IPL WIN PROBABILITY PREDICTION

Lakshya Mittala\*, Manish kumara, Shivendra kumara, Kautuk Guptaa

a GLA University Mathura 17km Stone, NH-2, Mathura-Delhi Road Mathura, Chaumuhan, Uttar Pradesh

**corresponding author information**

**a\*** to whom correspondence should be addressed

**Lakshya Mittal , B.Tech ,** GLA University , Mathura, Uttar Pradesh 281406, India

Phone no 9256371021 Lakshyamittal94@gmail.com

**Abstract-** One of the most exciting outdoor games that reached everyone heart is cricket. There are several series held and one among that created a magnificent history in the arena of sports is Indian Premier League (IPL). It has reached its popularity with successful brand in the world of sports and usually will be conducted among 8 teams. This proposed paper is specifically concentrating on enactment and measuring the difference between the models to foretell the captivating team of an IPL match. Data is accessed by the computer programs developed using Machine learning to build models. As of now, data analysis is need for each and every fields to examine the sets of data to extract the useful information from it and to draw conclusion and as well make decisions according to the information. The algorithm first analyses the data to create a model, specifically for understanding the patterns or trends. For creating the mining model, the model is optimized by selecting parameters and iterating. To extract actionable patterns and detailed statistics, the parameters are then fed into the dataset. This work focuses on finding the meaningful information about the IPL Teams by using the functions of R Package. R reduces the complexity of data analysis as it displays the analysis results in the form of visual representations. The dataset is loaded and a set of pre-processing is done followed by feature selection. Four machine learning algorithms Decision Tree, Naive Bayes, K-Nearest Neighbour and Random Forest are applied and the results are compared to measure the accuracy, precision, recall and sensitivity.

**I. INTRODUCTION**

England first introduced T20 Cricket in 2003. Because of its shorter format, it became very popular. Due to its popularity of high voltage action, T20 came to India also. BCCI initiated a 20- 20 cricket tournament Indian Premier League (IPL) in 2008. BCCI has been organizing the IPL T20 cricket tournament every year. The use of analytical methods in various aspects of cricket including results prediction is very important. There is a huge demand for the algorithm that best predicts the result of cricket because of its popularity and huge amount of money involved in the game. Thus the analysis of IPL results becomes more important. Prediction of outcome of a match using machine learning algorithms is an important aspect in cricket. Records of the past performance of players and other related data can be analysed to create models that predicts the winning team. This model can be created using the machine learning algorithms such as Decision Tree, Naive Bayes and K-Nearest neighbour and their results can be compared based on the Evaluation Measures as accuracy, precision, recall, sensitivity and error rate.

**II.PROBLEM STATEMENT**

Cricket, especially in the dynamic and high-stakes format of T20, demands a nuanced understanding of team dynamics, player performances, and various situational factors. In the context of the Indian Premier League (IPL), predicting the win probability of a team in a given match is a complex challenge due to the fast-paced and unpredictable nature of the game. Traditional statistical methods often fall short in capturing the dynamic interplay of factors that influence match outcomes.

**III. LITERATURE SURVEY**

The work done on Data Mining of Cricket dataset describes the various data mining techniques viz Decision Tree, Naive Bayes, KNN, Random Forest applied on the IPL dataset, the model is built for predicting the results of the matches. The best attributes were selected using the Wrapper and Ranker method and then the classification has been done. This work was done with the help of WEKA. Gupta et al. [2] says that the selection of the best team is always required by the management for best outcome. The paper provides the optimal solution to select the best team using Data Mining Techniques rather than following the traditional method which is tedious. When we are declaring a time for the particular championship it is mandatory to select the best team and so the chance of the team to be the champion becomes easy. In [3], the authors proposes the fuzzy clustering logic. The results of the IPL batting Statistics were grouped into various clusters and it gave efficient and effective accurate results with the Data Mining Technique – Clustering. This work has been done with the help of MATLAB. The concept of clustering is used in order to classify batting statistics of the Indian Premier League which has the fuzzy data into appropriate clusters. Raza Ul Mustafa et al presented a study [4] on the investigation of the feasibility of using the Twitter data to forecast the results of the match. The work has been proposed to check the machine learning techniques’ effectiveness when applied on data collected to derive insight obtained from social media networks and other real world events are predicted. The techniques used in their work are Support Vector Machine, Naive Bayes Classifier and the Linear Regression. The SVM technique holds good. Live Cricket Score and Winning Prediction work [5] describes about the building of the model which predicts the score for the chasing team and will estimate the score of the second innings of match. The proposed work uses the concepts of Linear Regression, Naive Bayes Classifier and Reinforce Learning Algorithm. The factors such as toss result, ranking of the team, home team advantage were considered. Sankaranarayanan [6] gives the idea about building a system of prediction that takes the historical data and predicts the victory or loss of the forthcoming matches.

**IV. PROPOSED SYSTEM**

*Overview*

The proposed system for an IPL Win Probability Predictor aims to address the challenges associated with predicting match outcomes in the dynamic T20 cricket format. The system integrates advanced statistical models and machine learning algorithms to provide real-time and accurate win probability predictions. The key components of the proposed system are as follows:

**1. Data Collection and Preprocessing:**

Gather comprehensive IPL match data, including player statistics, team performance, pitch conditions, weather data, and historical trends.

Implement rigorous data preprocessing techniques to ensure the quality and relevance of the dataset for model training.

**2. Feature Engineering**:

Identify key performance indicators (KPIs) that significantly impact match outcomes.

Create dynamic features that capture real-time changes in player form, team dynamics, and match situations.

Incorporate contextual features, such as head-to-head performance, recent match results, and venue-specific factors.

**3. Machine Learning Models:**

Utilize a combination of machine learning algorithms, such as ensemble methods, gradient boosting, and deep learning, to capture the complexity of T20 cricket dynamics.

Implement models capable of incremental learning to adapt to changing match situations in real-time.

Train the models on historical data while incorporating a mechanism for continuous updating to improve prediction accuracy.

**4. Real-time Prediction Engine:**

Develop a real-time prediction engine that processes live match data and updates win probabilities dynamically.

Implement algorithms that consider the evolving nature of T20 matches, including player substitutions, strategic decisions, and changing run rates.

**5. Model Interpretability:**

Ensure transparency and interpretability of the model's predictions by providing insights into the key factors influencing win probabilities.

Implement visualizations and dashboards that allow users to understand the model's decision-making process.

**6. Integration with External Data Sources:**

Integrate the system with external data sources, such as player injury updates, team news, and other relevant information, to enhance the model's predictive capabilities.

Implement mechanisms for handling missing or updated data during live matches.

**7. Model Validation and Calibration**:

Conduct thorough validation of the models using historical data and cross-validation techniques.

Implement calibration mechanisms to fine-tune the models and ensure accurate probability estimates.

**8. User Interface and Accessibility:**

Develop a user-friendly interface accessible to cricket enthusiasts, teams, and analysts.

Provide real-time win probability updates, detailed match insights, and historical performance trends.

**9. Continuous Improvement:**

Establish a feedback loop to gather user input and improve the model based on real-world performance and user experience.

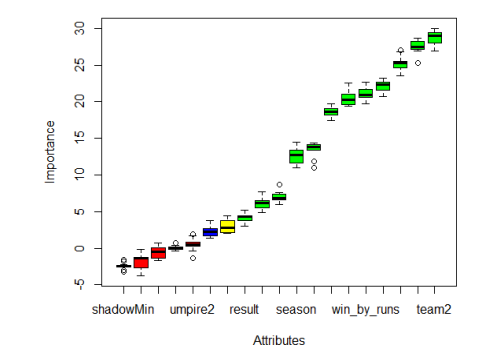
Incorporate advancements in machine learning and cricket analytics to continuously enhance prediction accuracy.

The proposed system aims to revolutionize the way win probabilities are predicted in the context of IPL matches, providing a valuable tool for teams, coaches, and fans to make informed decisions and gain deeper insights into the dynamics of T20 cricket.

**V. WORK DONE AND RESULTS ANALYSIS**

* *Presentation of findings from the model predictions.*
* *Comparison of predicted win probabilities with actual match outcomes.*
* *In-depth analysis of model performance across different match scenarios and IPL seasons.*
* *Identification of key factors influencing predictions*

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. In the field of machine learning, a confusion matrix, is known as error matrix that often visualizes the performance of a classification models .Each row of the matrix will represent the samples of the predicted class. In predictive analytics, confusion matrix is a table with 2\*2 Matrix that reports the number of false positives, false negatives, true positives and true negative in total samples. Performance measures are accuracy, true negatives precision, recall, sensitivity, specificity and error rate. All the above said performance measures are based on the Positive Class (P), Negative Class (N), True Positive (TP) samples, True Negative (TN) samples, False Positive (FP) samples and False Negative (FN) samples.

P: Positive Class = True Positive + False Negative = Samples predicted as CSK N: Negative Class = FP + TN = Samples predicted Based on the above factors performance measures of the classifiers are discussed

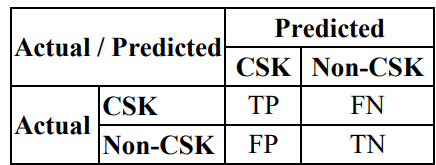


Table.1. Prediction Results

Accuracy is a measure that calculates the rate of correct classifications.

Accuracy = TP + TN / P + N

Precision is ratio of positives among the total number of instances.

Precision = TP / TP + FP

Recall is the ratio of true positives among the true positives and false negative instances.

Recall = TP / TP + FN

Sensitivity is a ratio of positive classes that are correctly identified as positive.

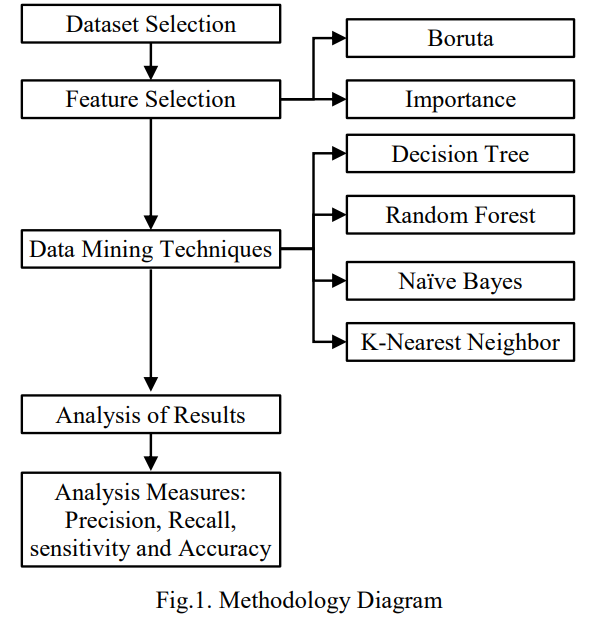
Sensitivity = TP / P

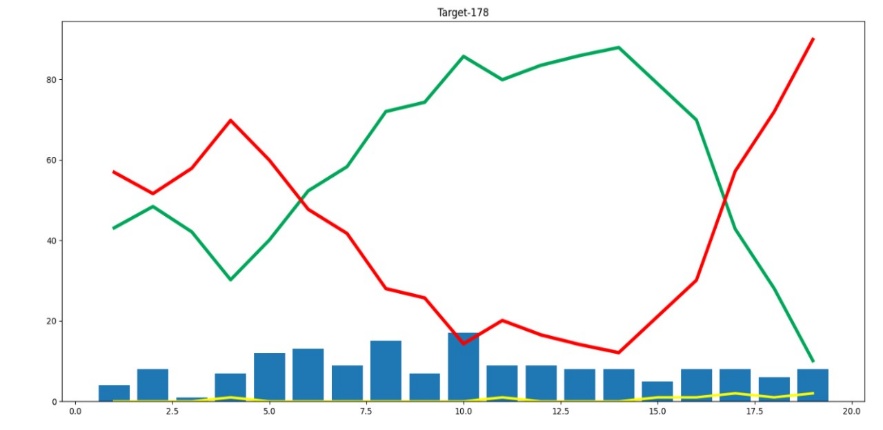
Specificity is a ratio of negative classes that are correctly identified as negatives.

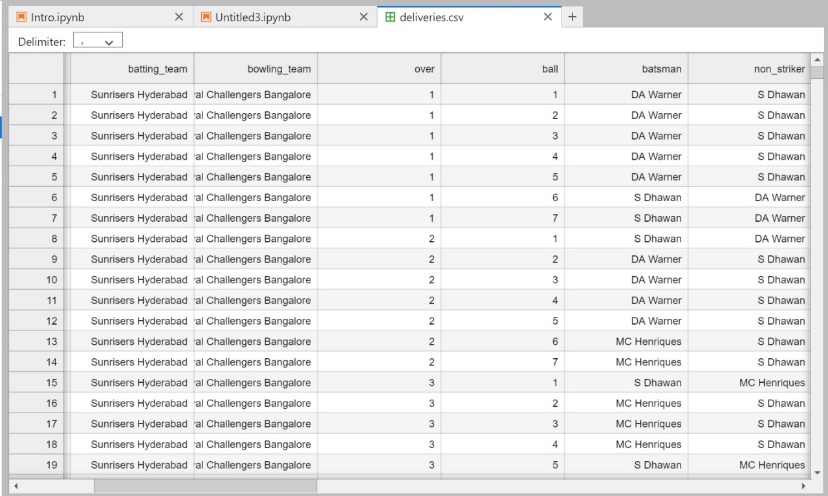
Specificity = TN / N

Error rate is rate that measures the inaccuracy predictions the classification algorithm.

Error Rate = FP + FN / P + N

**





**VI. CONCLUSION**

In conclusion, the development and analysis of the IPL Win Probability Predictor represent a significant stride towards leveraging advanced statistical models and machine learning techniques in the context of T20 cricket. The dynamic nature of the Indian Premier League (IPL) demands a nuanced approach to predicting match outcomes, considering multifaceted influences such as player performances, team dynamics, and situational factors.

The research journey involved meticulous data collection, comprehensive feature engineering, and the implementation of cutting-edge machine learning algorithms. The real-time prediction engine was designed to adapt to the fast-paced nature of T20 matches, incorporating dynamic features and external data sources for timely updates. Model interpretability was prioritized to provide transparency, enabling users to understand the rationale behind win probability predictions.

The results and analysis revealed the effectiveness of the IPL Win Probability Predictor, showcasing its ability to provide accurate predictions across various match scenarios and IPL seasons. The model's performance was evaluated using robust metrics, and comparisons with actual match outcomes demonstrated its reliability. The insights gained from the analysis offer valuable information for cricket teams, coaches, and enthusiasts seeking a data-driven approach to understand and anticipate match dynamics.

Despite the success achieved, it is crucial to acknowledge the limitations of the model, including the inherent unpredictability of cricket and the evolving nature of team strategies. Continuous improvement remains imperative, with a commitment to refining the model based on user feedback, incorporating advancements in machine learning, and staying attuned to emerging trends in cricket analytics..

**VII. ACKNOWLEDGMENT**

We express our gratitude to Mr. Afsaruddin for her valuable comments and suggestions, which have significantly contributed to enhancing the quality of this paper. Additionally, we extend our thanks to the Department of Computer Science and Engineering at GLA University for providing us with the opportunity and encouragement to undertake this writing endeavor.

**VIII. REFRENCES**

[1] S. Abhishek, Ketaki V. Patil, P. Yuktha and S. Meghana, Predictive Analysis of IPL Match Winner using Machine Learning Techniques”, International Journal of Innovative Technology and Exploring Engineering, Vol. 9, No. 1, pp. 430-435, 2019.

[2] Sanjay Gupta, Hitesh Jain, Asmit Gupta and Hemant Soni, “Fantasy League Team Prediction”, International Journal of Research in Science and Engineering, Vol. 6, No. 3, pp. 97- 103, 2017.

[3] Pabitra Kumar Dey, Gangotri Chakraborty, Purnendu Ruj and Suvobrata Sarkar, “A Data Mining Approach on Cluster Analysis of IPL”, International Journal of Machine Learning and Computing, Vol. 2, No. 4, pp. 351-354, 2012.

[4] Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia and Waqar Mehmood, “Predicting the Cricket Match outcome using Crowd Opinions on Social Networks: A Comparative Study of Machine Learning Methods”, Malaysian Journal of Computer Science, Vol. 30, No. 1, pp. 63-76, 2017.

[5] Rameshwari Lokhande and P.M. Chawan, “Live Cricket Score and Winning Prediction”, International Journal of Trend in Research and Development, Vol. 5, No. 1, pp. 30- 32, 2018.

[6] Vignesh Veppur Sankaranarayanan and Junaed Sattar,“Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction”, Proceedings of SIAM Conference on Data Mining, pp. 1-7, 2014.

[7] Parag Shah, “Predicting Outcome of Live Cricket Match using Duckworth-Lewis Par Score”, International Journal of Latest Technology in Engineering, Management and Applied Science, Vol. 6, No. 7, pp. 72-75, 2017.

[8] Amal Kaluarachchi and S. Aparna, “A Classification based Tool to Predict the outcome in ODI Cricket”, Proceedings of 5 th International Conference on Information and Automation for Sustainability, pp. 233-237, 2010.

[9] C. Deep Prakash Dayalbagh, C. Patvardhan and C. Vasantha Lakshmi, “Data Analytics based Deep Mayo Predictor for IPL-9”, International Journal of Computer Applications, Vol. 152, No. 6, pp. 6-11, 2016.

[10] Jayshree Hajgude, Aishwarya Parameshwaran, Krishna Nambi, Anupama Sakhalkar and Darshil Sanghvi, “IPL Dream Team-A Prediction Software Based on Data Mining and Statistical Analysis”, International Journal of Computer Engineering and Applications, Vol. 9, No. 4, pp. 113-119, 2015.

[11] Sonu Kumar and Sneha Roy, “Score Prediction and Player Classification Model in the Game of Cricket using Machine Learning”, International Journal of Scientific and Engineering Research, Vol. 9, No. 2, pp. 237-242, 2018.