

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 6 ISSUE : 9 Print / Issue Publication Date: 01-Apr-2022

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2022.v06i09.012

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 9, ISSN No. 2455-2143, Pages 104-107
Published Online January 2022 in IJEAST (http://www.ijeast.com)



IPL WINNER PREDICTION USING ML ALGORITHMS

Ruchitha M
Student
Department of Computer Application
Jain Knowledge Campus, Bangalore Jain University, India

Dr. Preeti Savant
Assistant Professor
Department of Computer Application
Jain Knowledge Campus, Bangalore Jain University, India

Abstract: **Cricket is a popular sport in India as well as the rest of the world. In recent years, the T-20 variation of this game has grown in popularity. The Indian Premier League (IPL), a tournament based on this structure, has risen dramatically in recent years. On the other hand, Cricket is known as the game of chance. Fans and followers are also concerned with predicting the victor of a tournament or match. Technology, on the other hand, is rapidly changing. After training a model, researchers always turn to machine learning algorithms to predict something. So, in this research, we use various supervised learning approaches to predict the winners of Indian Premier League matches.We have the following attributes in this system: team names, match venue, toss winner, toss decision, match winner, won by how many runs, and umpires present for the match. Logistic regression, Decision trees, Random Forest, SVM, Naive Bayes, Gradient boost, and KNN are examples of supervised techniques utilized in this research.**

Keywords**: - Indian Premier League, prediction of the match, cricket, Classification Algorithm, Dataset, machine learning.**

I. INTRODUCTION

Machine learning is a branch of artificial intelligence in computer science that uses statistical methods to allow computers to "learn" from data without needing to be explicitly programmed.

Following concurrent advances in computer power, large amounts of data, and theoretical understanding, ML techniques have experienced a resurgence in the twenty-first century, and ML techniques have become an essential part of the technology industry, helping to solve many challenging problems in computer science, software engineering, and operations research.

The major goal is to identify the essential parameters that influence the match outcome and to choose the optimal machine learning model that fits the data and produces the best outcomes. Some work has already been published in the domain of forecasting the outcome of a cricket match. Because only a few essential factors are used to forecast in certain papers, the accuracy is lower. However, the machine learning model is incorrect in other studies. As a result, it's crucial to think about all the important factors that could affect the match's outcome and the best model for training and analyzing the data. This will significantly improve prediction accuracy.

Many research papers have been published and a lot of work has been done in the last few years to forecast the outcome of a cricket match using supervised machine learning algorithms such as linear regression, support vector machines, logistic regression, decision tree, Bayes network, and random forest, as well as elements that influence the match outcome. Cricket is a broadly performed recreation all around the world. Because it is a fast-paced variant of the game, the Twenty20 format is immensely popularthat appeals to both on-field spectators and at-home viewers.

One of the long-term goals of the field is general intelligence. Statistical methods, intelligence based on computing, and Symbolic machine learning aresome of the ways. In machine learning, a variety of tools are employed, Search and mathematical optimization, artificial neural networks, statistics, probability, and economics techniques are just a few examples.

II. LITERATURE SURVEY

According to the findings of the literature review, there is a need for a machine learning model that can predict the outcome of a match before it starts. The IPL format, more than any other cricket format, sees a lot of changes in the game's momentum. A game can be radically changed by an over. As a result, predicting the outcome of an IPL match is challenging. Furthermore, creating a prediction model for a league that is entirely based on auction is a challenge. Matches in the IPL cannot be predicted only based on statistics collected from prior data.

Rabindra Lamsal and Ayesha Choudhary [1] used data from the official website of the Indian Premier League to compile this article. Because the data included a lot of features, they examined it and picked out a few of the most important ones. They pre-processed the data and applied a few models using the Scikit-learn machine learning package. Low

International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 9, ISSN No. 2455-2143, Pages 104-107
Published Online January 2022 in IJEAST (http://www.ijeast.com)

variance, univariate, and recursive characteristics were deleted. They found 5 critical features out of 15 by employing these feature selection models. Home team, away team, venue, toss winner, toss decision, and winner are the features. The Random Forests and Multiple Linear Regression models were trained using data from the first ten seasons of the IPL, followed by data from the eleventh season. Their model was able to accurately categorize 41 out of 60 matches. As a result, their accuracy is 68.33 percent, which isn't great. This model has limitations in that it only uses 5 features and two machine learning models.

Sankaranarayanan [2] presents the idea for creating a prediction system that uses previous data to determine whether upcoming matches will be won or lost. They employed Linear Regression, Nearest Neighboring, and Clustering approaches to show mathematical results and demonstrate the algorithm's performance in predicting model results.

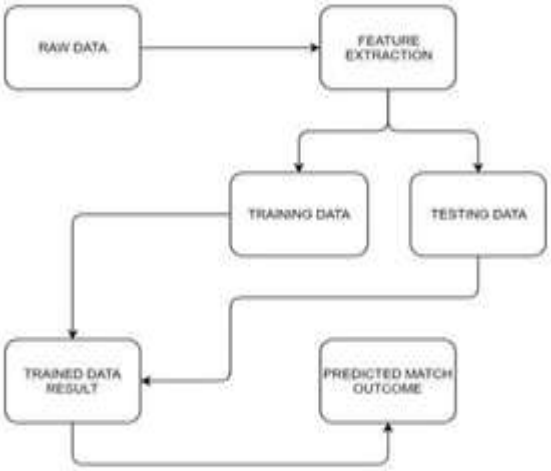
Elnaz Davoodi and Ali Reza Khanteymoori [3] presented a technical paper titled "Horse Racing Prediction Using Machine Learning Predictions." have given their thesis on machine learning predictions on horse racing where they used various models of machine learning which has given pictorial representations of the data using graphs, bar charts and used built-in function of data pre-processing which made the work easier evaluation of classification models.

Technical paper which is given by K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal [4] on the topic “Stock Market Prediction Using Machine Learning predictions” where the major goal of this research is to establish the best model for predicting stock market value. Considering many strategies and variables that must be considered during the procedure, I've discovered that tactics such as random forest, support vector machines were not exploited fully. They developed and reviewed a more practicable strategy for predicting stock movement with greater accuracy in this research. The dataset of stock market values is the first item they looked at from the previous year and predicted the accuracy from these models.

Using the ODI dataset, a predictive model is created to predict cricket score and player performance. Supervised approaches such as SVM and Naive Bayes were used. Clustering methods like KNN and MLP are used to accurately categorize. ML approaches are used to predict the winner of IPL matches [1]. Every year, there is lots of hypothesis approximately who might win the Indian Premier League's famed title. IPL is a game in which the outcome may be changed in a matter of seconds using machine learning techniques such as SVM, Naive Bayes, KNN, Decision Tree, and Logistic Regression to predict the winner.

Basketball outcomes are predicted using machine learning techniques [5]. Some machine learning approaches utilized in this study to construct a model for forecasting NBA game results include Simple Logistics Classifier, Artificial Neural Networks, SVM and Nave Bayes, and Random Forest. Data from 5 regular NBA seasons was collected for model training, while data from 1 NBA regular season was used as a score dataset, resulting in a compelling result. A data mart comprising NBA statistical data is generated when automated data collecting and cloud-based data management operations are completed. The above-mentioned machine learning models are then trained and tested by consuming data from the data mart. Simple Logistics Classifier eventually delivers the best result with an accuracy of 69.67 percent after applying the scoring dataset to evaluate the model accuracy.

III. SYSTEM ARCHITECTURE



The conversion of raw data into processed data is the first stage in the picture above. Because there are various qualities in the raw data obtained, but only a few of those attributes are important for prediction, this is done through feature extraction. The first step is feature extraction, which entails extracting the most important properties from the raw dataset's full list of attributes. Feature extraction starts with a set of measured data and works its way up to derived values or features. These characteristics are intended to be both informative and non-redundant, facilitating learning and generalization.
Following the feature extraction technique, the data created after feature extraction is separated into two portions by a classification procedure. Classification is the process of deciding which category a new observation belongs to. The

test data set is used to predict the model's accuracy, while the training data set is used to train the model. The training data is isolated from the test data in such a way that it maintains a larger proportion.
The random forest algorithm uses a set of random decision trees to analyze data. In layman's words, a cluster of decision trees searches the data for specified attributes from the forest's total number of decision trees. Data splitting is the term for this. In this situation, our proposed system's end goal is to predict the winner of the IPL.

IV. METHODOLOGY

a. Data Collection and exploration
Data gathering is a fundamental module and the project's first stage. It mostly concerns the gathering of the appropriate dataset. A variety of filters must be applied to the dataset that will be used to produce market forecasts. Data collection also helps to improve the dataset by incorporating more external data. We'll start by examining the Kaggle dataset, and then we'll use the model with the data to examine the predictions accurately based on the accuracy.

b. Data pre-processing
Machine learning relies heavily on data pre-processing to get highly accurate and insightful outputs. The more reliable the produced results are, the better the data quality is. Real-world datasets are characterized by incomplete, noisy, and inconsistent data. Data pre-processing improves data quality by filling in missing or partial data, reducing noise, and addressing discrepancies.

c. Data Cleaning
The process of eliminating errors and replacing them with genuine values is known as data cleaning. The data sets gathered contain noisy data, such as null values and inappropriate values, which must be cleaned. As a result, the data is cleaned by replacing null values with zeros, and the data is organized into correct columns so that we can properly analyze it.

d. Data Visualization
The data that has been gathered is used to visualize the information for better comprehension. The Matplotlib Library was used to visualize the graphs for team wins in various cities based on their venues and player strike rate.

e. Train and test splitting
To build an important prepared set, the issue for which it is being settled must be grasped. The preparation set, which is the larger of the two, is the one that is being used. When you run a set of preparations through an AI framework, it informs the net how to optimally weigh certain features, translating them into coefficients based on their likelihood of constraining errors in the outcomes. Those coefficients, which are referred to as boundaries, in any case, will be stored in tensors, which are collectively referred to as the model because they encapsulate a model of the data on which it is being trained. They are the most important lessons learned from putting together an AI framework. The test set comes after that. It serves as an underwriting seal and isn't used until the last end. The neural net can be tested against this last subjective judgment after it has been prepared and data has been set.

V. CONCLUSION

This project has helped us a lot in understanding the basics of prediction systems in machine learning and related concepts. This IPL prediction mini-project has also helped in learning a new data science operation in Python and its in built libraries.
The algorithm will be prominent for sports news analysis since it was chosen after being tested on sample data and has been trained on a large collection of historical data In comparison to previously implemented machine learning models, the project shows that the machine learning model can predict the match-winner with greater accuracy.

VI. 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]. Abhishek Naik, Shivanee Pawar, Minakshee Naik, SahilMulani, “Winning Prediction Analysis in One-Day-International (ODI) Cricket Using Machine Learning Techniques”, International Journal of Emerging Technology and Computer Science, Volume: 3 Issue: 2 (April 2018)
[3]. Arjun Singhvi, Ashish Shenoy, Shruthi Racha, and Srinivas Tunuguntla. “Prediction of the outcome of a Twenty-20 Cricket Match.” (2015).
[4]. Swetha, Saravanan.KN, “Analysis on Attributes Deciding Cricket Winning”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03 | (March 2017)
[5]. Goddam Jaishankar Harshit, Rajkumar S, “A Review Paper on Cricket Predictions Using Various Machine Learning Algorithms and Comparisons among Them”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), IJRASET17099 (April 2018)
[6]. Akhil Nimmagadda, Nidamanuri Venkata Kalyan, Manigandla Venkatesh, Nuthi Naga Sai Teja, Chavali Gopi Raju,
[7]. Research which is given by K. Hiba Sadia, Aditya Sharma, Adarsh Paul, SarmisthaPadhi, Saurav Sanyal



on the topic "Stock Market Prediction Using Machine Learning predictions"

[8]. Research paper by Elnaz Davoodi, Ali Reza Khanteymoori on" Horse Racing Prediction Using Machine Learning predictions".

[9]. Vignesh Veppur Sankaranarayanan and Junaid Sattar, "A Data Mining Approach to ODI Cricket Simulation and Prediction", Proceedings of SIAM Conference on Data Mining, pp. 1- 7, 2014.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE AND TECHNOLOGY

AUTHORS

Ruchitha M

Dr. Preeti Savant

ABOUT IJEAST

International journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high - quality research paper in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

JOURNAL METRICS

H-INDEX	i10-INDEX	IJEAST CITATIONS	PEER-REVIEWED & OPEN ACCESS
33	154	9,647 (IJEAST CITATIONS)	PUBLISHED MONTHLY

PEER REVIEWED

All submission are rigorously peer reviewed to ensure quality.

AUTHORS ARE OUR PRIORITY

We value our authors and recognize their contribution to the advancement of research and knowledge.

OPEN ACCESS

Free and unrestricted access to research for all.

GLOBAL REACH

Connecting researchers and professionals worldwide.

TIMELY PUBLICATION

We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

editor@ijeast.com

www.ijeast.com

India

2455-2143