We have analysed the dataset of the matches of IPL for every year. We have also applied different classification algorithms to predict the winner team of a particular IPL match based on factors such as venue of the match, weather, toss winner etc. This information can be used by gamblers or team owners to place bets on matches or analyse the team performance.

For making predictions, the data is preprocessed at first. It includes filling the missing values in the dataset and removing the records with incorrect data. The labels of the dataset are then encoded into numbers to enable the application of different classification algorithms on it. The dataset is then split into training and test set.

The dataset is tested for different algorithms at by taking different combination of features to analyse the predictability of the different features.

We tested the data using four algorithms, namely Naive Bayes, SVM, Decision Tree and Random Forest Classifier. After applying the algorithms for different combination of features, it was found that only five variables i.e, team 1, team 2, venue of match, toss decision, and stadium have the strongest impact on predicting the winner of the match.

Each algorithm was applied on the dataset 10 times and the average of the accuracy was taken. The accuracies obtained for different algorithms used are:

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Random Forest classifier | 68.3% |
| Decision Tree classifier | 54.63% |
| Naive Bayes Classifier | 22.48% |
| SVM classifier | 40.625% |

It was found that Random Forest Classifier provides best level of accuracy for the given data.

A table of the match outputs based on the participating teams, toss decision and venue along with the stadium is obtained.

