# FOOTBALL

**OVERVIEW**

Football or Soccer is one of the most interesting game played between two teams with eleven players. International Federation of Association Football (FIFA) organizes World cups for every four years.

This project analyzes list of football players with number of goals they scored. A complete data set of Soccer players who are involved in FIFA are listed along with a detailed information on their record of matches. There is an existing application “football.db” which determines all the details about the soccer matches including but this application didn't involve in retrieving the details of a specific player. This project will focus on analyzing the goals scored by the players.

**Problem:** The previous application doesn’t visualize or compare the goals scored by two different players.

# DATA

The data provided in this data package is extracted from the following sources: <https://github.com/footballdata>

<https://github.com/openfootball/de-deutschland> <https://github.com/openfootball/eng-england>

Some of the API's are also referred:

<http://api.football-data.org/v1/soccerseasons/351/teams> <http://www.sportsdb.org/sd>

**DATA ACQUISTION**:

It is very difficult to analyze all the details of a specific player among different players in the pool.

In this review we discuss various data- dependent and independent acquisition methods. Depending upon the datasets which gives the information about the number of matches conducted every year and number of players who are involved in FIFA. The number of goals each player earned can be determined and kept in record. So the data which is dependent here is the common data for most of the players who played same number of games. The independent data is the number of goals earned by each player. Visualization is one data mining tool which can be used here. Analysis of data can be also done by creating modules in separating data. So first all the datasets should be separated and then it will be easy to collect data for individual player.

**RESEARCH QUESTIONS:**

* Analyzing each player individually.
* Tracking the number of Goals from the first match.

# PROJECT MANAGEMENT TEAM

|  |  |  |
| --- | --- | --- |
| **Team member** | **Roles and** | **Contributions** |
| Rakesh Reddy Jammula | Analyzing the data sets collecting API's. | Collected the sources of data sets. |
| Arshad Mohammed | Analyze the performance of players. | Collected all the information about each player. |

**DELIVERABLES AND CHECKPOINTS**

|  |  |  |  |
| --- | --- | --- | --- |
| Check Point Date | Expected Deliverable | Responsible Team Member | Check point Result |
| 02/15/16 | Project Proposal | Rakesh Reddy, Arshad | Overview of Project. |
| 03/08/16 | Data Acquisition and Research Inquiry | Arshad, Rakesh Reddy | Initial Inquiry and Cleaned data acquisition. |
| 04/11/16 | Type of Data Mining Technique used | Arshad, Rakesh | Analyzed using text Mining |
| 05/01/16 | Final Documentation | Arshad, Rakesh | Documented. |

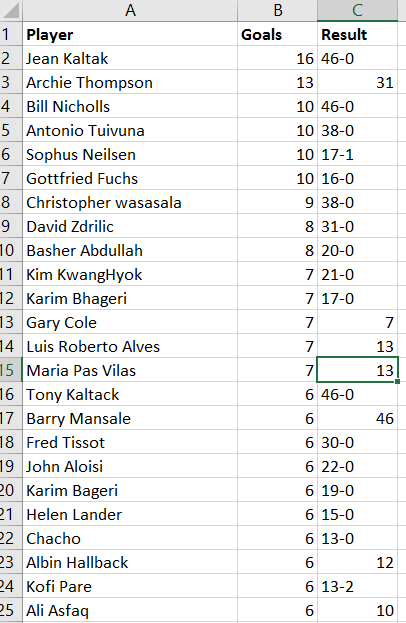
Data Mining:

Datamining hypothesis Data mining alludes to separating or "mining" information from a lot of information. It is the sort of innovation which removes the non-standard, verifiable, obscure, possibly valuable data from the expansive scale information. Information mining includes six basic classes of undertakings: i) abnormality recognition, ii) affiliation guideline learning, iii) grouping, iv) characterization, v) relapse, and vi) rundown. The affiliation standard system has been utilized as a part of game occasions examination to identify regular playing designs, since groups or players tend to display moderately stable playing designs. For instance, groups may have players who concentrate on sorting out the assaults, and advances who regularly make one on one circumstance by heading into specific zones of the assaulting range.

We have created an excel sheet which lists the details of players including how the application can be improved from its previous versions. Determined the data mining technologies need to be used for visualizing the data. We have represented them through graphs, and even took the sample screenshots of how the result looks like.

The data is collected from the open sources provided and stored in excels sheets for analyzing them. Text Mining is the data mining technology used here more analyzing the numeric data which is more important in this application.

In the next two weeks we will be submitting the excel sheets where the data has been stored and also how the text mining technology helps to evaluate the numeric data. Here are some of the screen shots and graphs:



The below graph represents the comparison of goals between the players to the total number of games played.

The below sheet will determine the comparison between Cristiano Ronaldo and Messe from 2015-2012. Between various kind of competitions, it gives a detailed information like total number of games played, goals and number of assists.



This Graph will compare the number of goals and results between few players.

Conclusion:

On overall this project will analyze the performance of players by comparing the number of goals secured to the number of games they played. Just by collecting the information of the players from the data sets and recording their entries into the excel sheets we can analyze them.

References:

1) Kuhn, Werner and Schmidt, Werner. Analyse und Beobachtung in Training und Wettkampf: Beitraege und Analysen zum Fussballsport IV. Sankt Augustin: Academia Verlag, 1991.

2) Jens. Bangsbo, Tomas. Reilly, and Charles. Hughes. Science and football. Spon Press, 1997.

3) N. Hirotsu, and M.Wright. An evaluation of characteristics of teams in association football by using a Markov process model. The Statistician. 2003, 52: Part 4, pp. 591–602

4) N. Hirotsu, and M. Wright. Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. Journal of the Operational Research Socitey. 2002, 53: 88-96.

5) C. Pan. Appliance of Apriori Algorithm on technical-tactics analysis of football. Computer Knowledge and Technology. 2010, 31(6): 8835-8837.