IPL PREDICTION USING MACHINE LEARNING

Cricket, especially the twenty20 format, has maximum uncertainty, where a single over can completely change the momentum of the game. With millions of people following the Indian Premier League, therefore developing a model for predicting the outcome of its matches beforehand is a real-world problem. A cricket match depends upon various factors, and in this work various feature selection methods were used to reduce the number of features to 5 from 15. Player's performance in the field is considered to find out the relative strength of the team. A Linear Regression based solution is proposed to calculate the weightage of a team based on the past performance of its players who have appeared most for the team. Finally, a dataset with the features: home team, away team, stadium, toss winner, toss decision, home team weightage and away team weightage, is fed to a different model classifiers to train the model and make prediction on unseen matches. Classification results are satisfactory. Problem in the dataset and how the accuracy of the classifiers can be improved is discussed.

PROJECT IDEA

Since the dawn of the IPL in 2008, it has attracted viewers all around the globe. High level of uncertainty and last moment nail biters has urged fans to watch the matches. Within a short period, IPL has become the highest revenue generating league of cricket. Data Analytics has been a part of sports entertainment for a long time. In a cricket match, we might have seen the score line showing the probability of the team winning based on the current match situation.

Being a cricket fan, visualizing the statistics of cricket is mesmerizing and amazing. Being a Data Scientist, it is also equally important to understand the way it works, tries to understand the science behind in it and to apply it in real life.

In Machine Learning, the problems are categorized into 2 groups mainly: Regression Problem and Classification problem. The Regression problem

deals with the kind of problems having continuous values as output while in the Classification problem the outputs are categorical values. Since the output of winner prediction is a categorical value, the problem which we are trying to solve is a Classification problem.

So where to start, and what to do?

1. Understand the dataset.
2. Clean the data.
3. Analyse the candidate columns to be Features.
4. Process the features as required by the model.
5. Train the model on training data.
6. Test the model on testing data.
7. Getting results for different model classifiers.
8. Predicting the outcome of match based on input variables.