MACHINE LEARNING BASED IPL WINNER PREDICTION ANALYSIS

Mrs.Ranjani.J[1]
AP
dept.of IT
Sri Sai ram Engineering College
Chennai,India
ranjani.it@sairam.edu.in

Mrs.B.Kamala[2]
AP
dept.of IT
Sri Sai ram Engineering College
Chennai,India
kamala.it@sairam.edu.in

Mrs.V.K.G.Kalaiselvi[3]
AP
dept.of IT
Sri Sai ram Engineering College
Chennai,India
Kalaiselvi.it@sairam.edu.in

KarthikRaja Y[4]
UG Scholar
Dept.of IT
Sri Sai ram Engineering College
Chennai,India
sec20it004@sairamtap.edu.in

Srikant A[5]
UG Scholar
Dept of IT
Sri Sai ram Engineering College
Chennai,India
sec20it049@sairamtap.edu.in

ABSTRACT

In India and all around the world, the Indian Premier League -IPL is the most popular league. The most recognized cricket format currently across viewers is T20. One of the reasons for the expansion of the T20 format of this game is a competition called the IPL, which uses the T20 format. Because anything can happen in this kind of cricket, experts used to refer to it as a humorous game. The squad with the best prospects of victory will be examined in our proposed study. SVM, logistic regression, KNN, decision trees, etc. are a few of the key methods we employ in our projects. Finally, the algorithm's conclusion indicated that decision trees and logistic regression had efficiency levels of over 89% and 97%, respectively. The best four methods for machine learning will be used to select the winner, and graphs of the findings will be displayed.

KEYWORDS:

IPL, Prediction, Machine Learning, SVM- Support Vector Machine, The Board Of Control For Cricket In India (BCCI)

1. INTRODUCTION

ML is a study of the wider category of AI which make use of mathematical structures to develop concepts. In simple words, ML uses algorithms which take statistical or previous data in and analyze it and produce the results according to the study. The main objective is to find essential parameters which affect the result of match and select the suitable ML model which matches the data and provides an excellent outcome. Few works are published on this category of predicting the match outcome. In some articles, only a small number of significant elements are utilized to forecast, but the accuracy is inferior. Billions of viewers from across the globe watch the T20 cricket competition known as ipl in india. There is an exciting and thrilling expectations around every match in a league, which consists some of top cricket players in the globe. With the development of artificial intelligence and machine learning. It is nowadays more interesting among people to predict the result of the match. Here we will talk about how to analyze the winning team of the ipl.

When the International Cricket Council(ICC), T-20 matches, which represent for 20-over cricket, started from the idea of fast play, cricket's popularity skyrocketed. India's interest in this sport surged when the first Twenty20 World Cup was hosted in South Africa in 2007The (IPL)-Indian Premier League emerged in 2008 when the The Board Of Control For Cricket In India(BCCI) grabbed on the chance and gained ICC approval. IPL is one of the best 20-over cricket leagues currently being performed and it is modeled after the NBA and the EPL (English Premier League) of football and basketball, respectively [5]. IPL's first season had tremendous success, providing possibilities to numerous stakeholders. Eight teams play against one another in the first stage of every IPL season. Four teams advance to the next level, the eliminator round, following which two teams play in the final match, where one team will eventually be crowned the victor. Each team is owned by a franchise organization, which is an organization. These franchises engage players after analyzing them depending on their performance and preceding national, international, and T-20 experience.[1].

Every IPL match's outcome is influenced by a number of elements, such as the player, venue player performance, toss etc. Only when previous player performance, the location of the match, and other match-related variables are available,the final result of a competition is often somewhat predicted. Three machine learning techniques (SVM)Support Vector Machine, Nave Bayes, and CTree are employed in this work to predict the outcome of an IPL match with data that is already available. [2]

2. LITERATURE SURVEY

Kansal et al. [3]. We have found that methods like random forests and support vector machines were not fully utilized due to the numerous strategies and variables that must be taken into account during the operation. In this study, they developed and validated a more virtual method for projecting scores with more probability. We gathered data from the previous year to validate these models' odds. utilised information from the IPL website's official statistics for this document. We examined the data and chose a few of the most important aspects because it contains multiple characteristics [2]. Jaiswal[6] explains the idea for developing a prediction engine that looks at historical data to anticipate the result of upcoming IPL matches. Wright [7]The teams, venues, and quantity of holidays before each match served to predict the likely schedule for a game of cricket in an acceptable and efficient manner. Using an evolutionary a plan of action the Subcost-Guided Simulation of Annealing method goes from a simple initial outcome to a challenging end solution.

EXISTING ANALYSIS

A formula for calculating the predicted score and an overall analysis based on a poll's and a team's winning % is attainable in the analysis of the present system. These methods rely on assumption and make prediction based on certain circumstances, thus the outcomes won't be excellent. This is the formula that will be used to calculate the projected score: Current Run Rate* Overs in an Inning = Projected Score

DEMERIT:

Following the suggested approach leads to a small rise in mean precision.

3. PROPOSED ANALYSIS

The main goal of our project is to use labeled data to anticipate the runs. Gathering information on all IPL matches is the first step toward achieving the goal because an immense quantity of data must be handled in order to obtain accuracy. Data that provides complete details about every IPL match has been collected. The attribute utilizes a limited amount of core knowledge to gather data on all IPL matchups. Our proposed paper primary objective is to use labeled data to predict the runs. To achieve the goal, data from all IPL matches must first be collected. A large amount of data must be evaluated and preprocessed in order to attain accuracy. Data that provides detailed details about every IPL match was gathered. Following preprocessing, the data will be trained on several models to produce the results. In order to predict the likely outcome of a match, we will examine multiple datasets and employ important variables like strike rate, bowler economy, etc. [4].

MERITS:

- One of the supervised training algorithms used for both

regression and classification is the decision tree.

- An method for supervised learning is random forest. If there are a greater number of trees in the random forest classification algorithm, the model's accuracy will be at its best.

4. METHODOLOGY

DATASET LOADING

The dataset is known as match.csv and it is 2,09,070 bytes in size. It is taken from the Repository of Kaggles and comprises IPL Matches data from 2008 to 2023. There are 20 attributes, and there are 686 records overall.

DATA PRE-PROCESSING

In machine learning, preprocessing of the data is crucial. It translates data that is raw into a format that can be used. It is typically used as the first stage in the data cleansing process. The preparation of information puts the data in a format that allows for error-free classification processing.

FEATURE SELECTION

The next step, feature selection, involves applying specific requirements from the data collection to improve performance. Variable selection is another term for feature selection.
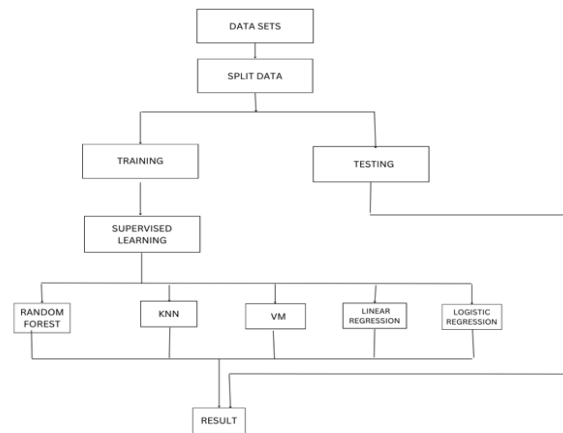
5. SYSTEM FLOW DIAGRAM:

FIG.1 SYSTEM FLOW DIAGRAM

6. EXPERIMENTAL RESULTS

Experimental Setup

The experiments were performed using hardware using with very least requirements, including an 1GB of free disk space, Intel Atom, macOS, Windows 7 or later, and Linux. Jupyter Notebook and Python 3.8 were used to implement the machine learning algorithm. From Kaggle, the datasets were downloaded.

Analysis of IPL datasets

A few features, like team winning statistics when batting first and when bowling first, were extracted from the IPL dataset after it had been evaluated.[4]
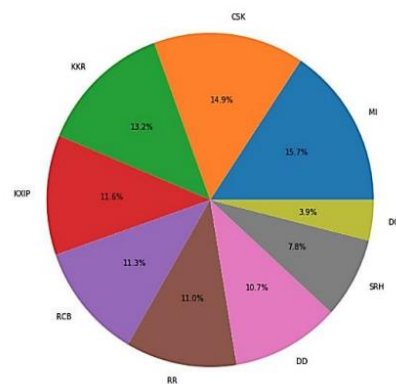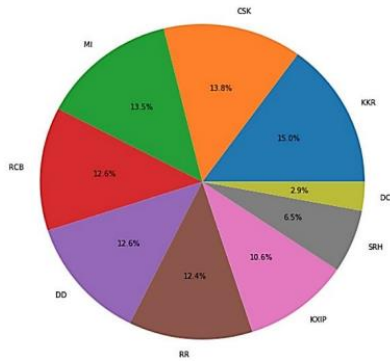

Fig:2 Batting first-winning percentage

Fig:3 Bowling first-winning percentage

Performance Evaluation

Accuracy:

Accuracy can be defined as the percentage of accurate forecasts among all predictions. The random forest classifier excels all other algorithms in this experiment by correctly predicting the outcome 88.10% of the time. The precision of each algorithm is shown in the following figure.
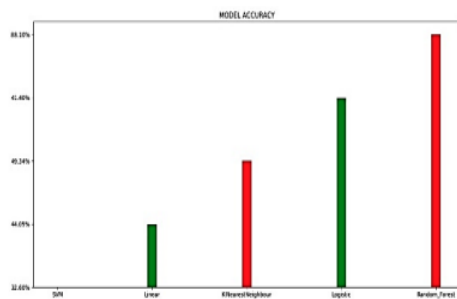

Figure:4 Prediction of algorithms

The Accuracy of the machine learning algorithms achieved shown below:

Algorithm	Accuracy
Random Forest	88.10%
K-Nearest Neighbor	49.34%
Logistic Regression	51.40%
SVM	32.6%
Linear Regression	44.05%

7. CONCLUSION

In a sport like cricket, selecting the winner is extremely challenging and includes multiple intricate procedures. But this can be significantly facilitated and reduced with the introduction of machine learning. In this essay, a number of factors that impact how matches in the Indian Premier League come out have been identified. An IPL match's result is greatly affected by the teams concerned, where it happens, the city, the individual who wins the toss, and the toss decision. We analyzed IPL data sets according to player performance and projected outcomes. Decision tree, LR-Logistic Regression, DT-Decision Tree, (SVM) -Support Vector Machine RFC-Random Forest Classifier, and K-nearest neighborhood are the methods utilized in the task to produce the final test. RFC classification carried out better than the different approach. Future considerations can include each player's performance and regular evaluation during the entire season. His bowling and batting ratings can also be anticipated. It may be possible to anticipate which player will help each team win the game.

8. REFERENCES

[1] Kumar, M. Keerthi, S. Prabu, B. D. Parameshachari,Silvia liberata Ullo. "Comparative Analysis to Identify Efficient Technique for Interfacing BCI System." In IOP Conference Series: Materials Science and Engineering, vol. 925, no. 1, p. 012062. IOP Publishing, 2020

[2] Balasundaram A, Jayashree D, Ashokkkumar S, Magesh Kumar S, "Data

Mining based Classification of Players in Game of Cricket", proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020), IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9.

[3] Prince Kansal, Pankaj Kumar, Himanshu Arya, Aditya Methaila, Player valuation in Indian premier league auction using data mining technique, International Conference on Contemporary Computing and Informatics (IC3I), 27-29 Nov 2014

[4] Prabu, S., Balamurugan Velan, F. V. Jayasudha, P. Visu, and K. Janarthanan. "Mobile technologies for contact tracing and prevention of COVID-19 positive cases: a cross-sectional study." International Journal of Pervasive Computing and Communications (2020).

[5] Manuka Maduranga, Hatharasinghe, Guhanathan Poravi, Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Links, 5th International Conference for Convergence in Technology (I2CT), Mar 29- 31, 2019.

[6] Arya Kothari, Harshit Barot, Romit Kankaria, Pramod Bide, Bhavya Ahir, "Analysis and Prediction of Indian Premier League", 2020 International Conference for Emerging Technology (INCET), 5-7 June 2020.

[7] M. B. Wright, Scheduling fixtures for New Zealand Cricket, IMA Journal of Management Mathematics 16, PP. 99−112,2005