Creating data repository for the game of cricket

Student Name: Saiteja Malisetty
Faculty Mentor: Mahbubul Majumder

A Description of project proposal

A.1 Abstract

Cricket is a bat-and-ball game played between two teams of eleven players on a cricket field. International Cricket Council (ICC) recognizes more than 125 countries that play cricket. To study the historical games and understand how the game has evolved in this modern form we need to explore all the games. One of the challenge is to gather data for this purpose. Fortunately the game data has been being stored in espn website called cricinfo.com. In this project we intend to build a data repository so that all the data related to cricket is readily available. We also want to create a data product so that other people can explore and query the data.

A.2 Project Description

The sport of cricket is being played since 16th century. The format of the game and techniques of the players has changed from time to time. There are many best cricketers who made their career as an inspiration to their young generations.The fusion of big data analytics and cricket is what we are looking for in this project. It empowers teams with the ability to make accurate decisions about a game and player that is known to be unpredictable. So here we are designing a dashboard in which we can compare the players with respect to the format they have played. This will be useful to know the best player played for a nation with respect to the format and era by considering the performance of fellow players of their time and the format of the game they have played. It will be also useful to estimate the performance ratio of one team towards another team.

To provide interactive environment, we intend to use the R [2] package Shiny [3]. This will allow us to create a dashboard where all the interactive tools will be present for convenient exploration of data. One such example of such a dashboard can be found in [1]. In our project we will explore the data from websites and identify the structure of data. Our challenge here is to gather all the data using an application designed in R. We also have to remove all the unwanted tables while we are gathering the data. Now we create a data repository where we will store all the data tables that are scrapped from websites. We do have to make sure that our repository is maintained well in order to store and handle huge amount of data.

A.3 Project timeline:

February 1, 2016 - June 30, 2016

Month	Task to finish
February	Explore the data at cricinfo.com and identify the data structure
March	Develop an application in R so that data can be scrapped automatically from the web. Create a
April	Design a data product to explore data online. Implement the plan using R and shiny. Testing a
May	Incorporate the suggestions obtained from feedback
June	Deliver the project and submit the report

References

[1] Mahbubul Majumder, *A Shiny application to explore USA crime data*, 2014, uRL: https://mahbub.shinyapps.io/usa-crime/.

[2] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org/

[3] RStudio, Inc, *Easy web applications in R.*, 2013, uRL: http://www.rstudio.com/shiny/.