**IPL PLAYER ANALYSIS**

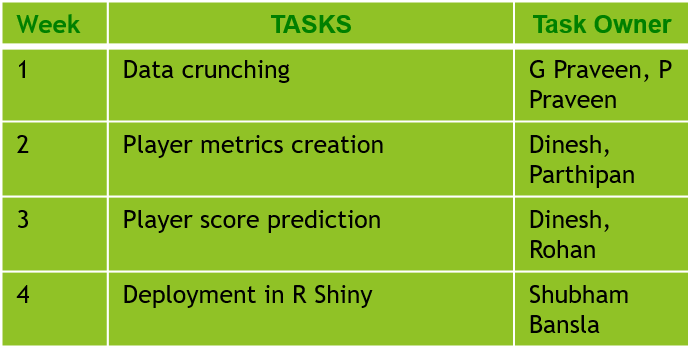
**By Shubham Bansla| Rohan Isaac| Dinesh**

1. **INTRODUCTION**

**IPL PLAYER SCORING**

G Praveen, Parthipan, Praveen P, Rohan, Sai Dinesh, Shubham Bansla

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Project Description**   1. **Problem Statement**   To rank the best 11 players of the last three seasons 2015, 2016, 2017.   1. **Benefits**   It will help to rank the players who performed well in last three seasons.   1. **Key Organizations**   In auction, it will help franchises to buy the players based on the ranking to build a better team. | **Risk and Mitigation**   |  |  | | --- | --- | | **Challenges and Risk** | **Mitigation Plan** | | Collecting data since data is not available directly. | Created master dataset with required metrics from four different datasets | | Prediction may not be accurate because some players didn’t play in some seasons. | Scores are predicted based on the No. of seasons each player played. | |
| **Project Plan**   |  |  |  | | --- | --- | --- | | **Week** | **Tasks** | **Task Owner** | | 1 | Data Crunching | Rohan Isaac, Dinesh | | 2 | Player Matrices Creation | Shubham Bansla, Dinesh | | 3 | Player Score Prediction | Rohan Isaac, Dinesh | | 4 | Deployment in R Shiny | Shubham Bansla | | **Solution Strategy**  **Prediction:**  For each player, batting, bowling and all-rounder performance metrics and ranks are calculated.  **Techniques used:**  Random forest for Recursive Feature Elimination,  R shiny for deployment  **Final Objective:**  It will help each franchise to select the best player for their team. |



1. **Dataset Description and preparation**

* Dataset downloaded till IPL season 2017:
  + Ball\_By\_Ball.csv --- 150k x 48
  + Match.csv --- 637 x 17
  + Player.csv --- 497 x 7
  + Player\_match.csv --- 14.0k x 22
  + Team.csv --- 13 x 3
* We considered only three seasons 2015, 2016 and 2017 in our project.
* Challenges we faced:
  + No consolidated data for player level.
  + Ball by ball data has each ball data and has bowler and striker name which is tedious to separate for player level.
* Unique players count is 497. So we have come up with player level data with batting and bowling metrics.
* As per our consideration of only last 3 seasons, we dropped those players who didn’t play in these seasons from the player level metrics data.

1. **Player level dataset description**

* Batting / Striker attributes:
  + No of fours, No of sixes, Balls played by Batsman and also No of boundaries
  + No of Innings played and No of Not Out Innings played
  + Total Runs scored, Total balls played
  + No of times out and also No of Run out
* Bowling attributes
  + Runs given, No of balls bowled
  + No of wickets taken and also No of 4 wickets and 5 wickets taken
  + No of Innings bowled
* Here, we calculated these attributes with respect to each season. i.e, 2015, 2016 and 2017.
* Also, we have created the same attributes for the cumulative of these three years, as 2017 & 2016 and 2017, 2016 & 2015.
* In the final data frame for the player level dataset “df\_ipl” have 497 rows for each player with respect to seasons and cumulative seasons as:
  + 2017
  + 2016
  + 2015
  + 2017&2016
  + 2017, 2016 & 2015

1. **Player Metrics**

**4.1 Batting**

* Hard Hitting ability
  + It is the measure of scoring fours and sixes in matches by facing the minimum No of balls.
  + Hard Hitting Ability = (4\* No of Fours + 6\* No ofSixes) / Balls Played by Batsman
* Finisher
  + It gives us how frequent a batsman gets out well before match closure.
  + Finisher = No of Not Out innings / Total Innings played
* Fast Scoring Ability
  + A Fast Scorer achieves a maximum score by facing the minimum No of balls.
  + Fast Scoring Ability = Total Runs / Balls Played by Batsman
* Consistency
  + This will give the striking consistency. A consistent player with a good run average is always an invaluable asset to his team.
  + Consistency = Total Runs / No of Times Out
* Running Between Wickets
  + It tells us how many runs scored by running without any boundary balls.
  + Running Between Wickets = (Total Runs – (4\*Fours + 6\*Sixes)) / (Total Balls Played - Boundary Balls)

**4.2 Bowling**

* Economy
  + It is the measure of how many runs given per over bowled.
  + Economy = Runs given / (No of balls bowled by bowler/6)
* Wicket Taking Ability
  + It gives us how frequent a bowler gets a wicket with respect to balls bowled.
  + Wicket Taking Ability = No of balls bowled / Wickets Taken
* Crucial Wicket Taking Ability
  + This attribute helps us to find the bowler who have taken more 4 and 5 wicket hauls than the average bowler.
  + Crucial Wicket Taking Ability = No of times Four or Five Wickets Taken / No of Innings Played
* Consistency
  + This will give the bowling consistency. Consistent bowlers help a team to to get the maximum wickets while keeping runs at bay.
  + Consistency = Runs Given / Wickets Taken
* Short Performance Index
  + This parameter factors in good bowling periods demonstrated by a bowler throughout the tournament.
  + Short Performance Index = (Wickets Taken – 4\* No of Times 4 Wickets Taken – 5\* No of Times 5 Wickets Taken) / (Innings Played – No of Times Four Wickets or Five Wickets Taken)

1. **Selecting Important features**

* The often quoted index in the case of IPL is the Most Valuable Player Index (MVPI).
* MVPI for batsmen and bowlers is defined as follows.
* Batting = [(PBA/TBA)\*PRS + (PBSR / TBSR)] \* PRS

PBA - Player Batting Average

TBA - Tournament Batting Average

PBSR - Player Batting Strike Rate

TBSR - Tournament Batting Strike Rate

PRS - Runs Scored by Player

* Bowling = [(TBWA/PBWA) + (TBWER/PBWER)\*2 ] \* PWT

PBWA - Player Bowling Average

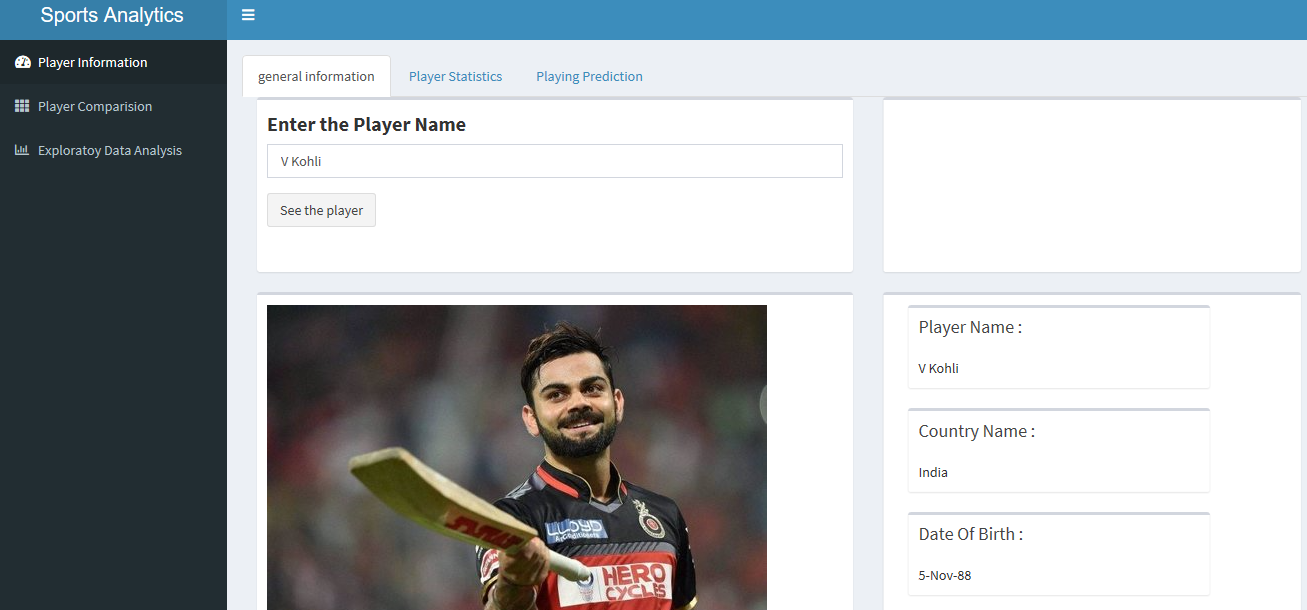
TBWA - Tournament Bowling Average

PBWER - Player Bowling Economy Rate

TBWER - Tournament Bowling Economy Rate

PWT - Player Wickets Taken

1. **R Shiny Sample Output**

****

* 1. Player Information
* General Information

In this section basic information of the player like player name, country name, date of birth etc. is shown.

* Player statistics

In this section variables like no of matches played no of 50’s, no of 100’s are shown.

* Playing Prediction

Indices which we calculated earlier based on that player rank will come like overall rank, batting rank, bowling rank etc.

* 1. Player Comparisons

We are comparing two players in this section according to their various statistics through graphs.



* 1. Exploratory Data Analysis of the Player

In this section we are analysis the data by putting the question like Number of Matches Played, Number of Matches won by each teams, Is winning the toss really an advantage, Batting Hand 5. Bowling Skills etc.

1. **References**

* IPL T20 –

<https://www.iplt20.com/>

* 360° Player Analysis –

<https://www.formcept.com/blog/360-player-analysis-now-at-your-fingertips-explore-how-formcept-is-transforming-cricket-into-the-new-face-of-big-data/>

* Dataset from Kaggle –

<https://www.kaggle.com/raghu07/ipl-data-till-2017>