iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69906

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

IPL Win Probability Prediction System using Machine Learning Techniques

Syed Mohammad Zohaib[1], Nikhil Sharma[2], Rohit Singh[3], Ms.Sonia[4]
Department of Computer Science and Engineering Galgotias College of Engineering and Technology Greater Noida

Abstract: This paper describe a machine learning framework for the estimation of the winning probability of a team in the Indian Premier League (IPL) [1]. First, historical data of matches, player data, teams and environments is compiled and cleaned. Important variables are established with help of EDA. The machine learning methods based on logistic regression and gradient boosting were used to construct the model. The model was evaluated using various assessment methods such as accuracy, precision, recall, with promising results. In later studies focus will be put on the model improvement and introducing new features.

Keywords: Machine learning,IPL,winprediction,exploratory dataanalysis,gradientboosting,logistic regression

I. INTRODUCTION

A. Overview

Cricketisplayedinseveralformats:ODI,T20s, and Test Matches. The IPL is a Twenty20 tournament formed by the Board of Control for Cricket in India, known as BCCI, in order topromote cricketing and grow talent. Conductedevery year, the competition is among teams representing the cities of India and these teams are chosen through an auction, a first of its kind inIndian sports. IPL has emerged as a profitable business with massive fan following, team spirit,and city loyalty.

The match outcome is critical for the stakeholders, which is determined by the toss, the abilities of the players, and their performance on the day. Natural parameters like historical player data are the major contributors to the outcome. Large datasets are necessary for precise predictions, and hence it is challenging. The tool to be presented in this paper evaluates player performance using visualization based on IPL T20 statistics. Predictive models based on variables derived from batsmen, bowlers, and match data help decision- makers assess the strengths of teams and guide team selection and match strategies.

B. Problem Statement

The IPL is a game that's become very data-heavy in performance analytics. The Indian Premier League is a tough Twenty20 cricket league comprising teams that represent the cities in India. In light of its Traditional cricket analyses are solely dependent on expertopinions and some simple metrics that may be popularity, the apt predictability of match outcome for the management, analysts, and fans is really worth much. It's not that easy to predict win probabilities in IPLmatches since T20cricket isthemost unpredictable game. Player form, pitch conditions, weather, and how the toss goes bring about a lot of unpredictability. Today, without proper prediction tools, decision makingamongcoaches,teamowners,andevenfans becomes boggled. Traditional cricket analyses are solely dependent on expert opinions and some simple metrics that may be far from capturing the intricacies of interplay of all the influencing factors.Dataregarding cricket isalsodiverse, encompassing detailed statistical figures of players involved, different metrics of team performance besides specific data of situations that would culminate in match outcome. Thus, handling such a plethora ofdatabecomes impossible with simple analysis. The effort put in to surmount these challenges is the creation of asystem that predicts the win probability of IPL matches using a machine learning-based approach. Historical match data, player performance records, and match context would be applied to algorithms for logistic regression, and gradient boosting. Data visualization tools would also be included to make intuitive interpretation. The approach should enhance theaccuracy ofprediction, inform strategic decisions,andprovidesomeentertainment valuefor the stakeholders involved in the IPL, thereby changing how cricket outcomes are forecast.

C. Proposed Solution

The significance of this research lies in its potential to revolutionize the way IPL match outcomes are predicted. Using machine learning techniques, this project makes an attempt to break out fromtraditional statistical models to factor in a muchmore diversified view of what determines match outcomes, including not only statistical data from previous matches played and the performance of the individual players but also factors related to the weather, type of pitch, and immediate game situations [6].

©IJRASET: All Rights are Reserved

The suggestion will present insightful inputs for various stakeholders. For example, in terms of team management and coaches, it can give them win probabilities,which can guide these authorities on in game strategies while making critical pressure decisions. The system could benefit analysts and commentators by providing a very resourceful storyline on data and better-informed conclusions would explain the expected outcomes of matches. And for the viewer, of course, the expected game predictions from the model will bring more entertainment to the match, greater insight into the game, and enhance the involvement of the fans inthe sport.

It not only meant for IPL, but also wider in scope .The methodologies developed in this study could possibly be applied to transfer and refer to different leagues o f cricket like T20, ODI, Test cricket thereby taking cricket analytics to a greater extent. Given that today data-driven decision making is acquiring a lot more importance in sports, findings from this research might open up ways for novel innovations in understanding and forecasting sporting events.

This research applys machine learning to predict the results of Indian Premier League matches as an essential point in sports analytics. They are in tandem with the current methodologies, along with providing some available tools for them to work with-analysts and fans including one-another.

II. STUDY DESIGN

This study is about building a machine learning modelthat would anticipate the likelihood of an IPL team winning. The model processes the historical match data along with player performance, team statistics, and game conditions to analyze the information and then, through statisticalmethodsandmachine learning, totheprediction of future winners. Such factor of the game is more of the form of players, team dynamics, surface conditions, and toss outcome. The purpose is to improve the algorithm's efficiency and make a contribution to the management, analysts, and fans as well. In the end, this kind of technology will assist coaches and managers to make well-timed decisions during games by being able to accurately predict data driven applications that can be applied in practice for strategy development.

A. Study Settings:

A predictive model for the outcome of matches will be created using data gathered from IPL through online databases and API For achieving the relevant data in the form of wide-ranging information from multiple IPL seasons, the data set will include player statistics, team performance, and match conditions.Over the years, analyzing IPL data has made it possible to detect trends, track player form, and understand key factors influencing match outcomes. These insights contribute to a stronger and more representative prediction model that reflects the complexity of IPL matches. This comprehensive data set ensures that the model works reliably.

B. Sampling:

We will examine the data that is associated with the previous seasonsofIPL,whichholdsfulldetailsaboutallthematches [2]. This would be enough data to critically examine which variables influence the outcome of matches; therefore, the model will be created based on various, representative variables. Hence, based on this past data, we try to constructa more accurate and robust model to forecast the outcome of the IPL.

III. STUDY METHODS

Various machine learning algorithms, such as Logistic Regression will be utilized to make predictions.Testing and trainingwillbedoneusingtechniques likecross-validationto evaluate performance [5][7]. Hyperparameter tuning will also be used to improve model accuracy.

IV. DATA COLLECTION

We will be collecting data from sports API and match history databases. After collection, the data will be preprocessed and cleaned to eliminate any inconsistencies and prepare it for detailed analysis

A. Data Analysis:

We are going to use important parameters such as precision, accuracy, and recall in measuring the performance of the model and the ability of themodel in correctly predicting the result of the IPL match. By placing these results through rigorous testing, we can identify the model's errors and thus make appropriate adjustments so that it becomes correct..

Algorithm fine-tuning, parameter fine- tuning, and continuous model tweaking based on novel data that form part of it. The model will be more effective and efficient in making real-worldIPLpredictionsifitisupdatedandtunedperiodically to adapt to evolving match patterns.

B. Controls:

Thestudywilltakeintoconsiderationvariableslike pitchtypeandvenue,makingsuretheirinfluenceon match results is appropriately taken into account in order to increase forecast reliability.

C. Variables:

Match results (win or lose), team statistics, player performance indicators, venue specifics, and match conditions are the main variables in this study [3]. To improve prediction accuracy,other factors including past head-to-headrecords and recent team performance will alsobe taken into account.

V. IMPLEMENTATION

A. Tools and Technologies

1) Numpy:

NumPy is used in the IPL probability of winning prediction system to effectively manage large numerical data such as player and team statistics. It supports fast mathematical operations that are very important while training machine learning models.

2) Pandas:

Pandas is most fitting for structured tabular data. It organizes and cleans IPL match data, handles missingdata values, transforms variables, facilitates feature engineering and data exploration, making insights readily available to be put to predictive modeling needs. Altogether, itstreamlines theworkflow ofdata processing and modeling.

3) Logistic Regression:

It is used in the IPL win probability prediction system to predict win probabilities by modeling the relationship between input features (like runs or wickets) and the outcome using the logistic function. It outputs probabilitiesbetweenOand1,efficientlyidentifyinghow each feature influences the result. It's fast, interpretable, and ideal for binary classification problems like win/loss predictions [4].

4) Pickle:

Pickle is a Python library that serializes and deserializes Python objects, such asmachine learning models, so that they can be saved to a file and loaded back into memory later.

5) Streamlit:

Streamlit is an open-source Python library used to create interactive web applications for data science andmachine learning projects with minimal coding effort. It allows us to tum data analysis scripts and machine learning models into shareable web apps.

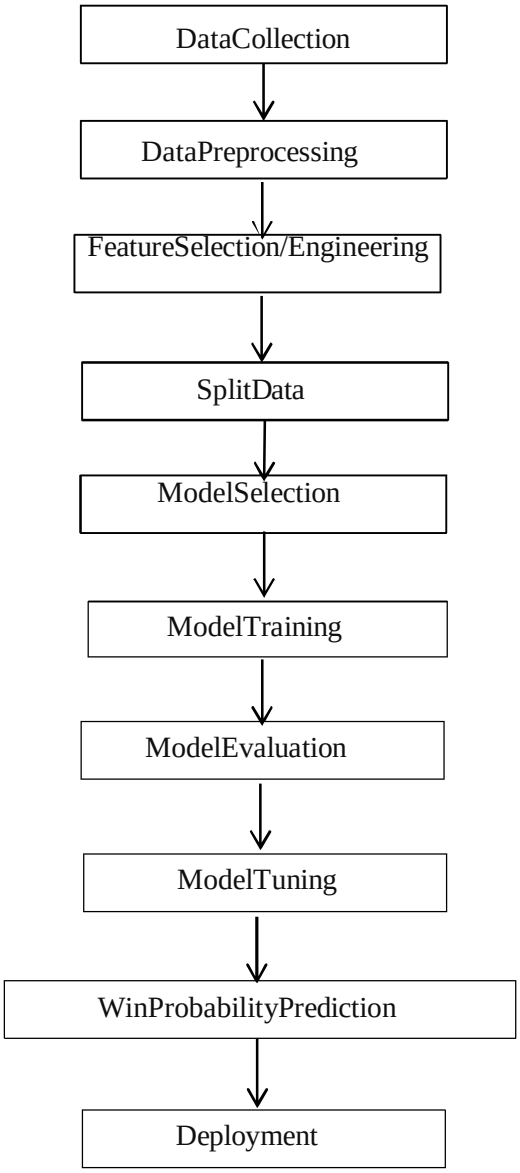
6) Heroku:

Heroku is a cloud platform that simplifies thedeployment and hosting of web applications, including machine learning models like an IPL win probability predictor. It allows developers to focus on building their app without worrying about server management or infrastructure.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

VI. FLOW CHART FOR THE WORK FLOW



VII. RESULTS AND ANALYSIS

The IPL Win Probability Prediction System applies Logistic Regression in calculating the Win Probability based onhistorical data, with the important factors of the matchbeingthebatting team,bowling team,wicketslost, and the run rate at that time as well as the number ofballs delivered at that time. It is effective at predicting win or lose conditions by providing the probability based on these inputs..

Based on past IPL match historical trends, the model predicts the impact of different factors on the match. Predictions are constantly updated as the game is underway, using new match data and decreasing win margins. The updates are based on trends learned during trainingandcurrent-timefactorssuchaspitchandground, making the predictions applicable to the current match.

The system also takes into account player performance, pitch location, and pitch type along with enhancing the accuracy of the predictions. The system predicts the outcome accuratelyandin timebased on historical trends and live match statistics. This assists the analysts, coaches, and fans to comprehend how the game statistics like the scoring rate and wicket falls influence a team's likelihood of winning.

©IJRASET: All Rights are Reserved

Thisapproachnotonlyimprovesstrategicdecisionmaking for teams but also keeps fans engaged, makingIPL matches more interactive and exciting.

VIII. CONCLUSION

In brief, the probability-based system for identifying IPL match winners employs advanced machine learning, robust data processing, and user interfaces to offer real-time match result prediction for the Indian Premier League (IPL). The system employs various technologies in a seamless manner to offer an end-to-end system. The system starts with NumPy and Pandas, two robust Python libraries used to process data as well as big data. The libraries offer data cleaning, manipulation, and preparation of required match data like players' performance data, team members, and match histories more efficiently. The system employs the numerical computation functionality of NumPy and the tabular data management functionality of Pandas to prepare allthe requiredfeaturesforprocessing in-depthandreadyto be provided to offer to provide for robust analysis.

This provides a necessary data base from which predictive models can be developed to provide meaningful and valid predictions.

The system's core is that it uses a LogisticRegression model, one of the most widely used machine learning algorithms for binary classification tasks such as finding the win or loss of a team. The model is dependent on the concept of learning from historical data through identifying patterns among different features such as runs, wickets, overs, and teamperformance thataffecttheoutcomeofmatches. Pickle is used to save and load the trained modelonce it is trained, so it does not have to be retrained every time a prediction needs to be made. This is a very essentialstep to optimizethe processandget the model ready for real-time predictions in fast- moving environments such as live match analysis. It serializes the model with Pickle. The system will make sure, by serializing the model with Pickle, predictions are made very efficiently without unnecessary calculations by saving time and computational power.

The system will be deployable on the cloud based Heroku, offering scalable hosting and simple deployment of web applications. With Heroku, the systemcanbemanagedeffectively todifferent_levels of user traffic without having to compromise on performanceand respond during peak usage.Heroku offers simple deployment with the complexities of server management and infrastructure being handled, thus enabling developers to concentrate on enhancing the prediction model and user interface. Streamlit is also employed to develop an interactive and user- friendly interface that enables the users to input real time match data like runs, wickets, and overs and get instant predictions of the win probability. Streamlit is best suited to develop web applications that are visually stunning and highly functional, offeringusers a smooth experience

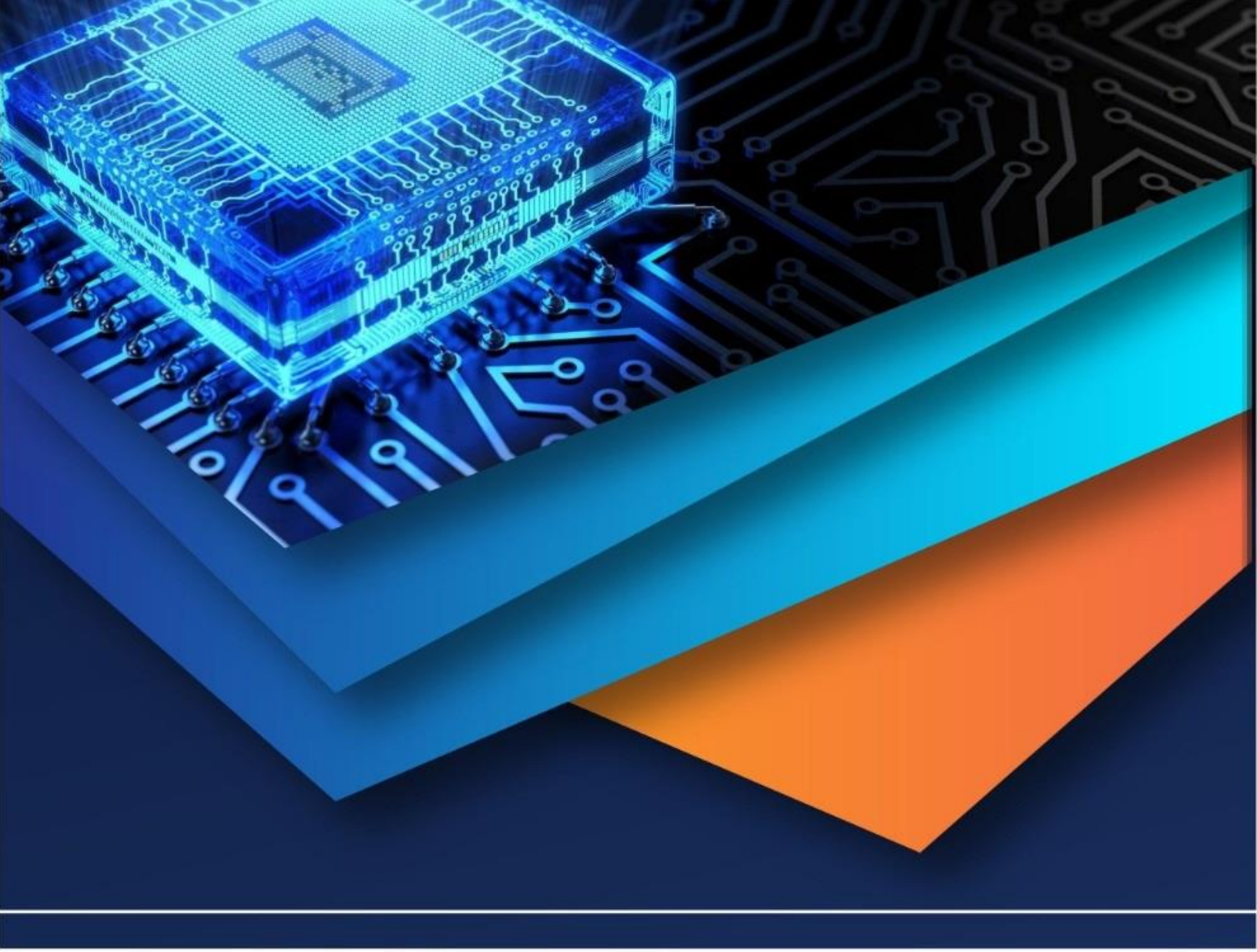
By including these new technologies, the IPL match win prediction system is an effective, scalable, and robust match prediction solution. It gives intelligent information about the match dynamics, offering agreat opportunity for viewers and analysts to examine and understand the forthcoming trends of an IPL match.

The system not only illustrates how data science and machine learning are applied in sports analytics, but it also shows how technology can improve the pleasureof watching cricket. The technology is a state-of-the- art platform that brings real-timedata-driven insightsto the world of IPL cricket, whether it is to help clubs make strategic decisions or to interact with fans more deeply.

REFERENCES

[1] Daniel Mago Vistro, Faizan Rasheed, Leo Ger trude David, "The Cricket Winner Prediction With A pplication of Machine Leaming And Data Analytics" International Journal of Scientific & Technology Research (2019)
[2] MadanGopalJhanwarandVikramPudi,"Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach" International Institution of Information Technology (2017)
[3] I.P.Wickramasingheet.al,"Predictingtheperformanceofbatsmenintestcricket,"JoumalofHumanSport&Exercise",vol.9,no.4,pp.(2017)
[4] R. P. Schumaker, 0. K. Solieman and H. Chen, " Predictive Modeling for Sports and Gaming" in Spor ts Data Mining, vol. 26, Boston, Massac husetts: Springer, (2016)
[5] J.McCullagh,"DataMininginSport:ANeuralN etwork Approach," International Journal of Sports Science and Engineering,vol. 4, no. 3 (2016)
[6] Bunker, Rory &Thabtah, Fadi. "A Machine Leami ng Framework for Sport Result Prediction. Applied C omputing and Informatics". (2017)
[7] Kulkarni,V.& Sinha,P., n.d.EffectiveLeamingand Classification using Random Forest Algorithm. International Journal ofEngineering andInnovative Technology (IJEIT).
[8] Lokhande, A., Chawan , R. &. &Pramila&, S., 2018. Prediction of Live Cricket Score andWinning. Computer and IT Dept, VeermataJeejabai Technological Institute, Mumbai,India, 5(4)(2394 -9 333)

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)