

Prediction and Analysis of Indian Premier League Using Machine Learning

Miss. Shivani VijayKumar Ambarge¹ Mr. Shivshant Vinayak Patil² Mrs. Sneha Pravin Madnaik³

Mr. Rushikesh Shripad Pandit⁴ Dr. Mrs. D. A. Nikam⁵

^{1,2,3,4,5}Department of Computer Science and Engineering

^{1,2,3,4,5}Dr. J. J. Magdum College of Engineering, Jaysingpur, India

Abstract— Cricket is one of famous sport in world. It contain large set of statistical data. IPL is played in 20 - 20 format. This game is very unpredictable. IPL match analysis is a ML based approach where the data sets and previous stats are trained in all dimensions covering all important factors such as: season, city, date, toss_winner, toss_decision, result, win_by_run, win_by_wicket, player_of_match etc. It focuses on measuring the outcome of Indian Premier League (IPL) matches by applying the existing data mining algorithms to the balanced as well as imbalanced dataset. This model is very much popular in analysing and predicting modeling.

The eight founding franchises were the Mumbai Indians, the Chennai Super Kings, the Royal Challengers Bangalore, the Deccan Chargers (based in Hyderabad), the Delhi Daredevils, the Punjab XI Kings (Mohali), the Kolkata Knight Riders, and the Rajasthan Royals (Jaipur). In late 2010 two franchises, Rajasthan and Punjab, were expelled from the league by the BCCI for breeches of ownership policy, but they were later reinstated in time for the 2011 tournament.

Keywords: Data Science, Numpy, Pandas, Matplotlib, Cricket Prediction, Supervised Learning, Decision Tree

I. INTRODUCTION

Indian Premier League (IPL), Indian professional Twenty20 (T20) cricket league established in 2008. The league, contain 8 teams, battle with each other. Analytics of sports would be procedure for gathering previous game information and investigating it extricate basic information from it, from an expectation which encourages powerful and dynamic judgement. It could be whether to buy a player or not in auction, else whom to set on the field in coming match, using more competitive task like, preparing the strategies to matches in future depending on the prediction being made using various factors from past matches.

In this paper we took some statement for analysis and its output in form of graphs will show. Our data set information is since 2008 to 2019 IPL season. This paper we took some statement for analysis and predicting and its output in form of graphs will show. Our data set information is since 2008 to 2019 IPL season. Today's level of sports analytics has evolved where both the technology which provides data, and the statistical methodologies which provide the tools for analyzing data, improved very rapidly. Though sports analytics has been rapidly developing, it has not been the case with cricket. Due to historical reasons where cricket was perceived as a leisurely gentleman's game played without remuneration to players (until recently), cricket was not subject to large financial transactions.

The game of cricket is played in three formats – Test Matches, ODIs and T20s. We focus our research on IPL. To predict the outcome of IPL cricket matches.

II. LITERATURE REVIEW

Different ML logics are used and they have proved its efficiency. Connection between ML, games goes previously to the underlying long periods for man-made consciousness. Man-made consciousness contemplated that machine learning draws near utilizing the game of checkers in cricket, to foresee a match's result, essential duty is separating basic highlights that influence after effect of a match. Fascinating works have been done in order to foreseeing result. This literature study wraps up most produced works which anticipated an after effect in a match coordinate earlier for 50 over cricket. The Cricket dataset describes the various data mining techniques like Decision Tree, Naive Bayes, Random Forest applied on the IPL dataset, the project is built for predicting the results of the matches. Selection of the best team is always required by the management for best outcome the results of the IPL batting Statistics were grouped into various Data and it gives effective accurate results with the Data Mining Technique.

The concept of Data Mining is used in order to classify batting statistics of the Indian Premier League Live Cricket Score and Winning Prediction work 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, ranking of the team, were considered. Parag Shah, in his work of predicting outcome of the live match proposed model which predict the match result after each ball. The par score concept has been used Duckworth & Lewis, the probability is calculated and it provides clarity of who will win the match.

Bandalasiri Jhanwar and Vikram have made predicting model for the outcome of a One Day International (ODI) match by using a supervised learning approach. They have used statistics of each active player as well as recent performance of each player. Thus, they have finalized KNN as their preferred algorithm for classification.

Rameshwari Lokhande and P.M.Chawan In this paper, match outcome prediction is done while the match is in progress i.e. live match prediction. Thus, features such as number of wickets fallen, match venue, team ranking, pitch report, home ground advantage, etc., are considered, rather than pre-match features such as past player performance, past team experience, etc. Thus, this paper discusses about various cricket elements gathered from various research papers related to cricket match outcome prediction.

Jaishankar and Rajkumar had predicted the outcome of a One Day (ODI) cricket matches using machine learning concepts such as supervised learning. This will also identify the champs of wrong and area of improvement in the model

so as to increase the accuracy of the model. So in these paper we had used various types of machine learning algorithms and comparison has been made in order the identify the best trained model with best accurate algorithms. We had considered the past statistics of matches along with team credibility in terms of performance like batting, bowling, etc. However the unpredictability of the match result will be there at some extent due to capacity of the players in terms of performance, different types of parameters. So in these paper we had trained model with help of four different types of algorithms and also comparing amongst them so as to get best and desired outcome with highest accuracy.

Akhil et al. has made a prediction system for a T20 matches in a particular IPL matches in dynamic nature (i.e. during match is in progress). Currently the score of first innings can be predicted on the basis of current run rate which can be evaluated by the basic formula however it does not include various features like home ground, venue, toss etc. which is too important during prediction because it can change the outcome of the match and also there is no prediction model to predict the outcome of the second innings of the match. In this system they have tried to predict the each inning of the match by taking all the features which is needed by using Multiple Variable Linear Regression along with Logistic regression and finally predicting the outcome of the match using Random Forest algorithm.

III. METHODOLOGY

A. Understanding Data

Understanding means gathering of the match data. The raw data was in an unstructured format with various fields.

We have taken the data and converted it into a structured .csv file format. This dataset is composed of all aspects.

Based on these aspects we have done our analysis.

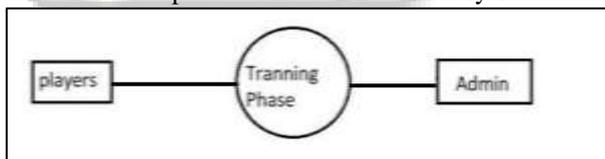


Fig. 3.1:

B. Data preparation

Data pre-processing is an important role in the process. It deals with modifying the data and then converting it into a usable format. Cleaning the data consists of numerous steps which comprises of eliminating null values, joining similar attributes. It plays a significant role and is very continuous and complicated stage.

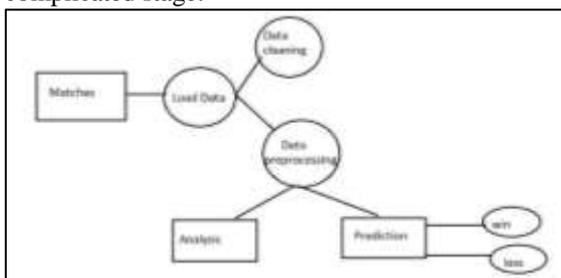


Fig. 3.2:

C. Data Visualization

The data depicting the various statistics related to the project which impacts the outcome of the winning team.

The bar graph is on:

- Likelihood of Toss Determining Win
- Favourite Venues of the League
- Most Man of the Match Awards

1) Likelihood of Toss Determining Win

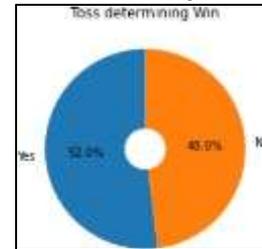


Fig. 3.3:

2) Favorites Venues of the League

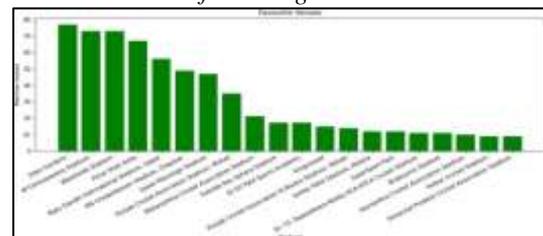


Fig. 3.4:

3) Most Man of the Match Awards

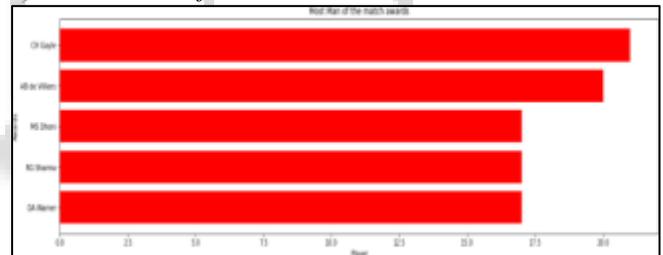


Fig. 3.5:

D. Prediction Phase Result-

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.

There are three types of Machine Learning.

- 1) Supervised Machine Learning
- 2) Unsupervised Machine Learning
- 3) Reinforcement Machine Learning

1) Supervise Machine Learning-

In supervise ML we are given dataset and already now what are our correct output looks like, Having idea that there is relationship between input and output.

Example-. Facial Recognition.

2) Unsupervised machine learning-

It is type of ML in which models are trained using unlabeled dataset and allowed to act on data without any supervision.

Example-Gmail differentiation. Grouping customers by purchasing behaviour.

3) Reinforcement Machine Learning-

In this learning is training of machine learning model to make a sequence of decisions.

Example -Video games, Amazon Alexa.



Fig. 3.6: Interface Phase or Prediction Phase



Fig. 3.7: Prediction Phase with Input and Output result.

IV. EXISTING SYSTEM

Existing system purely depends on how the team deals with the very important factors that influence the outcome there are pretty intelligent systems out there like Dream 11, which use Analytics of the IPL and other cricket data for predicting the outcome As it had limitations Visualizations and limitations Libraries In R because R language having some limitations in Algorithms definition .

V. FUTURE SCOPE

Though it is a fun machine learning project, it can be extended to corporate level also, where Sporting channels would like to show the metrics which can be used to increase the audience pulse, which shows the factors which can change even a losing game inclined to winning. It uses for creating online polls, Sports Business. IPL teams do their own analysis on each and every player as they spend millions on each of them, they use various analysis like Player Vs Ground Battle, before bidding the player.

VI. CONCLUSION

From the study it is observed that there are numerous elements which can impact result of any IPL match. Main factors that fundamentally impact any IPL match could be their host group, non-home group, arena, winner of toss and many more. This relatively helped in the calculation of strength. Different Machine learning techniques were handed down for IPL data set which contributed to this study. The data set consists of all the IPL matches that were held from the past 6 years that is from 2014 to 2019. The prepared models were utilized to foresee the result of IPL matches. The T20 cricket has a scope for changeability, because even few balls can totally change the game. IPL was started 12 years back, there were very less number of games played compared to 50-50 and test games. Thus, structuring Machine learning for anticipating game result with a precession of 75% is exceptionally good at this stage.

REFERENCES

- IPL Match Prediction using Machine Learning Sunil Bhutad¹, Subhani Shaik², Aishwarya Laxmi Nethi³, Pabba Adeeshwar⁴, Sairam Parshi⁵ 1 Professor, IT Department Sreenidhi Institute of Science and Technology Yamnampet, Hyderabad. 2 Associate Professor, IT Department Sreenidhi Institute of Science and Technology Yamnampet, Hyderabad. 3-5 B. Tech IV year, IT Department Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad.
- [1] L. Samuel, "Some studies in machine learning using the game of checkers. iirecent progress," in *Computer Games I*, pp. 366–400, Springer, 1988.
 - [2] Bandulasiri, "Predicting the winner in one day international cricket," *Journal of Mathematical Sciences & Mathematics Education*, vol. 3, no. 1, pp. 6–17, 2008.
 - [3] Indian Premier League Official Website
 - [4] P. Langley, W. Iba, K. Thompson, et al., "An analysis of bayesian classifiers," in *Aaai*, vol. 90, pp. 223–228, 1992.
 - [5] S. Kampakis and W. Thomas, "Using machine learning to predict the outcome of English county twenty over cricket matches," *arXiv preprint arXiv:1511.05837*, 2015.
 - [6] L. Passfield and J. G. Hopker, "A mine of information: can sports analytics provide wisdom from your data?," *International journal of sports physiology and performance*, vol. 12, no. 7, pp. 851–855, 2017.
- T. H. Davenport, "What businesses can learn from sports analytics," *MIT Sloan Man-agement Review*, vol. 55, no. 4, p. 10, 2014.