Score Prediction of IPL Dataset

M Vignesh(1NT18CS089)
Nalin Raj(1NT18CS105)
Kumara H (1NT18CS081)

Abstract

Indian Premier League (IPL) is the most popular sports event in India with huge amount of cash flow on and off the pitch. As IPL gains more popularity, it attracts more endorsement for each team. So, prediction of a match result is very much important for the sponsors as it will help them to put their investment in good hands. In this study we build a model by which it has been shown that for a particular team batting first, scoring as many runs as possible with a very good strike rate and a very good attacking bowling.

Introduction

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. Machine learning algorithms use historical data as input to predict new output values. The proposal of our project is to predict the scores of a team batting first based on the previous years observations using various models such as Decision Tree Regression, Linear Regression, Support Vector Regression.

Data Set

The dataset is chosen from Kaggle website which consisting of set of attributes such as:

• mid: Unique match id.

• date: Date on which the match was played.

• venue: Stadium where match was played.

• battingteam: Batting team name.

• bowlingteam: Bowling team name.

• batsman: Batsman who faced that particular ball.

• bowler: Bowler who bowled that particular ball.

• runs: Runs scored by team till that point of instance.

• wickets: Number of Wickets fallen of the team till that point of instance.

• overs: Number of Overs bowled till that point of instance.

• runslast5: Runs scored in previous 5 overs.

• wicketslast5: Number of Wickets that fell in previous 5 overs.

• striker: max(runs scored by striker, runs scored by non-striker).

• non-striker: min(runs scored by striker, runs scored by non-striker).

• total: Total runs scored by batting team at the end of first innings.

The challenge is lot of information is available which may not be necessary to predict our model we need to identify such attributes and trim them.

Machine Learning Methods

1)Decision Tree Regression:

A Decision Tree has many analogies in real life and turns out, it has influenced a wide area of Machine Learning, covering both Classification and Regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. Based on the fitting data we are going to calculate the accuracy and error in the model.

2)Linear Regression:-

linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. We are going to choose the dependent variable and predict the accuracy and error in the model.

3)**Support Vector Machine Regression:**

Support Vector Machines (SVMs) are well known in classification problems. We are going to calculate the accuracy and error in the model.

Finally, we choose a model with less error and more accuracy.

Assessment:-

Train and test split
k-Fold Cross-Validation

Presentation and Visualization

We will be showing both graph and prediction output.

Roles

Each one of us have dividing the equally.my role to collect the dataset trim the un-required attribute and perform some pre-processing of data, train the decision tree regression model calculate the accuracy and errors in the model. Team-mate role is to train the other two model and calculate the accuracy and error in the model. Finally ,we both can come to an conclusion that which model is has better accuracy and less error.

Schedule:

Date	Task to be completed
23/12/2021	Download and observe the dataset.
28/12/2021	Learn about the model.
05/01/2022	Execution of the model and predicting the output.
11/01/2022	Based on the accuracy, choosing the best model for Prediction.

Bibliography

- **https://www.kaggle.com/yuvrajdagur/ipl-dataset-season-2008-to-2017**
- **https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2**