PREDICTIVE ANALYSIS OF CRICKET TOSS BY USING MACHINE LEARNING

[1]B.SUJATHA, [2]K.MAHESH BABU, [3]CH.PARVATHINADH CHOWDARY, [4]G.UDAYA KEERTHI, [5]K.GOPI CHAND

[1]Assistant Professor, [2,3,4,5]UG Student
Department of Compute Science and Engineering, NRI Institute of Technology, Visadala (V), Medikonduru (M), Guntur, Andhra

Abstract: **A single over can completely alter the course of a cricket match, especially in the Twenty-20 format. The IPL is watched by millions of fans, making it a real-world challenge to create a model that can forecast the results of its games. Finally, a dataset was built to forecast each team's likelihood of winning and losing as well as its success as a team. Other variables that affect the team's performance include the outcome of the coin toss, the location of the game, the city, the pitch's condition, and the weather. The outcome of the coin toss has a big impact on how successful a team is. Therefore, by applying a machine learning algorithm and utilizing the outcome of the coin toss as a criterion, we can forecast the likelihood that each side will win. We are using Navies Bayes to train the model for this.**

Keywords**: neural network, multivariate regression, neural network, supervised learning, naive bayes classification, cricket prediction applications**

I. INTRODUCTION

A division of artificial intelligence called machine learning seeks to address actual technical issues. It provides the chance to learn without relying solely on programming and is focused on learning from data. We could not even notice it because it is used so frequently every day. Machine learning has the advantage of using mathematical models, heuristic learning, information gathering, and decision-making trees. As a result, it offers durability, control, and recognition.

Cricket is played in a variety of formats, including T20, One Day International, and Test Matches. The IPL, also known as the Indian Premier Tournament, is a 20-20 cricket league renowned for supporting cricket in India and hence inspiring young, talented players. Every year, the league is held. IPL teams are chosen to represent various Indian cities through an auction. Sports have long used player auctions. The first time a squad in India was chosen from an existing group of players through a player auction was, however, in the IPL. The outcomes of these matches are important for the stakeholders because of the large fan bases, strong sense of team, and financial investment. This in turn depends on the game's regulations, specifically the toss.

Cricket can be played in a number of different ways, including T20, which in turn depends on the game's rules, the team that wins the toss, the players' skill, and how well they perform on game day. Predicting the outcome of a cricket match involves a number of contextual elements, including player performance history. Predicting the outcome of a cricket match involves a number of contextual elements, including player performance history. A system that forecasts the results of games between various teams can help with team selection. However, it is extremely impossible to anticipate the exact result of the match due to the multiple aspects at play. Additionally, the size of the data affects how accurate the forecast is.

We processed the data and offered recommendations for this project study using a data analysis tool called Google Colab. Improved models can help decision-makers during cricket matches to compare a team's strengths to the other teams and environmental circumstances. Based on data collection, we intend to contribute to the planned project in the following areas:

- Offer player statistics analysis based on many variables.
- Predict team performance using player-specific statistics.
- Accurately estimate cricket game outcomes.
- Accurately forecast environmental variables influencing the cricket league.

II. LITERATURE REVIEW

The most crucial stage of the software development process is the literature review. The time factor, economics, and company strength must all be assessed before the tool is developed. Determine which operating system and language can be utilized to construct the tool in the following ten steps when these requirements have been met. As soon as they begin creating the tool, programmers require a lot of outside assistance. Senior programmers, books, and websites are good sources of this support. The aforementioned factors were taken into account before the system was built to design the suggested system.ODI cricket matches are forecasted using the Google API in the Existing System. A black box prediction method is Google API. Since it involves supervised learning, a lot of data must be provided in order to train the models. When making numerical predictions, the Google Prediction APIs use regression algorithms. And Classifiers when the application content allows for only a small number of possible values to be assumed for the target output, such as strings or numbers. with Only relevant data can be accounted for

using this API. A accurate probability curve cannot be formed if the attributes are unrelated to one another. Applying a CSV file of prior cricket matches, the necessary data can be extracted and the model may be trained by using the right queries.

Predicting outcome of the game has recognized some fundamental problem. In the existing method, by extensive literature survey many research papers have researched on this topic but most of them have used primitive machine learning algorithms like Naïve Bayes and logistic regression. They have taken into consideration factors like teams, toss winners, winners by runs and winners by wickets. We aim to develop an effective machine learning tool which is needed to predict the outcome of a game, taking humidity and wind speed into consideration. In the current situation, many franchise owners have lost money on the negative prediction results of players. Also, having gone through the works of different authors and researchers, a keen interest in the wide applications of Machine Learning algorithms has been seen. The significant reason for the application of random forest classifier and adaboost, is because these algorithm can take a large amount of input features, train the decision trees and stumps and can give a very accurate prediction. The old machine learning techniques like Naïve Bayes and Logistic Regression have run out of date and a little change in the application needs rebuilding the whole prediction system, according to the stakeholder's needs. With the advancements of parallel processing and increase in computational capacities the Machine Learning algorithms, tend to out per the older systems and hence are significant motivation for our work.

III. PROPOSED FRAME WORK

In sports, most of the prediction job is done using regression or classification tasks, both of which come under supervised learning. In simple terms, y = f(x) is a prediction model which is learned by the learning algorithm from a set of dataset: D = ((X1,y1),(X2,y2), (X3,y3), ... (Xn,yn)). Based on the type of output (y) supervised learning is divided further into two categories, viz., regression, and classification. Logistic Regression deals with predicting categorical values; however, classification deals with discrete kind of output. For predicting categorical values, Logistic Regression appeared to be quite effective, and for classification problems like predicting the outcome of matches or classifying players, learning algorithms like Naive Bayes, Logistic Regression, support vector machine, Decision Tree, K-Nearest Neighbor, Xg-Boost, Random Forests, are used.

1. This suggested approach uses an online data set of IPL matches from 2008 to 2019 that was collected.

2. This suggested work focuses on other characteristics including the match's location, city, etc. while also predicting the likelihood that toss winners will win the match.

3. To gain improved accuracy and a better model, the proposed work has chosen a variety of models. The Training model is tested on 2019 IPL matches for prediction.

Predicting the result of the game has shown several fundamental issues. Numerous studies have been conducted on this subject using the current methodology; however the majority of them employed crude machine learning methods like Naive Bayes and logistic regression. Teams toss winners, winners by runs, and winners by wickets are among the considerations they have made. Our goal is to create an efficient machine learning technology that will take humidity and wind speed into account when predicting a game's outcome. Numerous franchise owners have suffered financial losses as a result of players' poor projection performances in the current situation.

A significant interest in the varied applications of Machine Learning algorithms has also been observed after reading through the works of various authors and academics. Random forest classifier and Adaboost are widely used due of their ability to process enormous amounts of input information, train decision trees and stumps, and provide extremely accurate predictions.

The outdated machine learning methods, such as Naive Bayes and Logistic Regression, need to be replaced because even minor application changes necessitate rewriting the entire prediction system to meet stakeholder needs.

Machine learning algorithms tend to perform better than older systems due to parallel processing breakthroughs and increases in computational power, which is a major driving force behind our research.

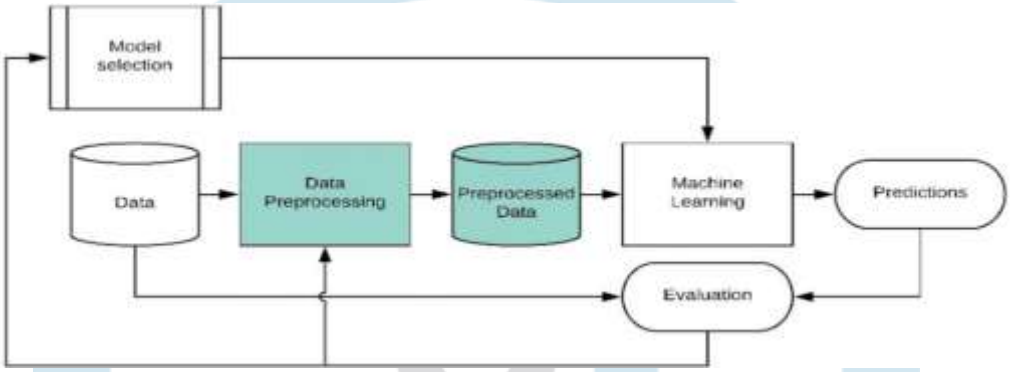
A significant interest in the varied applications of Machine Learning algorithms has also been observed after reading through the works of various authors and academics. The primary justification for using random Because forest classifier and adaboost can train models using a large number of input features, A fairly accurate prediction can be made using decision trees and stumps.

The outdated and sometimes ineffective machine learning methods like Naive Bayes and Logistic Regression application requirements have changed redesigning the entire prediction system in response to stakeholder requirements. Machine learning has advanced with parallel processing and increased computer power. Algorithms, which typically perform better than previous systems, are a major source of inspiration for our work.

IV. SYSTEM DESIGN

The purpose of this project research is to create data analysis tools to process cricket data with promising results. This project tutorial analyzes all the parameters used in the game namely player statistics, team statistics, environmental features to provide an effective solution to make the game more fun and keep the fun maintained. Cricket began in the 16th century in England. Cricket is a game with many formats, different levels of play and different lengths. The Twenty20 is one of three current types of cricket known to the International Cricket Council (ICC). In that format, two teams with one innings each team has more than 20 overs. Due to the short duration and the excitement, it is productive, twenty-two cricket has been so successful.

There are many competitions held annually at the local and international level. There is a strong commercial interest in predicting player performance in cricket leagues. This has encouraged a lot of analysis of individual and team performance, as well as predictions of upcoming games, across game formats. This project research aims to improve the application of accurate prediction in the game of cricket using machine learning about the game, the environment, and the players. We have proposed a system that overcomes the major weaknesses of time-consuming and labor-intensive work to keep individual player records and statistics.



In the proposed system the user can upload historical data collected from a machine learning tool such as Pycharm or Turkish Journal of Computer and Mathematics Education Vol.12 No.6 (2021), 5111-5124 Research Article 5115 Jupiter or Google Colab and perform data analytics to obtain the result. In data analytics we use the random forest classifier algorithm for building the prediction model defining conditions to process and produce a result.

Machine Learning (ML) has brought about a positive change in most of the fields. It can also be applied in sports like cricket.ML can improve the performance and accuracy of players and develop better strategies for the upcoming games. This can be done by predicting the runs scored by a player or the team, the wickets that can be taken and finally predicting the final result of the match. It is always important to select the correct variables so that the prediction is accurate. In the post we will focus on applying ML techniques to the dataset in R. As we are dealing the batting dataset of Virat Kohli and MS Dhoni, we will be dealing with the following variables:

- Runs which indicates the runs scored by the player.
- BF which indicates the balls faced by the player.
- Mins which indicates the minutes spent at the crease by the player.

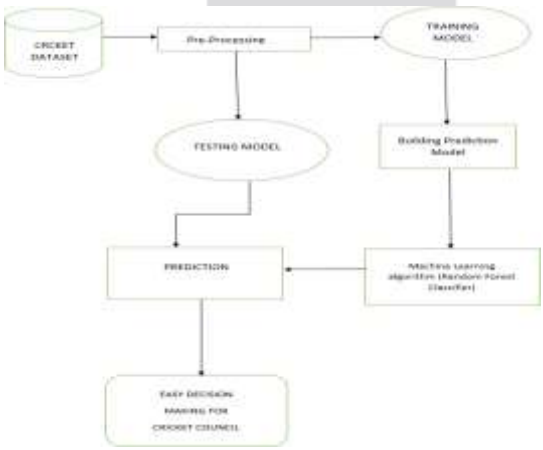


Fig 1: Architecture Diagram of the prediction Model

V. RESULTS AND DISCUSSIONS

The Software Environment:Colab is a free cloud-based writing platform. Colab can access electronic libraries that can be downloaded to your notebook. It is ready for everything from improving your Python coding skills to working with deep libraries, such as Py Torch, Cameras, Tensor Flow, and Open CV.

- Zero configuration is required
- Free access to GPUs
- Easy sharing

Google has been fast paced in AI research. Tens or Flow, which is an AI framework and Colaboratory, which is a development tool were made by google. Tens or Flow is an open-source software and Colaboratory is meant for public use for free and is otherwise called Google Colabor just Colab. One more feature offered by Google is that developers can make use of the GPU. To make its software a standard in machine learning and data science and also to build a broad customer base for Google Cloud APIs could be some of the reasons why google made its software free to the public For no apparent reason, there seems to be a reduction in the learning as well as the development of machine learning applications after colab was introduced. Colab is Jupyter's free note-taking site that works perfectly in the cloud. You do not need setup and your created notebooks are editable simultaneously by your team members in the same way documents are edited in Google Docs.

Python is a translated, high-quality programming language, with a common purpose. Python makes sure that code is readable with its remarkable utilization of white spaces. Its object oriented and language building method tries to assisted it or sin writing clear, logical code for small and large projects. Python is a programming language for multiple paradigms. Object-oriented programs and systematic programs, functional programming and interactive programs (including Meta programming and met objects) are supported. Contract programming and logic programming are also supported by the use of extensions. Powerful typing and a combination of reference counts and a garbage collector that takes care of memory management are used. It also incorporates dynamic word flexibility (late binding), which includes method and word changes during application. Python attempts to acquire syntax and simple, low-computer programming language while giving developers the opportunity to choose their own writing style. Python adopted "there must be one way-and probably the only one-to make it clear" philosophy for design.

The developers of Python avoid premature execution, and exclude layers in less important components of C Python reference that can provide a small increase in speed at a clear cost. With speed as the crucial factor, the Python program developer can submit time bound tasks by adding modules that are written in C-language languages, or by making use of Py, a timely compiler. Python can be used to convert the Python script into Can direct C-level API calls to Python interpreter.

RESULTS

With the application of Random Forest Classifier algorithm in our prediction model we have achieved an accuracy of98.14%for the training model and an accuracy of 89.47%forthe testing model. In contrast to Random Forest, when we applied Multinomial Logistic Regression for predicting the match outcome, the accuracy of training model was found to be 29.62% and that of the testing model was found to be27.63%.

Lastly, when we applied AdaBoost or Adaptive Boosting to our prediction model, we were able to achieve an accuracy score of 1.0 for the training model and accuracy percentage 84.21% for the testing models which are yet less than those of Random Forest Classifier algorithm.

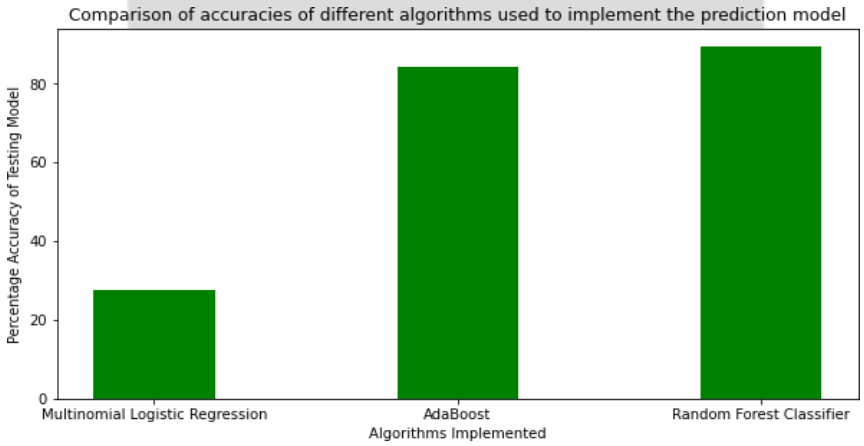


Fig 2 : Representing The Comparison Of Accuracies Of The Testing Models

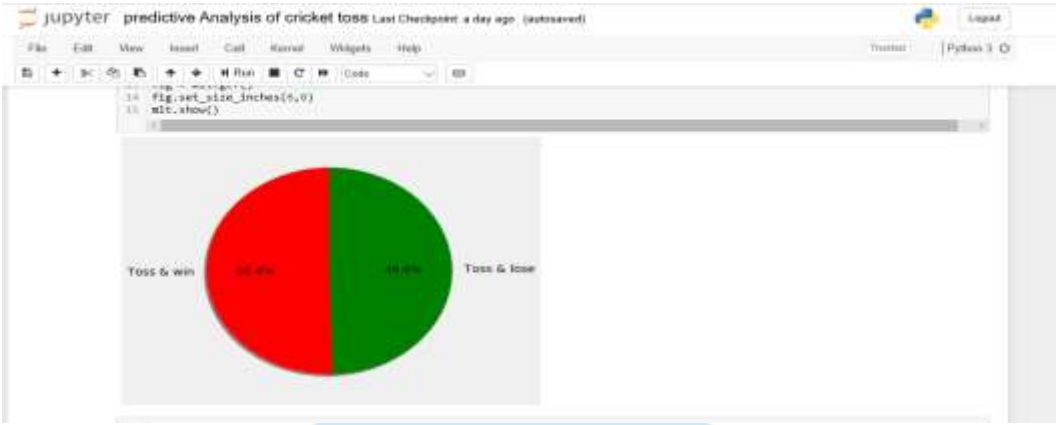


Fig 3 : Generalized Probability For Winning Match by Winning Toss

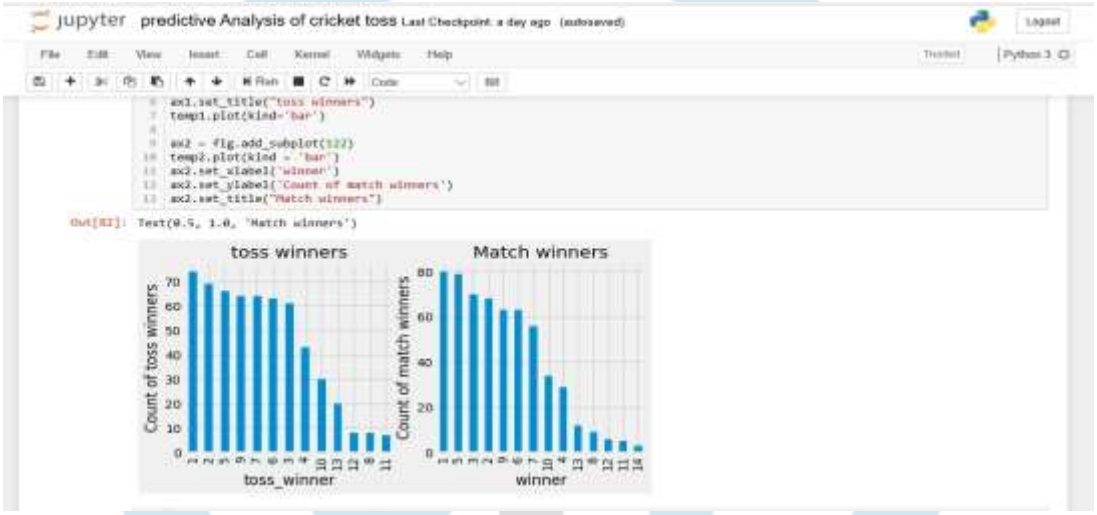


Fig 4: Visualizing Match Winners and Toss Winners

VI. CONCLUSION

In a previous project, the issue of predicting the results of One Day International cricket matches by using the Google prediction API was addressed. However, in the proposed project, we addressed the issue of the likelihood that toss winners will also win matches and predicted the success of a match based on the location of the match and the city in which it was played. For that, we took into account IPL games from 2008 to 2019. We have examined the obtained data and demonstrated how several characteristics, such as the outcome of the coin toss, the winner, the location of the match, etc., affect the outcome of the game. To make our prognosis more illustrative, we offer solid graphs and figures. The study also provides a framework for future research in the realm of cricket and may provide vital predictions for topics like best team of players, best venue, best city, best fielding decision to win a match. For building our Prediction Model we have chosen Random Forest Classifier Algorithm and implemented it to receive a good accuracy score as it can be seen in the Results. And, to provide a proof of concept, we have taken into account two more algorithms. First one is Multinomial Logistic Regression, which is one of the mostly used algorithms in the previous prediction models and the other one is AdaBoost which is less prone to over fitting and is considered as a good fit in such prediction model but has not been used in previous cricket match prediction models. As it can be seen from the results, Multinomial Logistic Regression is not at all suited for our model that tries to predict the outcome of IPL matches and AdaBoost is a good fit but not the best. Therefore, as cricket is a very unpredictable game and its outcome depends on number of factors hence, every percent increase in the model's accuracy will be counted as very important and also, we aimed to build a model that outperforms the past iterations of such model. Therefore, we went ahead with the Random Forest Classifier Prediction Model.

CONFLICTS OF INTEREST

There still are some scopes to improve upon. We are currently using the IPL game database to demonstrate the functionality of the future method we can create a separate database to predict world-class players. We can extend the system to manage flexible match data where the manager can add data after each game to make the analys is more dynamic. Furtherly this project opens a scope for future works in the field of cricket and may predicting other important things like best team of players, best venue, best city, and best fielding decision to win a match. Here we have used different algorithms like Logistic Regression, Support Vector Machine, Naïve Bayes, K-Nearest Neighbour, Gradient Xg-boost, Random Forest and we have achieved an accuracy of 89%. In future we can use different algorithm to reduce space and computational time.

REFERENCES

[1] Parker, D., Burns, P., & Natarajan, H. Player valuations in the Indian Premier League, Frontier Economics, 2008,1-17
[2] Gunjan Kumar. Machine Learning for Soccer Analytics. 2013

[3] Madan Gopal Jhawar & Vikram Pudi. Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in
[4] Databases, 2016

[5] Abel Hijmans. Dutch football prediction using machine learning classifiers

[6] C. Deep Prakash, C. Patwardhan &Sushobhit Singh. A new Machine Learning based Deep

Performance Index for Ranking IPL T20 Cricketers, International Journal of Computer
Applications 137(10):42-49, March 2016
[7] P. Kalgotra, R.Sharda, Chakraborty, Predictive Modelling in Sports Leagues: An application in Indian Premier League, SAS Global Forum, 2013
[8] P.K. Dey, D. N. Ghosh, A.C. Mondal. Multi-Criteria Decision Tree Approach to Classify All- Rounder in Indian Premier League, Journal of Emerging Trends in Computing and Information Sciences, 2011. 2 (11), 563-73

[9] https://www.crummy.com/software/BeautifulSoup/#Download

[10] http://pandas.pydata.org/

[11] https://www.tableau.com/

[12] Christopher Bishop. Pattern Recognition and Machine Learning. 2e.