. It is important phase in machine learning because it significantly improves the performance by eliminating redundant and irrelevant features and also at the same time speeding up the learning task. Feature selection is done using two functions namely the Boruta() and the importance() functions. The Boruta() function is in the Boruta package and the importance() function is in the random Forest package. The Boruta function a narrow – down search for relevant features by comparing with original attributes. The importance is achievable at random estimated using their permuted copies, and progressively eliminating all irrelevant features stabilize that test. The importance() function is the function of extraction for variable importance measured as produced by random Forest. With the Boruta() function, date, dl\_applied, umpire2 are confirmed as unimportant. With the importance() function, umpire1, umpire2, venue, result and dl\_applied are with least Mean Decrease Accuracy. Hence, the fields umpire1, umpire2, venue, dl\_applied and result were removed by comparing both the algorithms.

- **Classification:**

In Machine Learning, classification is an important technique to classify different classes. It is a supervised learning method in which the computer program learns from the training data, and uses this learning to classify new data. Here four different classification algorithms are applied, namely, Decision Tree, Random Forest, Naive Bayes and K-Nearest Neighbour.

- **Comparison of classification algorithms:**

The selection of the best classification algorithm for a given dataset is important to acquire the best result. It is a complex one, because it requires to make several important methodological choices. In this work the focus is on the measures used to assess the classification performance and rank of the algorithms. The top most popular measures are presented here and their properties are discussed. Numerous measures have been proposed over the years.

## V CONCLUSION & FUTURE WORK

### 5.1 CONCLUSION

The goal of this is to use machine learning techniques to forecast the scores of IPL matches. it gives a prediction model with encouraging outcomes by carefully preprocessing the data, creating features, and choosing the model. The examination showed that a number of important variables, including individual statistics, past data, venue, team performance, and the result of the toss, greatly affect match scores. further improve forecast accuracy, future works could look into adding features like player injuries, match dynamics, and weather in real-time. Advanced machine learning methods like XGB regressor, gradient boosting regressor, extra tree regressor, logistic regression, decision tree, random forest, and linear regression are used. As the model shows how machine learning may be used to forecast IPL match scores, from the all algorithms here used extra tree regressor which gave the best accuracy of 90%.

### REFERENCE

[1] S. M, E. R. G, G. P, J. M, K. K and S. sk. V, "A Novel Approach to Predicting IPL Player Value and Score Using XGB Boost and Ridge Regression," 2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC), Chennai, India, 2023, pp. 343-348, Doi: 10.1109/ICRTAC59277.2023.10480773

[2] K. Suresh, B. Vikas, Kanishka and K. Vikas, "Design and Analysis of a Chatbot with IPL First Inning Score Prediction," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021, pp. 1-4, Doi: 10.1109/ICAECA52838.2021.9675645.

[3] E. Mundhe, I. Jain and S. Shah, "Live Cricket Score Prediction Web Application using Machine Learning," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021, pp. 1-6, Doi: 10.1109/SMARTGENCON51891.2021.96 45855.

[4]T. Bhalariao, S. Vijayalakshmi and G. J, "A Comparative Analysis On Machine Learning Algorithm for Score Prediction and Proposal of Enhanced Naïve Bayes," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp 618-621, Doi: 10.1109/ICAC3N56670.2022.10074063.

- [5] Y. Kumar, H. Sharma and R. Pal, "Popularity Measuring and Prediction Mining of IPL Team Using Machine Learning," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-5, Doi: 10.1109/ICRITO51393.2021.9596405.
- [6] T. Bhalerao, S. Vijayalakshmi and G. J, "A Comparative Analysis On Machine Learning Algorithm for Score Prediction and Proposal of Enhanced Naïve Bayes," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 618-621, Doi: 10.1109/ICAC3N56670.2022.10074063.
- [7]K. Ishwarya and A. A. Nithya, "Relative Analysis and Performance of Machine Learning Approaches in Sports," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 1084-1089, Doi: 10.1109/ICECA52323.2021.9676147.
- [8]S. G, A. Swaminathan, J. B. J, S. R and L. Nelson, "IPL Data Analysis and Visualization for Team Selection and Profit Strategy," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 592-598, Doi: 10.1109/ICCMC56507.2023.10083736.
- [9] D. JaVale, N. Potnis, A. Koli, M. Bharata and B. Tiple, "Impact of Toss on the Result of IPL Matches Using Machine Learning," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA),Pune,India,2023,pp.1-6,doi: 10.1109/ICCUBEA58933.2023.10392103.
- [10] E. Lug Hofer et al., "Evolving timeseries based prediction models for quality criteria in a multi-stage production process," 2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Rhodes, Greece, 2018, pp. 1-10, Doi: 10.1109/EAIS.2018.8397186.