

Analysis Of Player Ratings Based On Intrinsic Factors To Support Team Selection

User Configuration Manual

MSc Reseach Project
Data Analytics

Sidharthan Selvaraj
x15009114

School of Computing
National College of Ireland

Supervisor: Dr.Paul Hayes

National College of Ireland
Project Submission Sheet – 2015/2016
School of Computing



Student Name:	Sidharthan Selvaraj
Student ID:	x15009114
Programme:	Data Analytics
Year:	2016
Module:	MSc Reseach Project
Lecturer:	Dr.Paul Hayes
Submission Due Date:	22/08/2016
Project Title:	Analysis Of Player Ratings Based On Intrinsic Factors To Support Team Selection
Word Count:	754

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	11th September 2016

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

1	Hardware and Software Requirements	1
1.1	Software Requirements	1
1.1.1	IBM SPSS Statistics	1
1.1.2	R Studio 64 Bit	1
1.1.3	Google Refine	1
1.1.4	MySQL	1
1.1.5	Tableau Desktop 9.2	1
1.2	Hardware Requirements	2
2	GETTING STARTED	2
2.1	How Data is extracted?	2
2.2	How the data is processed and transformed?	5
2.3	Running the model	5
2.4	Evaluating the Model	5

List of Figures

1	Data Extraction -1	3
2	Data Extraction -2	4

1 Hardware and Software Requirements

A detail overview of the software and hardware requirements to run the various artefacts are discussed in this manual.

1.1 Software Requirements

The software requirements used to run the artefacts are discussed in this section. The following softwares are utilized for this research purpose

1.1.1 IBM SPSS Statistics

Ibm Spss is a statistical modeller tool. This software is used to do the initial analysis on the cleaned dataset. For this research, descriptive statistics and bi-variate correlation are performed

1.1.2 R Studio 64 Bit

R is an open source statistical and analytical tool. In this research a majority of the artefacts are implemented using this tool. The necessary packages that need to be installed for implementing the artefacts are

1. XML
2. neuralnet
3. randomForest
4. H2O
5. caret

1.1.3 Google Refine

Google refine is a data cleaning and transformation open source tool. Every time when we attempt to open the tool we need to make sure that the bat file should be started.

1.1.4 MySQL

MySQL is a powerful tool in manipulating and transforming the large relational data sources the data. In our research, MySQL is used only to join the multiple data files which are extracted from different web sources.

1.1.5 Tableau Desktop 9.2

Tableau is one of the most powerful data visualization and analytical tool which is used in our project to evaluate individual models. Individual models evaluations and multiple model comparisons are performed.

1.2 Hardware Requirements

In this section the hardware that was used to run the artefacts are discussed. Below table shows the hardware requirements

Processor	Inter Core i7-5500U CPU 2.40Ghz
RAM	8 GB
Memory	500GB
Operating System	Windows 10

2 GETTING STARTED

In this section, We get in to step by step process on how to implement various modules of this research

2.1 How Data is extracted?

In this section, we utilize R studio for data extraction.

1. Open the installed version of R studio 64 bit version
2. Install the package XML
3. Load in the ESPN/BBC Sport/Whoscored.com link for individual team
4. Read the HTML Tables by choosing the Url data frame
5. Write the data in to csv file
6. Repeat the process for all the teams and players

Below are the screen shots showing a sample of data extraction

```

install.packages(XML)
library(XML) # Load the package
Player1_D <- "http://www.espnfc.com/player/126441/mikel-balenziaga"
Player2_D <- "http://www.espnfc.com/player/134306/eneko-boveda"
Player3_D <- "http://www.espnfc.com/player/130135/oscar-de-marcos"
Player4_D <- "http://www.espnfc.com/player/129419/xabier-etxeitia"
Player5_D <- "http://www.espnfc.com/player/159317/aymeric-laporte"
Player6_D <- "http://www.espnfc.com/player/222369/inigo-lekue"
Player7_D <- "http://www.espnfc.com/player/85300/benat"
Player8_M <- "http://www.espnfc.com/player/45927/raul-garcia"
Player9_M <- "http://www.espnfc.com/player/20497/carlos-gurpegi"
Player10_M <- "http://www.espnfc.com/player/222367/sabin-merino"
Player11_M <- "http://www.espnfc.com/player/108313/mikel-san-jose"
Player12_M <- "http://www.espnfc.com/player/109294/markel-susaeta"
Player13_M <- "http://www.espnfc.com/player/24285/aritz-aduriz"
Player14_M <- "http://www.espnfc.com/player/139419/iker-muniain"
Player15_M <- "http://www.espnfc.com/player/214412/inaki-williams"
Player16_M <- "http://www.espnfc.com/player/207718/javier-eraso"

# we need to tell the readHTMLTable function what table we want to read in.

LC_1 <- readHTMLTable(Player1_D,which=2)
LC_2 <- readHTMLTable(Player2_D,which=2)
LC_3 <- readHTMLTable(Player3_D,which=2)
LC_4 <- readHTMLTable(Player4_D,which=2)
LC_5 <- readHTMLTable(Player5_D,which=2)
LC_6 <- readHTMLTable(Player6_D,which=2)
LC_7 <- readHTMLTable(Player7_D,which=2)
LC_8 <- readHTMLTable(Player8_M,which=2)
LC_9 <- readHTMLTable(Player9_M,which=2)
LC_10 <- readHTMLTable(Player10_M,which=2)
LC_11 <- readHTMLTable(Player11_M,which=2)
LC_12 <- readHTMLTable(Player12_M,which=2)
LC_13 <- readHTMLTable(Player13_M,which=2)
LC_14 <- readHTMLTable(Player14_M,which=2)
LC_15 <- readHTMLTable(Player15_M,which=2)
LC_16 <- readHTMLTable(Player16_M,which=2)

```

Figure 1: Data Extraction -1

```

LC_10<- data.frame(PlayerPosition = c("Midfield"),LC_10)
LC_11 <- data.frame(PlayerPosition = c("Midfield"),LC_11)
LC_12 <- data.frame(PlayerPosition = c("Midfield"),LC_12)
LC_13 <- data.frame(PlayerPosition = c("Forward"),LC_13)
LC_14 <- data.frame(PlayerPosition = c("Forward"),LC_14)
LC_15 <- data.frame(PlayerPosition = c("Forward"),LC_15)
LC_16 <- data.frame(PlayerPosition = c("Midfield"),LC_16)

```

```

LC_1 <- data.frame(Playername = c("Mikel balenziaga"),LC_1)
LC_2 <- data.frame(Playername = c("Eneko boveda"),LC_2)
LC_3 <- data.frame(Playername = c("Oscar-de-marcos"),LC_3)
LC_4 <- data.frame(Playername = c("Xabier etxeitia"),LC_4)
LC_5 <- data.frame(Playername = c("Aymeric laporte"),LC_5)
LC_6 <- data.frame(Playername = c("Inigo lekue"),LC_6)
LC_7 <- data.frame(Playername = c("Benat"),LC_7)
LC_8 <- data.frame(Playername = c("Raul garcia"),LC_8)
LC_9 <- data.frame(Playername = c("Carlos gurpegi"),LC_9)
LC_10<- data.frame(Playername = c("Sabin merino"),LC_10)
LC_11 <- data.frame(Playername = c("Mikel-san-jose"),LC_11)
LC_12 <- data.frame(Playername = c("Markel susaeta"),LC_12)
LC_13 <- data.frame(Playername = c("Aritz aduriz"),LC_13)
LC_14 <- data.frame(Playername = c("Inker muniain"),LC_14)
LC_15 <- data.frame(Playername = c("Inaki williams"),LC_15)
LC_16 <- data.frame(Playername = c("Javier eraso"),LC_16)

```

```

LC_2 <- LC_2[c(-1),]
LC_3 <- LC_3[c(-1),]
LC_4 <- LC_4[c(-1),]
LC_5 <- LC_5[c(-1),]
LC_6 <- LC_6[c(-1),]
LC_7 <- LC_7[c(-1),]
LC_8 <- LC_8[c(-1),]
LC_9 <- LC_9[c(-1),]
LC_10 <- LC_10[c(-1),]
LC_11 <- LC_11[c(-1),]
LC_12 <- LC_12[c(-1),]
LC_13 <- LC_13[c(-1),]
LC_14 <- LC_14[c(-1),]
LC_15 <- LC_15[c(-1),]
LC_16 <- LC_16[c(-1),]

```

```

AtelticoBilbao <- rbind(LC_1,LC_2,LC_3,LC_4,LC_5,LC_6,LC_7,LC_8,LC_9,LC_10,LC_11,LC_12,LC_13,LC_14,LC_15,LC_16)
view(AtelticoBilbao)
write.csv(AtelticoBilbao , "AtelticoBilbao.csv", row.names = FALSE)

```

Figure 2: Data Extraction -2

2.2 How the data is processed and transformed?

Google refine is used to transform the data and merge them

1. Run the .exe file to start the refine
2. Now open this URL in a suitable browser <http://127.0.0.1:3333/>
3. Open the file in .xlsx /csv format to which needs to be transformed
4. Cluster the similar player names and modify the player names and teams
5. Save the project and export the file in a desired format either xlsx/csv

2.3 Running the model

In this section, R studio is used to run the model and predict the results.

1. Open the installed version of R studio 64 bit version
2. Install the package H2o ,neuralnet and randomForest
3. Load the train and test splits as H2O dataframe.
4. Normalize the data points in the range of 0 To 1
5. Train the model by initial using a trial and error approach by changing the ntree and hidden layers for Random Forest and neural networks respectively
6. Validate the training model
7. Predict the player rating for test data using the trained model
8. Plot the linear regression model and analyse the RMSE values

2.4 Evaluating the Model

Evaluation of the results are done in tableau 9.2.The csv file of the outputs are calculated and loaded in to tableau.

1. Open the installed version of Tableau Desktop 9.2 version
2. Load the calculated output file as excel or csv format
3. Visualize the individual model to know the performance
4. Compare the models and evaluate the performance between them