Big Data Analysis on Indian Premier League

Suhas Chikkanaravangala Vijayakumar
Mihir Manjrekar
Gautam Gadipudi

ABSTRACT

abstract goes here.

1. INTRODUCTION

Sports Analysis is game changer in the world of sports, instead of relying on institution and experience, we can now use different stats acquired by monitoring player performance and compare it in different game situations for which a player with particular stats will be ideal. We plan on using different analysis techniques on a game called Cricket, we plan on doing this analysis from data obtained from Indian Premier League(IPL) a professional cricket league with 8 teams from 8 different cities. This analysis can be used for building a team around core players and also it can give insights on features influencing IPL business.

The analysis on player's performance could be used to find his/her role in the match, metrics like bowling and batting are most basic metrics used to analysis player's performance, for example considering a batting metric which include finishing ability, hard-hitting ability, counter attacking ability, etc, such measures can be used to determine the role of player, a player with a good hard hitting ability indicates that the player may be able to recover runs in a match which his team is losing and a player with counter attacking ability can take the pressure off from other players. Based on this data analysis team owners are able to decide what their team lacks in and bid for players which can fill up that gap, in addition to that teams can also analyze other teams players to decide what kind of strategy they will bring in to the game and different roles of a each player.

we plan to combine two different datasets, The first data set consists ball by ball statistics of 756 matches played across all IPL seasons in yaml format[1].The second data set contains Player and club statistics of all seasons in csv format[2]. Our analysis focuses to mine two things, one of them is to mine the most successful batting partnerships, partnerships which have scored more runs and partnerships which have finished the games as this analysis could help IPL clubs buy players who have better understanding in the field which is instrumental for winning big matches. The second data mining technique focuses on clustering players into different Tiers based on their performances, which could help clubs manage their budget in auction.

Along with data mining techniques mentioned above, we also plan to visualise three attributes, runs scored, wickets taken and boundaries scored over 12 IPL seasons, through a time series plot, as this visualization could help improve IPL business, which they could use to check weather these attributes has a direct impact on Television Rating Point(TRP), number of tickets sold in an IPL season. At last we plan to do a pair wise comparison on teams performance in league stage and in playoffs. Performance is calculated based on different attributes like runs scored in power play and in death overs etc.

Section 2 describes the dataset section 3 presents a motivation for this project, and section 4 describes the work done in sport analysis 5 discusses the implementation 6 presents the result. Section 7 describe the current state of the project and what else could be done in the future.

2. DATASET

There are primarily 2 datasets, one is based on matches[1] which consists 181,440 instances and the other is based on players[2] which consists of 562 instances. The players dataset initially consisted of basic player information such as age, batting style and bowling skill, it was then updated with new attributes like runs scored and balls faced extracted from matches dataset[1]. Two new datasets partnerships and team progression were extracted from primary sources[1][2]. Partnerships data consists of batting partnership attributes such as runs scored by each partner, number of balls faced by each partner,strike rate of each partner and total runs scored by the partnership between 2 players. Team progression data consists of progression of teams score at each over in the match, it also includes other attributes like home team, away team and season. The Partnerships data consists of 9925 instances which can be clustered using total runs and strike rate of the partnership from which player-pairs can be clustered into different tiers. Similarly team progression dataset which consists 28654 instances can be clustered based on teams score and overs.

3. MOTIVATION

4. RELATED WORK

5. IMPLEMENTATION

6. RESULTS

7. FUTURE WORK

8. REFERENCES

[1] Indian Premier League data. `https://cricsheet.org/downloads/`. Accessed: 2020-06-08.

[2] IPL Player stats. `https://data.world/raghu543/ipl-data-till-2016-set-of-csv-files`. Accessed: 2020-06-08.