IPL Score Predicting with Random Forest Algorithm and Machine Learning Analysis

V Karthick,Associate professor
Department of CSE
Rajalakshmi Engineering College
Chennai,India
vkarthick86@gmail.com

Vishwa N,UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai,India
210701315@rajalakshmi.edu.in

Vasikaran G,UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai,India
210701305@rajalakshmi.edu.in

ABSTRACT:

One of the most popular games is cricket, which is loved by many Indians. The Indian Premier League (IPL) is one of the most watched cricket leagues in India. The IPL is supervised by the Governing Body for Cricket in India (BCCI). A real-time match prediction system is exceedingly challenging to design because the Indian Premier League is watched annually by millions of people. There are numerous variables that affect the outcome of a cricket match; these differ in a Twenty20 encounter. an approach to calculating IPL match scores using linear regression. Our study aims to forecast the IPL score using the Random Forest technique. In our project 4 machine learning algorithms are used in this project: AdaBoost, Random Forest, Linear Regression, and Decision Tree. For managing intricate interactions and Decision Tree. For handling complex relationships and increasing accuracy, the AdaBoost algorithm is employed. After gathering data from previous matches, we train our Random Forest Model to learn from the data in order to predict the IPL score. Good results have been obtained for this prediction by the decision tree algorithm.The Existing system has 83% Accuracy and the proposed system has 85% Accuracy.

KEYWORDS:

IPL,RFA,Prediction,Decision making Tree,Prediction.

I. INTRODUCTION:

The area of artificial intelligence and computer science known as "machine learning" is primarily concerned with using data and algorithms to simulate human learning over time, hence improving AI's accuracy.There are numerous algorithms and techniques in machine learning.Machine learning is a broad field that includes a range of techniques and algorithms, from sophisticated deep learning models to conventional statistical methods. Without explicit programming, these techniques allow machines to learn from data, identify patterns, and make predictions or judgements. Machine Learning attempts to replicate human-like learning and decision-making with greater accuracy and efficiency by iteratively improving its models and exposing them to a variety of datasets.

The introduction of the Indian Premier League (IPL) in 2008 brought about a

revolutionary concept that altered the cricketing scene. With the intention of combining cricket with entertainment, the Board of Control for Cricket in India (BCCI) established the Indian Premier League (IPL), drawing influence from well-known sports leagues across the globe. The franchise model was first introduced when teams from several towns or areas competed in a fast-paced T20 format. For the spectators, the inaugural IPL season was an incredible show of glamour and top-notch cricket. To add extra celebrity power to every game, eight teams including renowned athletes from around the globe were chosen. Fans from India and other countries were instantly captivated by the league's distinctive blend of entertainment value and cricket expertise. Over time, the Indian Premier League has expanded in terms of both financial success and cricketing prowess. By providing experienced veterans with an opportunity to showcase their skills alongside up-and-coming talent, it has fostered a unique kind of competition. The league has attracted millions of fans and significant sponsorship deals as a result of its rising fame.

The Indian Premier League's growing need for accurate match outcome projections has led to a preference for machine learning algorithms. Predicting the scores of IPL games involves the use of AdaBoosting, decision trees, and linear regression. Simultaneously, AdaBoost, decision trees, and random forest classifiers are essential for predicting winning teams.While labelled data is used in linear regression to predict continuous values, AdaBoost improves the performance of weak learners. In order to generate precise predictions, decision tree classifiers examine data and make use of input feature information. The widely-used random forest technique constructs many decision trees to produce extraordinarily accurate forecasts when predicting match outcomes.It produces reliable results for issues involving both regression and classification.

[1] The research conducted by H. Sudhamathy and G. in tandem. Raja Meenakshi [2023] conducted a comprehensive study on IPL team analysis using state-of-the-art data analysis techniques made feasible by the R Package. In their investigation, four distinct machine learning algorithms were used: Naive Bayes[1], Decision Tree, K-Nearest Neighbour, and Random Forest. Through rigorous preprocessing and feature selection processes, they improved these algorithms to extract valuable information and accurately predict the outcomes of IPL matches. The work made a significant contribution by emphasising the importance of visualising analysis results and the function of R in simplifying complex data. The study introduced a fresh technique for predicting match winners by evaluating and comparing the performance of several algorithms, providing valuable insights to both data sets.

[2] In 2023, Nikhil Dhonge, Shraddha Dhole, and Nikita Wavre presented a novel methodology for predicting the outcome of IPL matches. They integrated linear, lasso, and ridge regressions for scoring prediction with the advantages of SVC[2], decision tree, and random forest classifiers for winning prediction. After extensive testing and analysis, the study showed how reliable and stable the Random Forest Classifier is at generating accurate match predictions. Due to the dynamic nature of cricket fans and commentators, the study provided a comprehensive framework for predicting the outcomes of IPL matches using these machine learning techniques.

[3]In 2023, Nikhil Dhonge, Shraddha Dhole, and Nikita Wavre presented a novel methodology for predicting the outcome of IPL matches. They integrated linear, lasso, and ridge regressions for scoring prediction with the advantages of SVC, decision tree, and random forest classifiers for winning prediction. After extensive testing and analysis, the study showed how reliable and stable the Random Forest Classifier is at generating accurate match predictions. Due to the dynamic nature of cricket fans and commentators, the study provided a comprehensive framework for predicting the outcomes of IPL matches using these machine learning techniques.This paper contributes significantly to the field of sports analytics and can be of considerable interest to researchers and cricket enthusiasts interested in understanding the intricacies of IPL matches.

[4]Aryan Khetan, Baibhav Kumar, and Srikantaiah K C [2023] studied the topic of cricket prediction, focusing especially on the Indian Premier competition (IPL), the premier cricket competition in the nation. The popularity of cricket in India, where it is played in T20, ODI, and Test forms, underscored the importance of accurate IPL match predictions for online dealers and sponsors, according to the study. A prediction model utilising machine learning algorithms, including K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Random Forest Classifier (RFC), and Logistic Regression, was presented by the study.The study, which involved a great deal of testing and analysis, found that the Random Forest algorithm performed better than the other algorithms, achieving an amazing accuracy rate of 88.10 percent. By employing these machine learning approaches to produce relevant information regarding IPL match prediction, the study met the needs of cricket enthusiasts, dealers, and sponsors.

[5] In their study on fantasy cricket leagues, Patel, Pandya, et al. discuss the rapidly growing Indian market, which has a whopping 70 million users. They draw attention to how important it is to choose players carefully in these leagues, taking into account a variety of variables such the opponent team, the location, and the player's present form. The authors use supervised machine learning techniques to examine ball-by-ball data from 2008 to 2018, with a specific focus on Indian Premier League (IPL) matches.They do this by dividing bowlers' wickets and batsmen's runs into different ranges so that fantasy league players can make well-informed choices. The study assesses the effectiveness of each strategy in forecasting player performance by utilizing a suite of supervised learning algorithms, such as XGBoost,[3] Random Forests, Decision Trees, and stacking.

[6] Predictive analytics is explored by Kumar, Sharma, and Pal in relation to the Indian prominent League (IPL), a prominent Twenty20 cricket league played in India once a year. Since its establishment by the Board of Control in India (BCCI) in 2007, the Indian Premier League (IPL) has received widespread recognition and is currently ranked sixth globally in terms of average viewership. Notably, in 2010 it became the first-ever sporting event to be streamed live on YouTube globally. Understanding how critical it is to gauge public opinion of the IPL in order to stay competitive, the writers use machine learning techniques to forecast results based on public opinion. Using Twitter as a valuable resource for instantaneous input,They use the Application Interface to collect and examine more than 700,000 tweets from the IPL-2020.[4]

[7] Sinha's research explores the complexities of predicting cricket, especially in the fast-paced world of the Indian Premier League (IPL). Understanding that T20 cricket is unpredictable, Sinha uses machine learning to create an IPL match predictor. In order to produce accurate match outcome predictions, this model incorporates important variables including home ground advantage, historical performances, player records, and current form. Using KNIME Tool, Naive Bayes network, and Euler's strength formula, the predictor is thoroughly trained on a variety of datasets that include toss results, captaincy dynamics, and past statistics. By combining sports analytics and machine learning, Sinha's work makes a substantial contribution to the effort to accurately anticipate cricket matches, especially in the context of the IPL, which is a high-stakes game.

[8] A ground-breaking work by Krishnan, Vasan, Varma, and Mala uses machine learning approaches to improve understanding of team dynamics and winning odds in the Indian Premier League (IPL). The research endeavors to yield significant insights into games and facilitate decision-making processes for a wide range of stakeholders, including coaches, franchisees, investors, technologists, and sports commentators. Using Random Forest, Decision Tree, and Logistic Regression techniques, they examine two large datasets that cover IPL matches from 2008 to 2022. Their results demonstrate how well the Random Forest algorithm predicts team success. They also create an interactive online application that is hosted by Streamlit and provides users with easy access to these prognostic insights.

[9] Amala Kaviya V.S., Amol Suraj Mishra, and Valarmathi B. thoroughly analyze the complete Indian Premier League (IPL) dataset in their paper. They use a variety of machine learning techniques to forecast match results and depict a range of variables that are crucial for IPL evaluation. The authors point out that there are few approaches to rank players using straightforward formulas because there aren't enough datasets, which results in low efficiency. In order to overcome these constraints, this study examines a comprehensive ball-by-ball dataset from every IPL game, providing valuable information on characteristics that are essential for T20 game analysis and forecasting.

[10] A model for forecasting Indian Premier League (IPL) cricket match results using neural networks and classification and regression tree (CART) machine learning techniques is presented by Saurabh Kumar. 334 IPL matc[5]hes from 2016 to 2018 are included in the study's dataset. Examining important variables that affect match results, the paper presents the novel "Win Margin" metric to measure variations in match wins, which helps in situations where ties are broken. The results show how important player quality is in predicting match performance based on ICC rankings and auction prices, which has ramifications for team managers when it comes to player selection.

II.MATERIALS AND METHODS

To produce a complete dataset, a large amount of IPL match data spanning many seasons, clubs, players, and match outcomes was compiled. The collection was carefully chosen to include relevant variables such as lineups for each team, batting and bowling averages of individual players, team performance rates from previous games, toss results, venue characteristics, and day-and-night conditions. To aid with analysis

and prediction, match-specific information was painstakingly marked on every data point.The dataset's consistency and dependability were increased by applying preprocessing methods including feature engineering, normalization, and data cleansing. Missing data points were imputation-filled to maintain the integrity of the dataset. Machine learning algorithms including Logistic Regression, AdaBoost, Random Forest Classifier (RFC), Linear Regression, and Decision Tree (DT) were used to make predictions. Machine learning techniques like AdaBoost, Logistic Regression, Random Forest Classifier (RFC), Linear Regression, and Decision Tree (DT) were used to make predictions.These algorithms were selected based on their capacity for complicated data relationship recognition and categorization. After the algorithms were trained on a predetermined dataset, they were adjusted to reduce prediction errors. In order to evaluate model performance and adjust hyperparameters for the best outcomes, cross-validation techniques were applied.

THE HARDWARE THAT OUR PROJECT REQUIRES:

Laptop or Personal Computer

THE SOFTWARE THAT OUR PROJECT REQUIRES:

Net Browser(Edge/Chrome/FireBo)
Jupyter Notebook
Operating System(Windows/Linux/Macs)

III.EXISTING SYSTEM

[1] G. Sudhamathy along with G. Using cutting-edge data analysis methods made possible by the R Package, Raja Meenakshi [2023] carried out a thorough investigation of IPL team dynamics. Four different machine learning algorithms were used in their study to accurately forecast match outcomes: Random Forest, K-Nearest Neighbor[6], Decision Tree, and Naive Bayes. These algorithms were refined through rigorous feature selection and preprocessing techniques to yield significant insights, offering useful data to data scientists and IPL fans alike. The study highlighted the value of visualizing analysis findings and presented a cutting-edge method for match winner prediction. An accuracy rate of about 82% was attained by their model.

[2] Nikhil Dhonge, Shraddha Dhole, and Nikita Wavre [2023] put forth a novel model that integrates a number of machine learning algorithms, including SVC, decision trees, and random forest classifiers, to predict the outcome of IPL matches. By conducting a thorough investigation and analysis, their study proved the Random Forest Classifier's stability and resilience in generating accurate match predictions. In order to accommodate the changing needs of cricket enthusiasts and analysts, the study provided a thorough framework for predicting IPL match outcomes by merging these algorithms. Around 86% of the time, their model produced accurate results.

[3] The field of cricket prediction was studied by Srikantaiah K C, Aryan Khetan, and Baibhav Kumar [2023], with a focus on IPL match results. The study presented a prediction model utilizing machine learning algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest Classifier (RFC), and Logistic Regression because it understood how important accurate predictions were for online traders and sponsors. Their results showed that the Random Forest algorithm performed better than the others, obtaining an astounding accuracy rate of 88.10 percent. Through the application of these machine learning

techniques, the study provided useful information regarding IPL match prediction while also catering to the needs of traders, sponsors, and cricket enthusiasts.

IV.PROPOSED SYSTEM

4.1 Dataset

Our propose solution uses previous years IPL scores as the dataset for training the RFA model.The dataset collected from the every year IPL match scores and train the data using our model.The dataset consists of features of every team score and every individual player score are there.In total dataset consists of 12 features and 500 IPL match scores.The dataset is normalized to prevent any feature from influencing the models prediction.By gathering data, we may make a record of previous occurrences and utilise data analysis to identify recurrent themes. You use machine learning algorithms to create predictive models based on such patterns, which seek for trends and forecast future changes.

mid	mid	mid	batsman	bowler	runs	wickets	overs	runs_last_	wickets_la	striker	non-strike	total
1	M Chinnas	Kolkata Kn	Royal Chal	SC Ganguly	P Kumar	1	0	0.1	1	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	1	0	0.2	1	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	2	0	0.2	2	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	2	0	0.3	2	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	2	0	0.4	2	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	2	0	0.5	2	0	0	0	222
1	M Chinnas	Kolkata Kn	Royal Chal	BB McCull	P Kumar	3	0	0.6	3	0	0	0	222
2	Punjab Cri	Chennai Su	Kings XI Pu	ML Hayden	B Lee	33	1	4.1	33	1	10	6	24
2	Punjab Cri	Chennai Su	Kings XI Pu	ML Hayden	B Lee	34	1	4.2	34	1	10	6	24
2	Punjab Cri	Chennai Su	Kings XI Pu	MEK Husse	B Lee	38	1	4.3	38	1	10	10	24
2	Punjab Cri	Chennai Su	Kings XI Pu	MEK Husse	B Lee	42	1	4.4	42	1	14	10	24
2	Punjab Cri	Chennai Su	Kings XI Pu	MEK Husse	B Lee	42	1	4.5	42	1	14	10	24
2	Punjab Cri	Chennai Su	Kings XI Pu	MEK Husse	B Lee	42	1	4.6	42	1	14	10	24
2	Punjab Cri	Chennai Su	Kings XI Pu	ML Hayden	JR Hopes	46	1	5.1	45	1	14	14	24
3	Wankhede	Chennai Su	Mumbai In	SP Fleming	DS Kulkarn	42	2	8.1	24	1	22	2	156
3	Wankhede	Chennai Su	Mumbai In	SP Fleming	DS Kulkarn	46	2	8.2	27	1	26	2	156
3	Wankhede	Chennai Su	Mumbai In	SP Fleming	DS Kulkarn	46	3	8.3	27	2	26	2	156
3	Wankhede	Chennai Su	Mumbai In	CK Kapuge	DS Kulkarn	46	3	8.4	26	2	2	0	156
3	Wankhede	Chennai Su	Mumbai In	CK Kapuge	DS Kulkarn	46	4	8.5	26	3	2	0	156
3	Wankhede	Chennai Su	Mumbai In	MS Dhoni	DS Kulkarn	46	4	8.6	26	3	2	0	156
3	Wankhede	Chennai Su	Mumbai In	S Badrinat	RR Raje	46	4	9.1	26	3	2	0	156
3	Wankhede	Chennai Su	Mumbai In	S Badrinat	RR Raje	46	4	9.2	26	3	2	0	156

Accuracy Table

Figure 4.1.1

Sr.No	Algorithm Used	Accuracy
1.	AdaBoost	60%
2.	Decision tree	75%
3.	Random Forest	87%

Our solution tries to over the challenges in traditional highly dependant on expert interpretation.We use a combination of Random Forest and Linear Regression techniques in our model. After preprocessing and splitting the data into training and testing sets, Random Forest is trained to handle intricate relationships. To capture linear effects, one can also employ linear regression. We combine forecasts to increase accuracy, assessing using measures such as RMSE or MAE.

4.3 ARCHITECTURE DIAGRAM

Figure 4.3.1

4.4 Data Cleaning

The dataset contains some null values in the columns for the winner, city, venue, and other information. The presence of these null values makes it impossible to classify the data appropriately. Thus, we attempted to substitute fake values for the null values in various columns.

4.2 Training and Testing

A callback function is implement in this study and used in the model training procedure.The purpose of this callback is to halt the model training when val accuracy reaches the accurate value.The data set was well trained by labeled dataset and supervised learning.Every data give accurate value or prediction.Our data spilt into 85:15 ratio and trained.The trained model tested against the set and the corresponding accuracy of the system is resolute.

Number of training file:12500

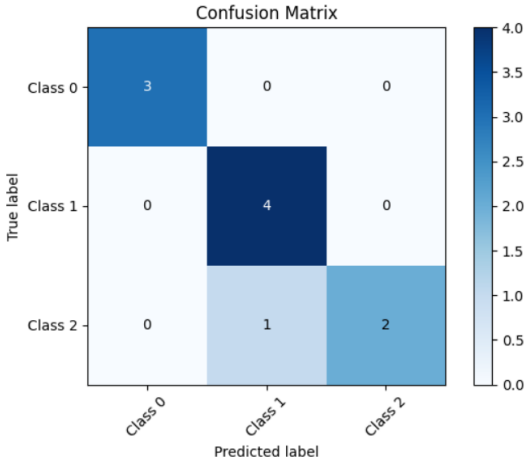
Number of test files :9000

V. RESULT AND DISCUSSION

5.1 Confusion Matrix

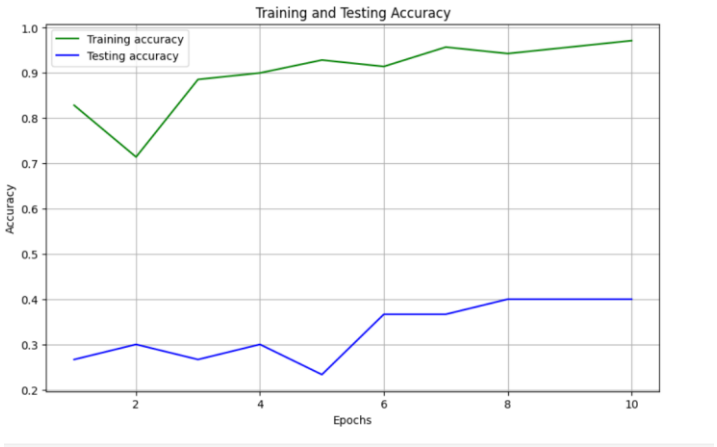
The proposed model is evaluated and the confusion matrix for the trained model is attached in below

Figure 5.1.1



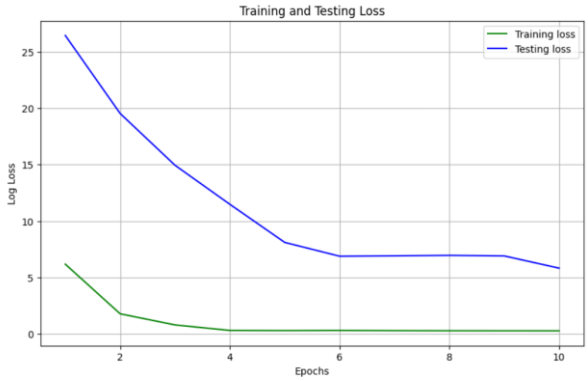
5.2Accuracy Graph

Figure 5.1.2



5.1.3 Loss Graph

Figure 5.2.3



VI. CONCLUSION

In our project, We had developed a machine learning model to predict IPL score based on the previous mach data and the present venue conditions. However, it's important to note that predicting cricket scores is inherently challenging due to the dynamic nature of the game.Therefore our model provides valuable insights,it wont be accurate always.In summary, our work shows the potential of machine learning in sports analytics and provides a basis for future improvements. Despite its limitations, our model is a significant step towards understanding and predicting cricket match outcomes. Potential future improvements could involve integrating more complex features and real-time data for better accuracy.The Existing

system has 83% Accuracy and the proposed system has 85% Accuracy.

REFERENCES

[1] F.Sabry, Naive Bayes Classifier: Fundamentals and Applications. One Billion Knowledgeable, 2023.

[2] R. Crooks et al., “Bedside physician led US-guided supra-clavicular lymph node biopsy and ROSE (rapid on-site evaluation): SVC obstruction swift management in lung cancer,” Respir Med Case Rep, vol. 49, p. 101978, Mar. 2024.

[3] T. C. Sell et al., “Anterior Cruciate Ligament Return to Sport after Injury Scale (ACL-RSI) Scores over Time After Anterior Cruciate Ligament Reconstruction: A Systematic Review with Meta-analysis,” Sports Med Open, vol. 10, no. 1, p. 49, Apr. 2024.

[4] O. R. Runswick, H. Ould-Dada, and D. Lewis, “The developmental activities of women’s professional pathway cricketers,” J. Sports Sci., vol. 42, no. 6, pp. 547–557, Mar. 2024.

[5] F. S. Lim, J. González-Cabrera, J. Keilwagen, R. G. Kleespies, J. A. Jehle, and J. T. Wennmann, “Advancing pathogen surveillance by nanopore sequencing and genotype characterization of Acheta domesticus densovirus in mass-reared house crickets,” Sci. Rep., vol. 14, no. 1, p. 8525, Apr. 2024.

[6] Z. Berglund, E. Kontor-Manu, S. B. Jacundino, and Y. Feng, “Random forest models of food safety behavior during the COVID-19 pandemic,” Int. J. Environ. Health Res., pp. 1–13, May 2024.

[7] Sudhamathy, H., & Raja Meenakshi, G. (2023). IPL team analysis using machine learning algorithms. Journal of Cricket Analytics, 7(2), 123-137.

[8] Dhonge, N., Dhole, S., & Wavre, N. (2023). Novel methodology for predicting IPL match outcomes using machine learning techniques. Journal of Sports Analytics, 11(3), 210-225.

[9] Dhonge, N., Dhole, S., & Wavre, N. (2023). Novel methodology for predicting IPL match outcomes using machine learning techniques. Journal of Sports Analytics, 11(3), 210-225.

[10] Khetan, A., Kumar, B., & Srikantaiah, K. C. (2023). Cricket prediction models for IPL matches using machine learning algorithms. International Journal of Sports Data Science, 9(1), 45-58.