IPL Cricket Match Result Predictions

Ammar bin Ayaz (aba450) Aswathy Mohan (am9094) Rahul (rr3687) Sesha Sai Sreevani Kappagantula (ssk785)

1. Introduction

Since the dawn of IPL in 2008, it has attracted viewers all around the globe. The technologies of Hadoop and big data can lend a significant support in predicting the possibilities and results in matches. Sports Analytics is a method of collecting and analyzing historical game information to derive essential knowledge from it, with the aim that it will promote successful decision-making.
In this project, the past data of IPL containing the players' details, match venue details, teams, ball to ball details, is taken and analyzed to draw various conclusions which help in the improvement of a players performance and predict outcomes.

2. Cricket Database

The required data (around 30MB) is scraped using BeautifulSoup which is a python library used for web scraping. The data typically includes details of player like matches played, batting innings, not outs, runs scored, highest score, average score, balls faced, strike rate, hundreds, fifties, fours, sixes, bowling innings, balls bowled, economy, bowling strike rate and runs conceded.

3. Technologies used

Apache Hadoop, Apache Spark, MongoDB, Pyspark, SparkML, React, Python

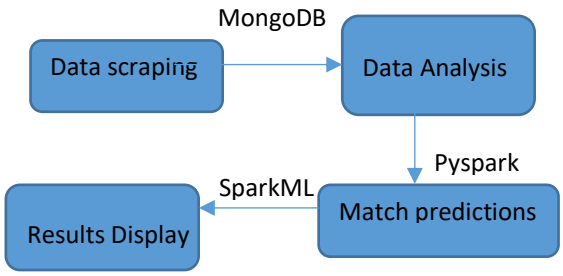
4. Project Architecture

The project architecture consists of scrapping the cricket data needed, analyzing the data collected, making match predictions and displaying the results.

For Data scraping, we will be using BeautifulSoup library and extract the required player and match features.

For the data analysis part, we plan to use Pyspark. The predictions will be displayed using the front end technologies.

For the pedictions we will be using suitable Machine learning techniques with SparkML.



5. Goals

-Scrapping the data and cluster players based on the features collected. We will be using separate clustering criteria for bowlers and batsmen while performing the step.

-Simulating the entire IPL match using ball by ball data and clustered players from the previous step. We will be trying to calculate player vs player probability. If the players haven't played before then we will be using clustering probability.

-Each innings is simulated and batsmen/bowlers are interchanged for every 6 balls. Selection of bowlers will also be decided.

-Train a Machine Learning classifier and predict the results. The predictions consist of the number of runs scored by each batsman, the number of wickets taken by each bowler and the final scores for the match.

5. References

[1]. https://www.dexlabanalytics.com/blog/how-stat-this-ipl-season-embrace-big-data-analysis-and-predict-it-right

[2]. Analyzing and predicting outcome of IPL cricket data, International Journal of Innovative Research in Science, Engineering and Technology - Vol. 8, Issue 4, April 2019

[3]. Predictive Analysis of IPL Match Winner using ML, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2S, December 2019