Analyzing and Predicting Outcomes of IPL Cricket Data

Kilari Jaswanth[1], Kencham Arun[2], B Deva Deekshith[3], Kilari Ashwik[4], Pavithra J[5]

[1,2,3,4]Student, Department of Computer Science and Engineering, K S Institute of Technology (Affiliated to VTU), Bangalore, Karnataka, India
[5]Assistant Professor, Department of Computer Science and Engineering, K S Institute of Technology (Affiliated to VTU), Bangalore, Karnataka, India

---

ABSTRACT: The Indian Premier League (IPL) is a professional men's Twenty20 cricket league in which 10 teams compete from ten different locations. Millions of people, especially Indians, are obsessed with the Indian Premier League (IPL), and our job involves data analysis and match prediction for IPL matches. In recent years, analytics has been used to predict and draw various insights in the field of sports. IPL Data Analysis is all about utilizing data science, machine learning to analyze data that is already existing in a data collection. This application design will be implemented for the purpose of analyzing the IPL data by fetching different attributes and building a predictive model that could predict the score, batsmen run, predict the winner, overall performance of the team, and the players, head to head-analysis.

Keywords: **Indian Premier League (IPL), Machine Learning, Analyzation, IPL Prediction**.

I. INTRODUCTION

The use of statistical analysis in sports has been rapidly increasing over the years. As a result, how game methodologies are framed or the player's assessment measures changed, yet the crowd interest in cricket has expanded. Presently cricket is one of the most followed games on the planet. The Indian Premier League is one of the world's most famous T20 cricket associations. The association began in 2008 and was established by BCCI. It is normally held in March and May consistently. IPL is the most well-known cricket association on the planet positioning 6th among all sports associations as far as normal participation in 2014.

With total of eight teams, each team faces each opponent twice in the league. At the final stage, only top four teams get qualified for the playoff. The top two teams from the league will face each other in the first qualifier match, with the winner directly promoted to the Ipl finals and the losers gets a second chance by playing in the second qualifying match. Third and fourth teams in the points table compete each other in the eliminator math. All teams pick more than 150 participants, with each squad consisting of 11 players, four international players, and seven local players.

The eight founding franchises are Kolkata knight riders, Punjab kings, Royal Challengers Bangalore, Sunrises Hyderabad, Mumbai Indians, Chennai super kings, Delhi capitals, and Rajasthan Royals. One of the key goals of BCCI in launching the IPL is to strengthen the ability of domestic players by providing a more competitive and better platform than the domestic cricket circuit. In IPL, it is difficult to know which player to be select in the team, who to buy and auction, and team balance in all aspects. Manual prediction of the team winning is not accurately possible.

The performance of any team is decided by key player performance, team conditions, and other important aspects that influence a cricket team's performance. The model will take into account all of the factors that can affect the outcome of a cricket match. The performance of a team's players determines its victory factor. Player performance is important since every team relies on their players to perform well. The performance of the players is a crucial component in deciding starting lineup of the team. In cricket, pitch conditions are crucial. Matches are held on different types of pitches.

Since it's important to gather the possible factors that are affecting the outcomes of IPL matches. The goal of the work is to collect and analyze data from previous matches in order to extract meaningful information that will assist in the prediction of match results. Prediction and analysis can be done in a variety of ways including taking into account the performance of individual players as well as the performance of the entire team. The object of this research is to provide an effective model to predict and analyze team performance, player performance.

This paper is presented as follows: Section II is Literature Survey Section III is the description of the Dataset, Section IV describes the proposed method for building a predictive model, Section V Includes the conclusion and future work of the paper.

II. LITERATURE SURVEY

[1] Nikhil Donge, Shraddha Dhol, Nikita Wavre, Mandar Pardakhe, Amit proposed two models the first one is a prediction of the score and the second one is a team winning prediction. In these, the score prediction includes linear regression, lasso regression, and ridge regression whereas in winning prediction SVC classifier, decision tree classifier, and random forest classifier are used. They have imported the dataset from Kaggle and selected the required attributes. In Score Prediction analysis accuracy of Linear

Regression is more than Ridge and Lasso Regression and in winning prediction analysis among SVC, Decision tree classifier and Random forest classifier, they got random forest classifiers accuracy more than other 2, with all 90%, 80%, 75%, 70% training data.

[2] Praveen Banasode, Minal Patil, Supriya Verma performed a work about the analysis of the data and predicting of the IPL matches. IPL data analysis is all about analyzing the data that is present in the data set and the predicted player runs, team winners. The algorithm used by them provided an accuracy of over 95%. They have collected datasets from Kaggle and espncricinfo.com which contains data from 2008 to 2019. This application can help in selecting the best players, bowlers, and fielders from each team and predict their future performance.

[3] Amala Kaviya V.S, Amol Suraj Mishra, Valarmathi B has brought out an application that analyzes and visualizes the various aspects of IPL matches in all possible ways and gives useful results to the user. Random forest and other tree-based algorithms are outperformed by the likes of JRIP and SVM. Amongst all the algorithms they have applied, JRIP seems to have the most promising. With an accuracy of 75.86%. The SVM and FDA also gave them a good result with an accuracy of 72.41% respectively.

[4] G. Sudhamathy, G. Raja Meenakshi uses Decision tree, Naive Bayes, K-Nearest Neighbor, and Random forest to build a model that is used to predict the winner and visualize the results as graphs. The dataset are been taken from the Kaggle repository. The number of attributes is 18 and the total number of records is 637. Based on the outcomes, they have concluded that the random forest is performing well than the other algorithm. It is with high accuracy and less error.

[5] Pallavi Tekade, Kunal Markad, Aniket Amage . Bhagwat Natekar prepared a machine learning model for predicting the outcomes of its matches. They have used Decision Tree, Random Forest Regression, Logistic Regression, Naïve Bayes. They have taken 11 seasons datasets as a training dataset which consists of 580 matches. The highest Prediction accuracy is 90% and they have also briefs about the key factors that affect the result of the cricket match.

[6] Daniel Mago Vistro, Faizan Rasheed, Leo Gertrude David uses Random Forest, SVM, Naive Bayes, Logistic Regression, and Decision tree to build a model that predicts IPL match-winner before the match starts. The prediction model will have benefits for many cricket lovers to evaluate the team's strengths and weaknesses and also for gambling applications. The Decision tree classifier, random forest classifier, and XGBoost classifiers show an accuracy of 94.87%, 80.76%, 94.23% respectively. They have taken data from seasons from 2008-to 2017.

[7] Christopher R.Brydges creates a machine learning technique that uses ball-to-ball data from the IPL to predict the match outcomes based on events occurring in the first inning of a match. The algorithm used are AIC, BIC, Random forest, Naive Bayes. The data has been taken from Kaggle. All the data from 2008 to 2020 is being taken. It was found that the AIC model had the highest accuracy(67.18%), followed by the NB model (65.64%), the RF model (62.56%), and the ABC model(62.05%). The only limitation is that it has low accuracy.

III. DATASET

We are using an ML -based approach over here. So an ML

The algorithm's core prerequisites are datasets, training the dataset with the algorithm, and assessing the model.

As a result, we've imported data from Kaggle, Data world, the official website of IPL. The past seven years data of IPL contains various attributes such as venue details,

Runs, and many features to draw the various conclusion that helps in the improvement of players' performance. Some of attributes are a venue, toss winner, the player runs, wickets, winners and we can include more factors while predicting various features.

IV. METHODOLOGY

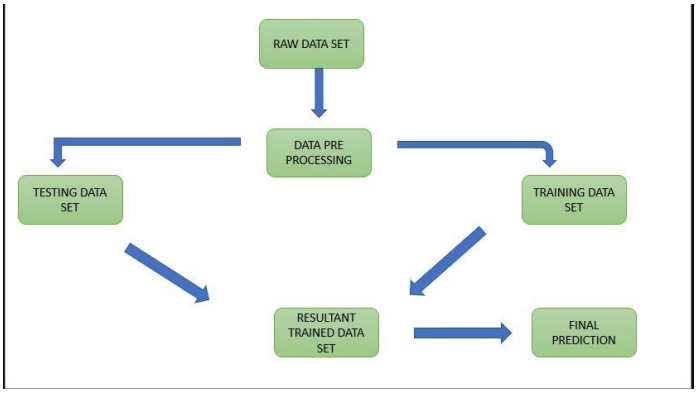

Fig.1 Methodology used for prediction using machine learning.

Dataset: The IPL dataset has been taken from Kaggle, Data world, and the official website of IPL.

Data Pre-Processing: Data preprocessing can refer to the manipulation or dropping of data before it is used in order to ensure or enhance performance. It is a data mining technique that is used to transform the raw data in a useful and efficient format.

Splitting of Dataset: A train-test split is a tool for estimating a machine learning algorithm's performance. Taking a dataset and separating it into two subgroups is the technique.

1. Train Dataset: Used to fit the machine learning model.
2. Test Dataset: Used to evaluate the fit machine learning model.

The objective is to estimate the performance of the machine learning model on new data: data not used to train the model.

Feature Selection: It is the process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and in some cases, to improve the performance of the model. Feature selection is the process of selecting the best collection of features from a set of input features by minimizing data dimension and improving time and space complexity by deleting unnecessary characteristics. Choosing features increases the model's performance while also conserving time and space.

Model Training: The process of training an ML model includes providing training data to an ML algorithm(learning algorithm). The learning algorithm searches the training data for patterns that connect the input data attributes to the goal(desired output that we want to predict).

Regression: Regression analysis computes using a variety of algorithms and predictive value depending on the results. It is a Statistical technique used for estimating the association between the independent variable and independent variable. There are many different regression techniques. Linear regression is one of the regression techniques that will be used.

Classification: Classification is used when the target variable we present particular category. The classification system that we use for IPL match winning predictions, such as "winners" or "losers". Out of which Logistic regression and Random forest are the techniques that will be used.

Proposed System: The complete work done has been compactly organized into this architecture. It starts with the preparation of datasets and their loading into the backend. The user interface is then provided with several features that can be used on the players or matches. It may also be used to make predictions.

We can implement the following modules for analysis, prediction, ranking, and visualization.

- Overall team performance
- Batman analysis and ranking
- Bowler analysis and ranking
- Match analysis
- Team ranking
- Head-to-head analysis

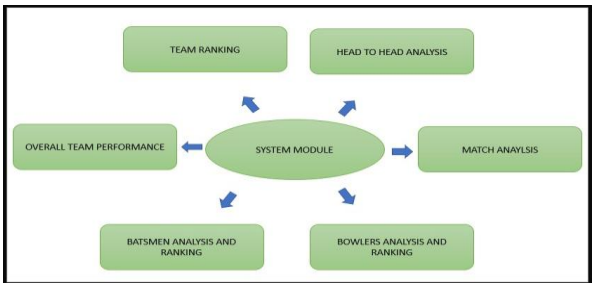

Fig.2 Proposed system

V. CONCLUSION AND FUTURE WORK

This survey helps to propose a model that helps in the prediction and analyzes of the team and player performance. The useful applications are online fantasy games, used by team analyst, which provides stats to cricket lovers and they can also use to access an opponent's strengths and weakness. The IPL prediction helps people who are willing to play online fantasy games, such as dream11, MPL, and other online platforms.

Future research should analyze different supervised and unsupervised machine learning techniques and feature selection techniques with additional performance metrics for better IPL prediction. Hence, the scope of the project is to build predictive model that works with maximum accuracy and includes all the important factors that influences the results is taken, which will work with maximum accuracy and it should consider all important factors that could influence the result.

REFERENCES

1. Nikhil Donge, Shraddha Dhol, Nikita Wavre, Ammit. "IPL Cricket Score and Winning of Stock Market Prediction Using Machine Learning Approach", IRJMETS 2021.
2. Praveen Banasode, Minal Patil, Supriya Verma. "Analysis and Predicting Results of IPL T20 Matches", IOP 2021.
3. Amala Kaviya, Amol Suraj, Valaarmathi. "Comprehensive Data Analysis and Prediction on IPL using Machine Learning Algorithms" IJET 2020.
4. G.Sudhamathy GR, Aja Menakshi. "Prediction on

IPL Data using Machine Learning Algorithms" IJSE 2020.

5. Pallavi Tekade, Krunal Marka, Aniket Image, Bhagwat Natekar." Cricket Match Outcomes Prediction using Machine learning", IJASRE 2020.

6. Daniel Mago, Faizan Rasheed, Leo Gertrude David. "The Cricket Winner Prediction With Application of Machine Learning and Data Analytics", ISTR 2019.

7. Indian Premier League - https://www.iplt20.com/, 2020.

8. https://machinelearningknowledge.ai/ipl-data-analysis-and-visualization-project-using-python/ ,2020

https://github.com/anujvyas/IPL-First-Innings-Score-Prediction-Deployment, 2020

9. https://stats.espncricinfo.com/ci/engine/records/i ndex.html?id=117;type=trophy,2020

10. Christopher R.Brydges "Analytic of Batting First in Indian Premier League Twenty20 Cricket Matches", C.R.(2021)