International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

IPL SCORE PREDICTION USING MACHINE LEARNING

Dr K Ramesh babu*, Rachakonda Pujitha **, Padabakkala Sowmya *** .
Turaka Bhakthavatsalam ****, Vintha Preethi *****
CSE Department, GVR&SCET, JNTUK, Andhra Pradesh, INDIA.

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 % .

Keywords: Linear Regression, logistic regression, decision tree, random forest, gradient boosting regressor, extra tree regressor, XGB regressor IPL Winning Prediction, IPL Score Prediction, ball-to-ball Statistics.

1.INTRODUCTION

The introduction of technology has brought about a revolutionary alteration in cricket, one of the most popular sports in the world, redefining our perceptions and experiences of the game. The Indian Premier League (IPL) is the centre piece of this cricket revolution; it is a spectacular display of skill, cunning, and pure enjoyment. Machine Learning (ML) presents itself as a game-changer in the ongoing quest to forecast match outcomes as the cricketing globe changes. The use of machine learning techniques to sports analytics has created new opportunities in recent years for analysing and predicting match results. Cricket's unpredictable nature, particularly in the T20 format of the IPL, offers machine learning algorithms a difficult but fascinating platform to demonstrate their ability to predict. Combining sports and technology together promises to improve spectator experiences while also providing teams, coaches, and analysts with insightful information on the dynamics of games. This paper explores the fascinating field of machine learning-based IPL score prediction. To identify the patterns and trends that affect match outcomes, we use a variety of characteristics, including pitch conditions, player statistics, historical data, and other variables. We want to explain the intricacies of the game and offer a view into the future, one in which projecting IPL scores will no longer be only a speculative endeavour, thanks to machine learning.

2.LITERATURE SURVEY

[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. In [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. In [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.

3. SUMMARY OF LITERATURE SURVEY

Studying all of the aforementioned publications revealed gaps in knowledge, limits in current methods, and lower accuracy when compared to one another. Their obtained results show a lack of accuracy. Furthermore, some algorithms to estimate IPL score prediction encounter difficulties due to regional diversity in some publications. The dataset may have an impact on the accuracy. The following regressors were employed in this paper: gradient boosting, extra tree, XGB, decision tree, logistic regression, and linear regression.

4.PROPOSED ARCHITECTURE

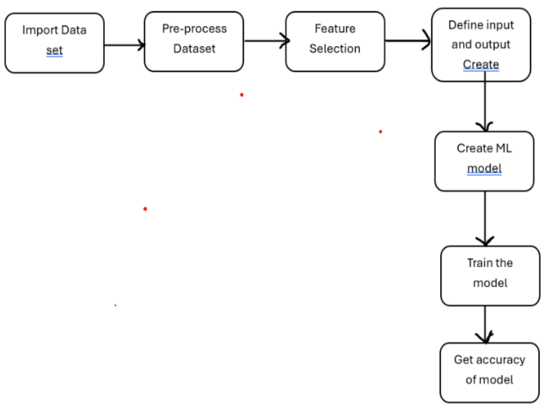


Fig 4.1 IPL score prediction Architecture

Data collection

In order to capture the several aspects influencing match outcomes, a comprehensive strategy is used in the data

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

collection process for IPL score prediction. First, a compilation of past match data from several IPL seasons is made, which includes information on team lineups, batting orders, individual player performances, pitch and weather conditions, venue statistics, and match outcomes. From the fig 4.1 IPL score shown the predictive models are trained using this historical dataset, which helps them find patterns and connections between various factors and match results. Up-to-date information on player injuries, team tactics, and other dynamic aspects preceding each match is also incorporated using real-time data feeds. Sources of structured data include official IPL sources, statistical databases, and cricket analytics platforms; news outlets and social media offer additional perspectives on public opinion and team dynamics. The predictive models can be trained to provide educated predictions about team scores in upcoming IPL matches by combining and preprocessing this wide range of data sources. This would enable data-driven decision-making for stakeholders in the cricket ecosystem.

ALGORITHMS USED

In machine learning mainly there are two types of algorithm models they are: Classification Algorithms, Regression Algorithms.

Classification Algorithms

Classification algorithms are used to predict categorical labels or classes for new data points based on past observations. The goal is to learn a mapping from input features to predefined categories.

i). Logistic Regression Fig 4.2 shows below explain Logistic regression is a statistical method and a type of predictive analysis used in machine learning tasks. Logistic regression statistical method used for binary classification problems, where the outcome variable y is categorical and has only two possible outcomes (0 and 1, or "yes" and "no"). It models the probability that a given input belongs to a certain category. For example, in medical diagnosis, logistic regression can be used to predict whether a patient has a particular disease or not based on various features such as age, gender, blood pressure, etc.

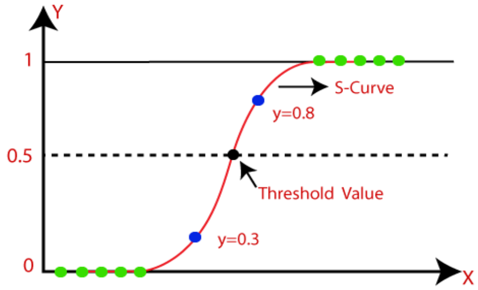


Fig 4.2 logistic regression

ii). Decision tree Fig 4.3 shows below explain A decision tree is a machine learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences by creating a tree-like structure of decisions, making it an intuitive and visually interpretable method. Decision trees split the data into branches to make predictions, using the structure of a tree consisting of nodes and leaves. A Decision Tree is a versatile and intuitive machine-learning algorithm used for both classification and regression tasks. It's a tree-like structure where an internal node represents a feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome (or class label). The below diagram explains the general structure of a decision tree:

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

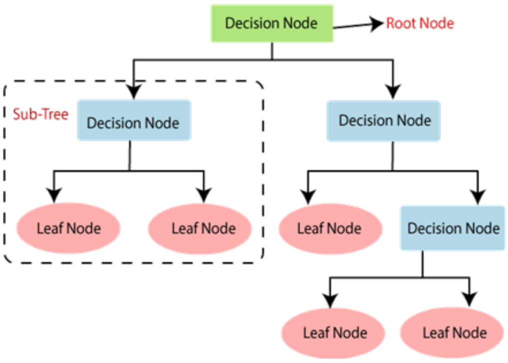


Fig: 4.3 Decision tree algorithms

iii). Random Forest Fig 4.4 shows below explain A Random Forest is an ensemble learning method used in machine learning that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Random Forest is an ensemble learning method used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or the mean

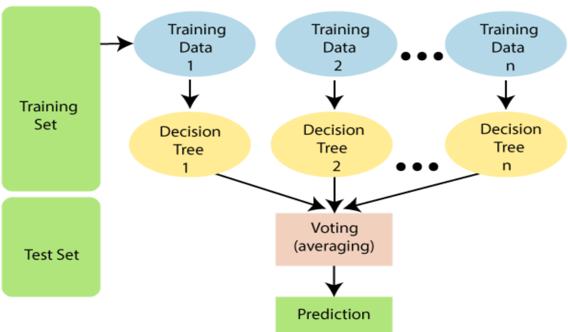


Fig: 4.4 Random Forest algorithms

prediction (regression) of the individual trees. Here is an explanation of the Random Forest algorithm: The below diagram explains the working of the Random Forest.

iv). Linear regression Machine Learning is a branch of Artificial intelligence that focuses on the development of algorithms and statistical models that can learn from and make predictions on data. **Linear regression** is also a type of machine-learning algorithm more specifically a **supervised machine-learning algorithm** that learns from the labelled datasets and maps the data points to the most optimized linear functions. which can be used for prediction on new datasets

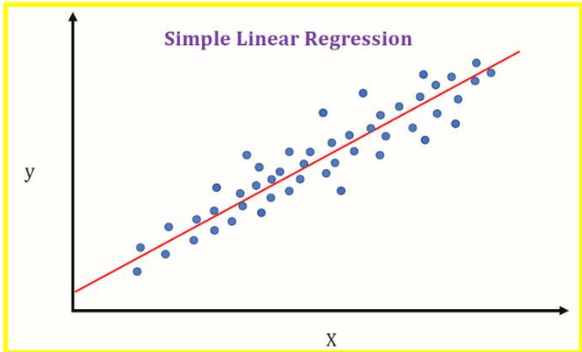


Fig 4.5 linear regression

v). Gradient boosting regressor Gradient boosting is a machine learning ensemble technique that combines the predictions of multiple weak learners, typically decision trees, sequentially. It aims to improve overall predictive performance by optimizing the model's weights based on the errors of previous iterations, gradually reducing prediction errors and enhancing the model's accuracy.

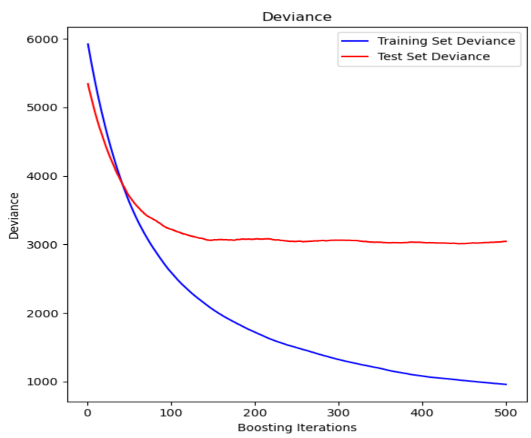
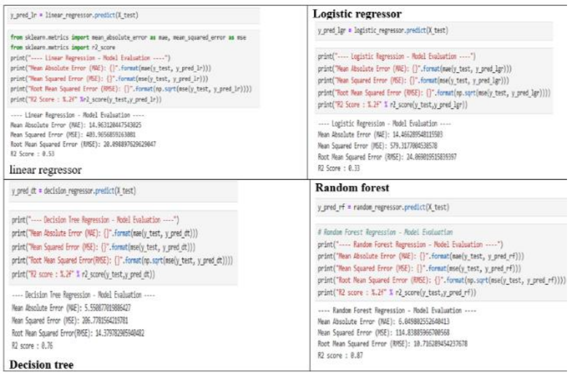


Fig 4.6 gradient boosting regressor

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

vi).Extra tree regressor an extra-trees



regressor. This class implements a meta estimator that fits several randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Read more in the User Guide. The number of trees in the forest.

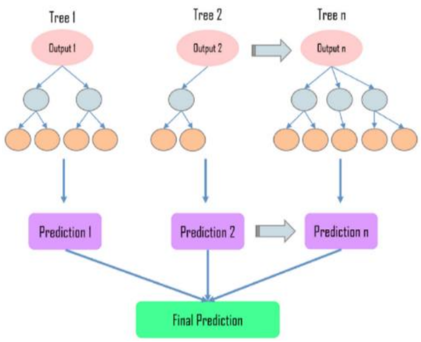


Fig 4.7 extra tree regressor

vii). XGB regressor Extreme Gradient Boosting (XGB Boost) is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm. Shortly after its development and initial release, XGB Boost became the go-to method and often the key component in winning solutions for a range of problems in machine learning competitions.

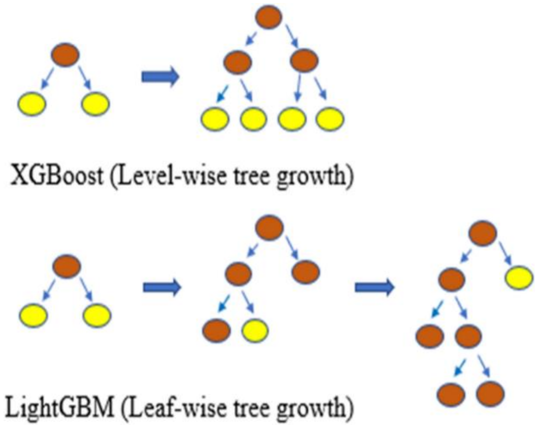


Fig 4.8 XGB regressor

5. RESULTS

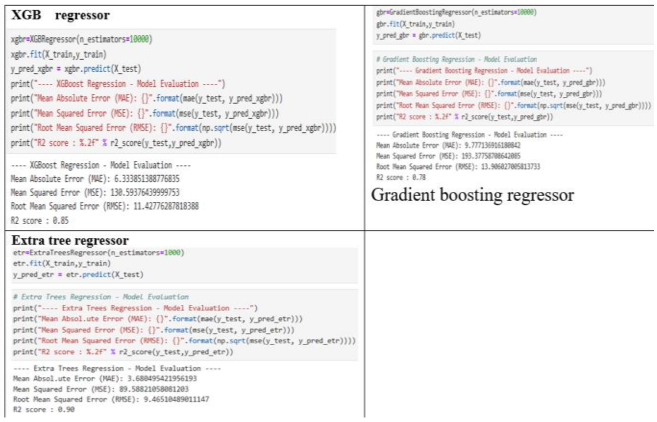


Fig 4.9 executed results comparison.

from the above it is compared that the seven machine algorithms are used with the code execution results.

S.NO	ALGORTHIM	ACCURACY
1	linear regression	0.53
2	logistic regression	0.33
3	decision tree	0.76
4	random forest	0.87
5	**Extra tree regressor**	**0.90**
6	Gradient boosting regressor	0.78
7	XGB regressor	0.85

Fig 4.10 accuracy comparison table

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

From the above comparison table it concluded that the extra tree regressor gives the best accuracy of 90%.

6. 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%.

7. FUTURE SCOPE

There are many opportunities for research and development in the broad and exciting field of IPL score prediction using machine learning. One area of interest is improving current models to include more variables and data sources, like team dynamics, player tiredness, and even sentiment analysis from social media sites for fans. Through the integration of these parameters, models are able to provide more accurate and forecasts, better reflecting the awareness of cricket matches. Additionally, there's a chance to create ensemble models that mix many methods and algorithms to improve the robustness and dependability of predictions. By combining the advantages of several models, ensemble approaches can mitigate the shortcomings of individual models and offer more thorough insights on match results.

8. 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/SMARTGENCON51891.2021.9645855.

[4]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.

[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 time-series 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.