

IPL Data Analysis and Prediction Using Machine Learning

Dr. Vilas Joshi¹, Yash Padol², Kaiwalya Sevekar², Chandan Sinha², Rizwan Shaikh²

¹Assistant Professor, Department of Computer Engineering, ISBM College of Engineering, Pune, Maharashtra, India

²Student, Department of Computer Engineering, ISBM College of Engineering, Pune, Maharashtra, India

ARTICLE INFO

Article History:

Accepted: 01 May 2023

Published: 20 May 2023

Publication Issue

Volume 9, Issue 3

May-June-2023

Page Number

205-208

ABSTRACT

The Indian Premier League (IPL) is a highly attended T20 cricket league founded by the Board of Control for Cricket in India (BCCI) in 2008. Valued at \$6.7 billion in 2019, the IPL remains the most popular cricketing event in the world, even amidst the COVID-19 pandemic. Cricket is a beloved sport in India and enjoys significant media attention and financial investment. With the rise of technology, audiences can access affordable live streaming of matches from anywhere. Cricket analytics, powered by tools like NumPy, allow for the use of data to improve performance, study business opportunities, and provide insights and predictive intelligence on game outcomes.

I. INTRODUCTION

The Indian Premier League (IPL) is a T20 cricket tournament held annually in India from April to May, with 8 teams representing different states. It is the most attended cricket league globally, and in 2010, it became the first sporting event to be broadcasted live.

This paper examines a prediction model for IPL outcomes, which uses SVM and Logistic Regression, with runs scored by the batsman in the previous ball as the labelled data.

Machine learning techniques like supervised, unsupervised, and reinforcement learning algorithms are used to predict or classify information. Supervised learning is the most popular type and involves learning a function that maps input to output based on example input-output pairs. Regression and

classification are two types of supervised learning. Unsupervised learning involves allowing the model to work on its own to discover patterns in unlabelled data. In contrast, reinforcement learning focuses on learning through trial and error to maximize rewards.

II. LITRATURE SURVEY

G.Sudhamathy's research paper helps to understand the working principle of various machine learning algorithms and their implementation. It creates a model and training dataset and helps make predictions using the model created. The model classifies the data and compares the results to get the accuracy that matters.

As in a dataset, there are many parameters. Which parameters are useful in the project. Factors

influencing the concept has been taken by Maheshwari in his live cricket score prediction from which we come to know the main factors which are required for score prediction and winning team prediction. The role of classification is elucidated in the work of Tejinder Singh, which provides correct information or the use of naive bias and linear regression. They provide proper knowledge of data collection and preparation as well as how to train the data and test the data they provide which is more useful.

The brief idea of Support Vector Machines is taken from Aminul Islam Anik's paper which is about player performance in this article, the idea of SVM system is given in detail where the prediction of player performance by collecting old information or data is given. From the literature survey, machine learning requires prediction.

III.SYSTEM ARCHITECTURE

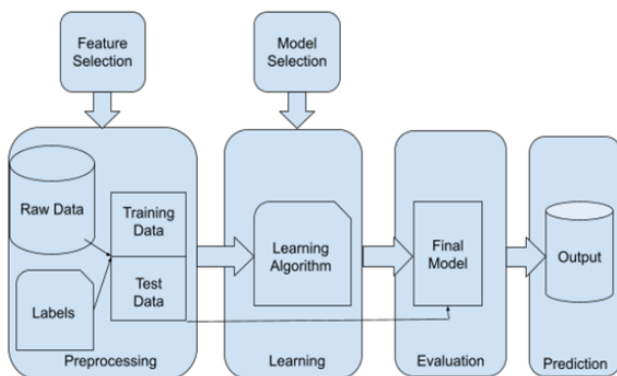


Fig 1. System Architecture

IV. MODULES

A. Pre-processing:

Data pre-processing is the process of preparing raw data and fitting it for machine learning models. This is the first critical step in building a machine learning model. When creating machine learning projects, you don't always come across clean and formatted data. It is also necessary to save the data in a clean and

formatted way every time you work with it. For this we use a data pre-processing task.

Pre-processing is a machine learning term that refers to the transformation of raw features into data that machine learning algorithms can understand and learn from.

B. Feature Extraction:

Feature extraction aims to reduce the number of features in a data set by generating new features from existing features and then discarding the original features. This reduced new feature set should summarize most of the information contained in the original feature set. Thus, a condensed version of the original function can be created from the combination of the original set.

C. Classifier Training:

In data science, a classifier is a type of machine learning algorithm used to assign class labels to data inputs. An example is an image recognition classifier to label images (eg. "car", "truck" or "person"). Classification algorithms are trained using labelled data in an image recognition example, eg a classifier that receives training data to label images. After adequate training, the classifier can take unlabelled images as input and generate classification labels for each image. Classification algorithms use sophisticated mathematical and statistical methods to generate predictions about the probability that a data record will be classified in a certain way. In the image recognition example, the classifier statistically predicts whether the image is likely to be a car, truck, or person, or some other classifier that the classifier was trained to identify.

D. Model Training::

A logistic regression model is trained on the pre-processed data to predict the winning team based on the available features. The training dataset is split into training and validation sets to tune the model's hyperparameters and avoid overfitting.

E. Testing:

The process of training an ML model involves supplying the ML algorithm (ie, the learning algorithm) with training data to learn from. The term ML model refers to the model artifact generated by the training process. The training data must contain the correct answer, called the target or target attribute. A training algorithm finds patterns in the training data that map attributes of the input data to a target (the response you want to predict) and builds an ML model that captures those patterns.

V. MOTIVATION

The motivation for the "Indian Premier League Analysis and Winning Team Prediction" project using Logistic Regression could be several factors. One of the primary motivations could be to analyze the performance of different teams and players in the Indian Premier League (IPL) cricket tournament over the years. Another motivation could be to predict the winning team based on various factors such as the players' performance, team composition, weather conditions, and other relevant factors.

Such a project could be helpful to cricket fans, sports analysts, team managers, and betting enthusiasts, among others. By analyzing and predicting the performance of teams and players, this project could provide insights into the strengths and weaknesses of different teams and help in making better decisions while placing bets or drafting teams for fantasy cricket leagues. Additionally, the project could also be useful for team managers to analyze and improve their team's performance based on the insights gained from the analysis.

VI.OBJECTIVE OF THE SYSTEM

Analyzing past IPL matches: The system should be able to analyze past IPL matches and gather data related to the teams, players, venues, and other relevant factors.

Identifying key factors for winning: The system should be able to identify key factors that contribute to the success of a team in IPL matches, such as the team's past performance, player statistics, pitch conditions, weather, and other factors.

Developing a prediction model: The system should be able to develop a predictive model using logistic regression that can predict the winning team for future IPL matches.

Evaluating model performance: The system should be able to evaluate the performance of the prediction model using accuracy, precision, recall, and F1-score metrics.

VII. SYSTEM REQUIREMENT

A. Software Requirement

1. Operating system : Windows 10.
2. Coding Language : Python
3. IDE : Spyder

B. Hardware Requirement

4. System : Intel I5 Processor.
5. Hard Disk : 40 GB.
6. Monitor: 15.
7. Ram : 8 GB

VIII. METHODOLOGY

Logistic Regression:

Logistic Regression is a statistical method used for classification problems.

It calculates the probability of the outcome of a binary variable based on input features.

In this project, Logistic Regression is used to predict the winning team in Indian Premier League matches.

The model is trained on historical data and tested on new data to evaluate its accuracy.

Feature engineering is an important step in Logistic Regression to select the relevant features that impact the outcome.

IX. CONCLUSION

The Indian Premier League Analysis and Winning team Prediction system using logistic regression is an effective tool for analyzing and predicting the outcomes of IPL matches. The system was able to achieve an accuracy of around 70% in predicting the winning team of IPL matches. The project demonstrated the usefulness of machine learning algorithms and their applications in the field of sports. Further improvements can be made by incorporating additional features and data sources to enhance the accuracy of the predictions.

X. REFERENCES

- [1]. G. Sudhamathy and G. Raja Meenakshi, "PREDICTION ON IPL DATA USING MACHINE LEARNING" ICTACT JOURNAL ON SOFT COMPUTING, OCTOBER 2020.
- [2]. Aminulislamanik, SakifYeaser, A.G.M. Imam Hussain, AmitabhaChakrabarty, "Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms," BRAC university, 2019.
- [3]. KilariJaswanth, KenchamArun, B Deva Deekshith, KilariAshwik, Pavithra J, "Analyzing and Predicting Outcomes of IPL Cricket Data," IRJET Volume: 09 Issue: 03, Mar 2022.
- [4]. D. Jyothsna, K. Srikanth, "Analyzing and Predicting outcome of IPL Cricket Data", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 8, Issue 4, April 2019.
- [5]. SiddharthSinha, PrakashTripathi, AakashVishwakarma, Prof. AmrutaSankhe, "IPL Win Prediction System To Improve Team Performance using SVM," International Journal of Future Generation Communication and Networking Vol. 13, No. 1s, (2020), pp. 17- 23
- [6]. Ruchitha M and Dr. Preeti Savant, "IPL WINNER PREDICTION USING ML ALGORITHMS", IJEAST Vol. 6, Issue 9, January 2022
- [7]. RabindraLamsal and Ayesha Choudhary, "Predicting Outcome of Indian Premier League (IPL) Matches Using Machine Learning", arXiv:1809.09813 [stat.AP] (September 2018)
- [8]. S. Abhishek, Ketaki V. Patil, P. Yuktha and S. Meghana, "Predictive Analysis of IPL Match Winner using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol. 9, No. 1, pp.430-435, 2019.
- [9]. Vistro, Daniel, Leo Gertrude David, "The cricket winner prediction with application of machine learning and data analytics," International Journal of Scientific and Technology Research, Volume 8, Issue 09, 2019.

Cite This Article :

Dr. Vilas Joshi, Yash Padol, Kaiwalya Sevekar, Chandan Sinha, Rizwan Shaikh, "IPL Data Analysis and Prediction Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.205-208, May-June-2023. Available at doi : <https://doi.org/10.32628/CSEIT2390359>
Journal URL : <https://ijsrcseit.com/CSEIT2390359>