International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 210-213 www.ijaem.net ISSN: 2395-5252

Predicting Cricket Outcomes: A Comparative Analysis of Machine Learning Models for Optimal Accuracy

Mayur Chavan, Aditya R. Vyawahare, Ansh R. Bhutada, Parth P. Ostwal, Pranav K. Nimbalkar

Asst. Professor , Dept. of Computer Engineering Pune Institute of Computer Technology Pune, India
Dept. of Computer Engineering Pune Institute of Computer Technology Pune, India
Dept. of Computer Engineering Pune Institute of Computer Technology Pune, India
Dept. of Computer Engineering , Pune Institute of Computer Technology Pune, India
Dept. of Computer Engineering, Pune Institute of Computer Technology Pune, India

Date of Submission: 05-02-2025 Date of Acceptance: 15-02-2025

ABSTRACT—The project will focus on the development of a predictive model of cricket match outcomes on the basis of high advanced machine learning techniques and highly massive data concerning player statistics and match history with conditions prevailing on game day to build an intelligent system capable of forecast more accurate results than ever produced before. The model will compare several machine learning algorithms, including Decision Trees, Random Forest, Logistic Regression, and Neural Networks, in terms of which is likely to be most ef fective for prediction purposes.Some key features involve dynamic datasets, where live data from match situations and performances are fed into the model to be in real time, keeping it as accurate as possible. Another unique feature is the ability to weigh various match conditions, including pitch conditions, weather, and team form, which it offers users for a better understanding of how results can depend on a number of factors. The fairness and reliability aspects for the system will be of prime importance to ensure no imbalance arises from any single model influencing the results.It is aimed at providing a new tool both for teams, analysts, and cricket fans that would offer an opportunity to take strategic decisions and enjoy the sport even more. Let the game of cricket become more transparent, data informed, and a little less behind the lines for a larger audience with the help of machine learning.

I. INTRODUCTION

It is an Indian Premier League league that has become to be of the top-most watched and followed cricket leagues around the world. It is really hard to predict what would result from a match given the fast-paced high-stakes nature of such a league, which encompasses a dynamic environment with thousands of influencing variables during the time of result. The methods currently employed for outcome prediction often settle on experts' opinions or very simple statistical models not well suited to capture the complexity in this game. Current interest is to utilize the computational model incorporating ML for better prediction accuracy on predictive modelling.

The paper surveys machine learning techniques applied to predicting IPL match results, centering on the array of models and methods applied to this domain. It explores the type of data used, which may range from player statistics and team performance to venue conditions, real-time match data, and points of discussion on how such features impact the predic tion process. It also reviews the most widely used machine learning models, including random forests, gradient boosting machines (GBM), and long short-term memory (LSTM) net works, challenges, and progress so far in model evaluation and real-time predictions. The purpose of this survey is to provide an overview of the current state-of-the-art in IPL outcome prediction and create space for further development in this area which may have massive application domains such as in sports analytics, fantasy sports, and decision-making by teams and analysts.

II. LITERATURE REVIEW

Now, with such complexity and vast data collection, it has become trendy to predict cricket

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 210-213 www.ijaem.net ISSN: 2395-5252

match outcomes with machine learning models. The approach has been followed through numerous research works and existing approaches to predict sports outcomes, including cricket. This literature survey reviews key contributions in this area to understand models, techniques, and datasets used along with challenges faced.

A. Historical Sports Outcome Prediction

Machine learning has been applied to sports outcome prediction for decades and earlier work focused on more traditional models such as linear regression and decision trees. Cricket being more dynamic and non-linear, weather factors, form of players, type of pitch, and in-match decisions led to seeking more complex models.

Other related studies include the one by Rao and Shah (2019) who tried to predict cricket match outcomes using decision trees, random forests, and logistic regression models. The above study gave a relatively good per formance of the random forest models in capturing the wide influences that determine the outcome of matches because its ability to handle both non linear relationships and an interaction between features. Sankaranarayanan et al. (2017) used logistic regression for predicting match outcomes by con sidering innings, scores, and wickets in hand. It shed some light on the performance of players and teams but was not able to treat non-linear data interactions quite well.

B. Data Collection and Feature Engineering

The biggest hurdle with any such prediction task related to the sports outcome is the collection of well-preprocessed data that encloses all the rel evant features. Cricket is one of the games with huge variables, like player performance statistics, team compositions, match conditions, and contextual data, like pitch condition, weather, etc. As noted by Bunker and Thabtah (2019): This presents a tremendous challenge for feature engineering because it involves both substantial struc tured historical data-like match scores and player stats-and unstructured data-like social media sentiments and weather reports-to enhance the prediction models.

Koul et al. (2019) managed to demonstrate the effect of feature selec tion. The removal of the most irrelevant features caused a highly significant improvement in predictive models by the machine learning model. They identified that contextual match features like venue, whether it is a toss related decision, and match format type (ODI, T20, or Test) provided an improvement in prediction results.

C. Machine Learning Models for Outcome Prediction

There are four types of such models-most notably: decision trees, ran dom forest, logistic regression, and neural networks which have been used extensively in sports analytics. Each model has some strengths and some weaknesses in the pre diction of match outcomes as follows:

Decision Trees: Such simple yet very interpretable models are effec tive when there are sharp decision-making criteria in terms of the conditions of the match played, along with player statistics. Decision trees tend to over f it, especially in cases of highdimensional data, as pointed out by Reddy et al. (2016). Random Forest: It is an ensemble of decision trees that avoids over f itting due to the fact that many trees are created and takes the majority vote for predicting the answers. As Kalpana et al. (2018) have stated, ”the random forest technique ensures higher accuracy than individual decision trees for cricket match prediction”.

Logistic Regression: Logistic regression is a good number to base predictions on in situations where one wants to classify something into two categories. In match outcomes, logistic regression has been applied as a function of the most important features including lost wickets and runs and batting partnerships, although that logistic regression mainly fails in han dling non-linear data relationships, as pointed out by Gupta et al. (2020).

Neural Networks: Deep learning approaches like neural net works have become very popular recently, considering the effectiveness with which it learns complex nonlinear inter relations, for the task of sports outcome prediction. Mishra et al. (2020) applied neural networks to cricket predic tion. The authors mentioned that although they achieved promising results in terms of accuracy, they demanded vast datasets and immense processing power.

D. Comparative Analysis of Models

There are many comparative studies that exist comparing the perfor manceof amachine learning model on cricket outcomes prediction. Chakraborty et al. (2021): The paper compares a random forest, logistic regression and a neural network with the metrics including accuracy, precision, recall, and F1-score. This conclusion drawn from the analysis gives evidence of ran dom forests having more accuracy but for the closer games, neural networks worked better in precision and recall and thus their output can be more advisable in predicting the possible outcome based on their current performance.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 02 Feb. 2025, pp: 210-213 www.ijaem.net ISSN: 2395-5252

Pandey and Kothari (2018) conducted an exhaustive study of machine learning models for the prediction of sports. Accord ing to their conclusion, no single model was found to perform best across all performance metrics. Where accuracy has been prioritized and transparency overrated, random forests do well as the powerful neural networks miss out on the transparency of the decision-making process.

III. PROPOSED SOLUTION

A. Data Collection and Preprocessing:

IPL match history, result of games played, team compo sition, player's batting and bowling performance, economical participation along with venue details. All this comes along with weather information and pitch reports. The last section of preprocessing the data would be handling missing values and outlier detection, handling feature normalization to generate consistency and reliability.

B. Feature Engineering:

This covers player statistics, such as form, for instance, runs scored or wickets taken, team strengths in batting and bowling combinations, the context of the match, such as overs, wickets, and run rate, as well as conditions particular to the venue, like type of pitch and weather. These are designed to capture the evolving nature of IPL matches that affects results in both pre-game and game phases.

C. Model Selection and Training:

Several machine learning approaches were explored, which include Random Forest, Gradient Boosting Machines (GBM), as well as Logistic Regression, to classify match outcomes in order to identify win/loss. A LSTM (Long Short-Term Memory) network is also used to enable learning of temporal
properties in match trending, with most analyses taken of the impact of mid-match events. Model performances are optimized by using grid search for hyperparameters

D. Model Evaluation:

Hence, extensive model evaluation is performed using sev eral metrics, such as accuracy, precision, recall, F1-score, and AUC. Also, in order to gauge the strength of the models and to prevent overfitting, K-fold cross-validation is performed. For determining the best predictor of IPL match outcomes, a comparison of the performances of the models is conducted.

E. Real-Time Prediction Framework:

A real-time prediction system was thus designed, updating the predictions on real-time match data (runs, wickets, overs). Dynamic match outcome forecasting for teams, analysts, and fans using this framework helps to entail better decision making for teams, analysts, and fans, while it forms a platform that is interactive for the user to track the progress of a match.

IV. CHALLENGES AND LIMITATIONS

Despite its advantages, the proposed system faces several challenges:

• Generally, data quality and availability may prove a challenge with missing or inconsistent data, which would reflect inaccurately in the real situation. Certain factors like injury to players and match day conditions will not be covered.

• The game being unpredictable, in most cases it becomes complicated to consider other factors like player form or a sudden change of weather that can impact matches in a huge way.

• The developed machine learning models might suffer with overfitting, model generalization, and biases in the history data, making it unable to make accurate predic tions regarding future matches.

V. CONCLUSION

The research into the potential of machine learning to predict the outcomes of IPL matches is relevant to modern sports analytics in light of the centrality of data-driven insights. It applies modeling with advanced techniques to historical match data and player statistics, which could enhance the accuracy of predictions about the outcome or provide more practical applications to teams, analysts, or fans in formulating strategy, informing fantasy league decisions, and increasing fan engagement within the IPL ecosystem.

REFERENCES

[1]. Rabindra Lamsal and Ayesha Choudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning", arXiv:1809.09813 [stat.AP] (September 2018)

[2]. Monoj Ishi(2022) "Winner Prediction in one day International cricket matches using Machine Learning Framework: an Ensemble Approach" ,Indian Journal of Computer Science and Engineering (IJCSE) Vol. 13 3 May-Jun 2022.

[3]. Renata Ntelia. Fortnite as Bildungsspiel? Battle Royale Games and Sacrificial Rites. In Researchgate articles (pp. 41-45).

[4]. Arjun Singhvi, Ashish Shenoy, Shruthi Racha and Srinivas Tunuguntla. "Prediction of the outcome of a Twenty-20 Cricket Match." (2015).
[5]. Swetha, Saravanan.KN, "Analysis on Attributes Deciding Cricket Winning", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03 — (March 2017)
[6]. Geddam Jaishankar Harshit, Rajkumar S, "A Review Paper on Cricket Predictions Using Various Machine Learning Algorithms and Compar isons among Them", International Journal for Research in Applied Sci ence Engineering Technology (IJRASET), IJRASET17099 (April 2018)
[7]. Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju, "Cricket score and winning prediction using data mining", International Journal of Advance Research and Development, Volume: 3 Issue: 3 (2018)
[8]. Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia, 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, Volume: 30(1) (2017)
[9]. A.N.Wickramasinghe, Roshan D.Yapa, "Cricket Match Outcome Pre diction Using Tweets and Prediction of the Man of the Match using So cial Network Analysis: Case Study Using IPL Data", International Con ference on Advances in ICT for Emerging Regions, ICTer: 442 (2018)
[10]. Ayush Kalla, Nihar Karle, Sushant Wagle, Sandeep Utala, "AutoPlay Cricket Score Predictor", International Journal of Engineering Science and Computing, Volume: 8 Issue: 4 (April 2018)