Capstone Project

Project Title: Analysis and Predicting Results of IPL T20 Matches

Abstract:

The craze of Indian Premier League (IPL) is always there in millions of people including Indian and our work is about the analysis of the data and prediction of the IPL matches. IPL Data Analysis is all about the analysing the data that is already present in data set using data science, machine learning and python. This is an application design for the purpose of analysing the data by fetching the attribute from the data set and predicting the future of the match and as well as of the players. This will help in the selection of the IPL team that the team should perform good and win the match. Prediction is done for anything like which player will play well in tomorrow's match, which team will win toss and even match etc. The prediction can be done with the help of the analysis on that data set collected and by displaying proper data that is useful for the future prediction. The algorithms have given accuracy over 80%.

Capstone Project –

IPL WIN PROBABILITY PREDICTION

Introduction

Creating a legendary milestone in the world of a sport like cricket, the BCCI introduced Indian Premier League (IPL). As the volume of cricket enthusiasts across the globe increases, the curiosity on prediction of the future matches led to creation of many websites like Dream11, IPL fantasy league etc. IPL predictions can be right or wrong. Sports analytics is not new in cricket. Cricket being an attractive and profitable sports, there are so many stakeholders, involved in these sports. That's why decision-making process is very critical in cricket. Analytics is a supporting pillar for every decision maker in this game. The IPL predictor visualizes useful insights and predicts outputs for instances provided by the user. The product caters to the needs of sports analysts, broadcasters, sports enthusiasts, business corporates, team management and so on . The objective of this research is to observe impact of different Machine learning models in Prediction of an IPL match. Another objective of this study is to explore information, pattern related to Matches, Player etc. using descriptive analysis so as to increase the decision-making effectiveness. The main objective of our project is to analyse the IPL statistics of various matches and teams. Dashboards are created for individual teams to provide an insight of their performance from the beginning of IPL.

DATASET DESCRIPTION

The section describes the IPL dataset providing information on team performance from 2008 to 2019. The section provides details on the attributes, instances, missing values in the dataset.

Number of attributes: 18

Number of instances: 756

ID –The attributes contain the information about the unique id for a match.

SEASON –The attribute contains the information about the year when the match has been conducted.

CITY - The attribute holds the information about the city where the match took place.

DATE – The attribute holds the information about the date when the match has been held.

TEAM 1 – The attribute describes that which team is going to bat first.

TEAM 2 – The attribute describes that which team is going to bat second.

TOSS_WINNER – The attribute holds the information about who wins the toss in that match.

TOSS_DECISION – The attribute contains the information about the decision (bat/field) taken by the toss winner. RESULT – The attribute contains information about the result (normal/tie) of the players.

DL_APPLIED – The attribute describes whether the. Duckworth Lewis (DL) rule is applied.

WINNER – The attribute holds the information about the winner of the match.

WIN_BY_RUNS – The attribute describes that which team had win by runs.

WIN_BY_WICKETS – The attribute describes that which team had win by wickets.

PLAYER_OF_MATCH – The attribute contains information about the man of the match.

VENUE – The attribute contains information about in which place the match has been played.

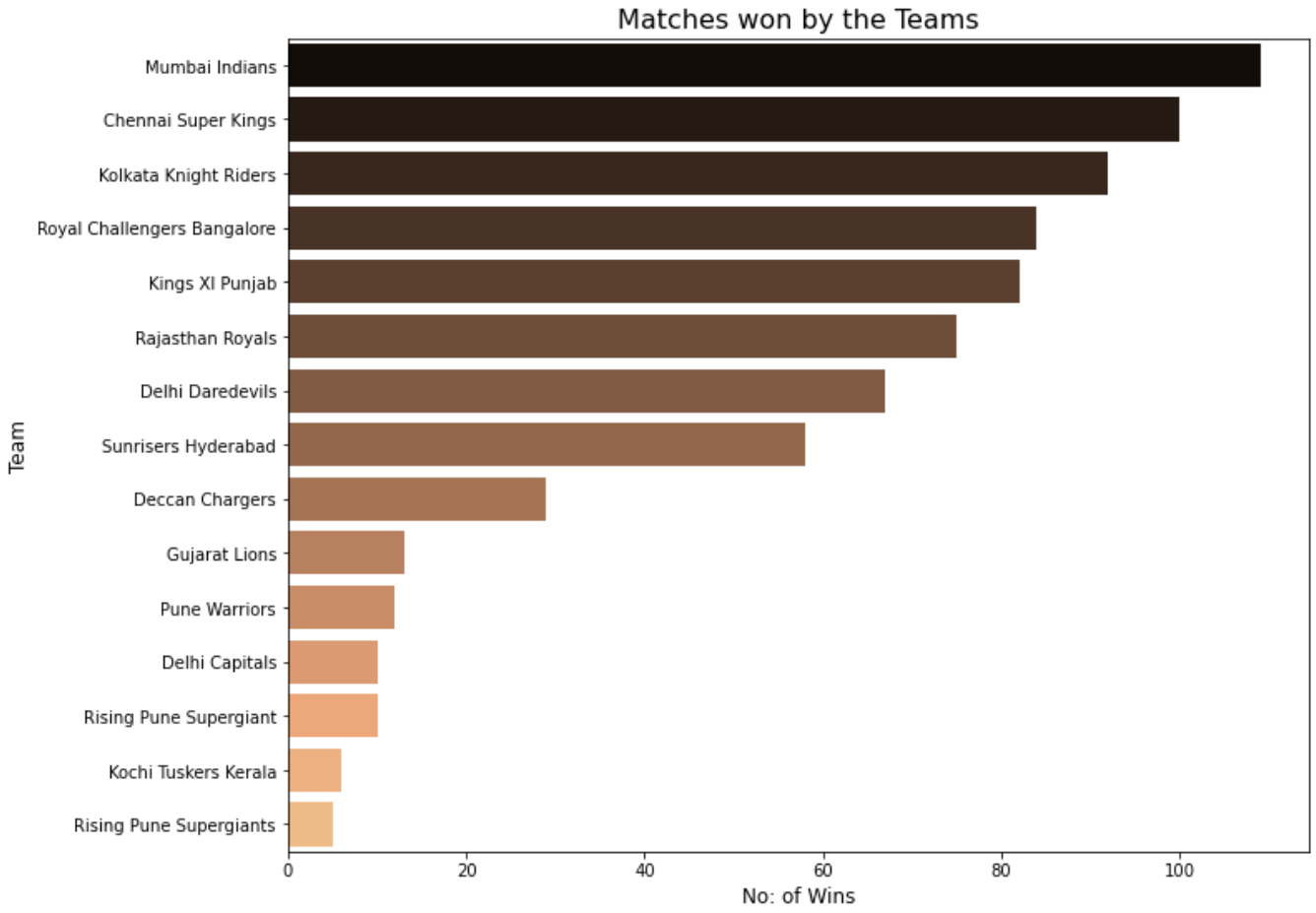
UMPIRE 1 – The attribute contains information about the names of the umpire 1.

UMPIRE 2 – The attribute contains information about the names of the umpire 2.

UMPIRE 3 – N/A.

Data visualization

Analyzing and visualizing the data gives us insights about the data and helps us to make future decisions based on the data. Once the data is preprocessed it is ready for visualizing and finding some interesting facts and hidden insights from the data. From 2008 until 2019 756 matches were played.

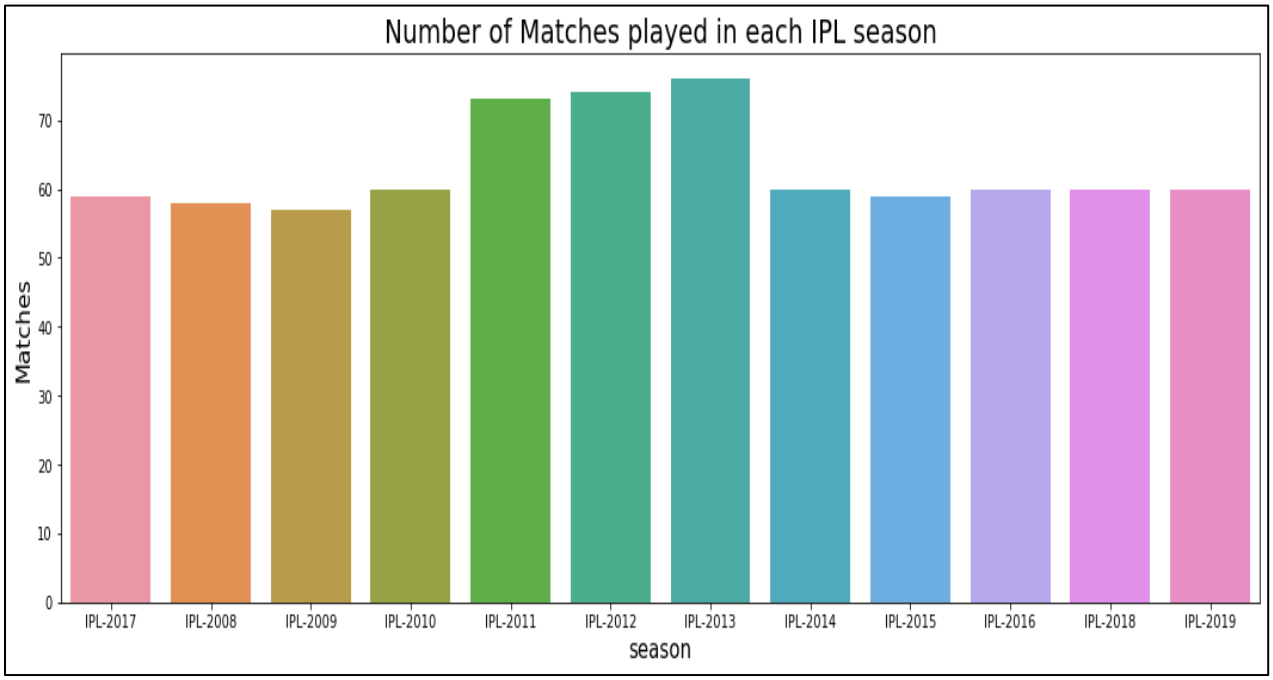


* Mumbai Indians has maximum number of winning matches followed by Chennai Super Kings.

* In match DataFrame, "city" column has 32 unique values while "venue" column has 41 distinct values.

* Let's find out which city has many numbers of venues.

During 2008 to 2010 there were less than 60 matches played in every season. From 2012-2013 there were about 74 matches played and in 2014 it reduced to 46 and from 2015 to 2019 the number of matches played were less than 60 as it was in the initial seasons. It can be easily observed from the visual below.



Steps performed

1. Installing the necessary packages: ----

a. Numpy

b. Pandas

c. Matplotlib

d. Seaborn

e. Sklearn

2. Fetching the required packages from the library

3. Loading the data in Python from your working directory

The data matches.csv is a windows comma separated value (csv) file that contains 18 varia -bles and 756 observations.

4. Data Preprocessing

Merging of Match and delivery dataframe

Current Score

Runs Left

Balls Left

Current Run Rate

Required Run Rate

Wickets

Result

Extraction of relevant features and getting final dataframe

5. Data Cleaning

Dropping Null values

Removing outrageous values from rrr(required run rate) col

Shuffling of final_df

6. Model Building

Splitting of Training and Testing Data

Column transformation using OHE

Build Pipeling

Predicting accuracy

Predicting Probability score

7. Analysis of Match Progression

Plots regarding match progression

BUSINESS / SOCIAL IMPACT

As I am forecasting the predictive analysis of IPL, it will help the Sponsors to know whom they can sponsor and which players they can buy in auction.

CONCLUSION

Analytics can be used for Cricket match Prediction and its analysis in very easy way. For IPL game, Teams, Venue, Winning Toss, Venue of the Match and Decision after winning the toss are important influencers to win a match. Different Machine Learning helps to predict outcome of a match. Right selection of Machine Learning Model helps to increase Accuracy of Prediction. With this we can predict the IPL match through machine learning models.