DATA MINING - FINAL PROJECT

Tennis Major Tournament Match Statistics Data Set

Name : Sharanya Dave

INTRODUCTION

I have selected a dataset on US Open Men's (2013) Statistics to predict the winner of the tournament which is classified as either player 1 or player 2. Tasks associated with this dataset are classification, Regression and clustering. These characteristics are determined with the help of a tool called Weka. Here Classification, Linear regression are used to determine the variety of network features. Classification techniques like ZeroR, NaïveBayesMultinomialText, Bagging, ClassificationViaRegression, J48 and Clustering techniques like LVQ, Filtered Clusterers.

DATA SET DESCRIPTION

Data Set Characteristic:	Multivariate	**Number of instances:**	127	**Associated Tases:**	Classification, Regression, Clustering
Attribute Characteristic:	Integer, Real	**Number of attributes:**	42	**Missing Values?**	Yes

This dataset is retrieved from the following URL:
http://archive.ics.uci.edu/ml/machine-learning-databases/00300/

Attribute Information:

Player 1 Name of Player 1
Player 2 Name of Player 2
Result Result of the match (0/1) - Referenced on Player 1 is Result = 1 if Player 1 wins (FNL.1>FNL.2)
FSP.1 First Serve Percentage for player 1 (Real Number)
FSW.1 First Serve Won by player 1 (Real Number)
SSP.1 Second Serve Percentage for player 1 (Real Number)
SSW.1 Second Serve Won by player 1 (Real Number)
ACE.1 Aces won by player 1 (Numeric-Integer)
DBF.1 Double Faults committed by player 1 (Numeric-Integer)
WNR.1 Winners earned by player 1 (Numeric)
UFE.1 Unforced Errors committed by player 1 (Numeric)
BPC.1 Break Points Created by player 1 (Numeric)
BPW.1 Break Points Won by player 1 (Numeric)
NPA.1 Net Points Attempted by player 1 (Numeric)
NPW.1 Net Points Won by player 1 (Numeric)

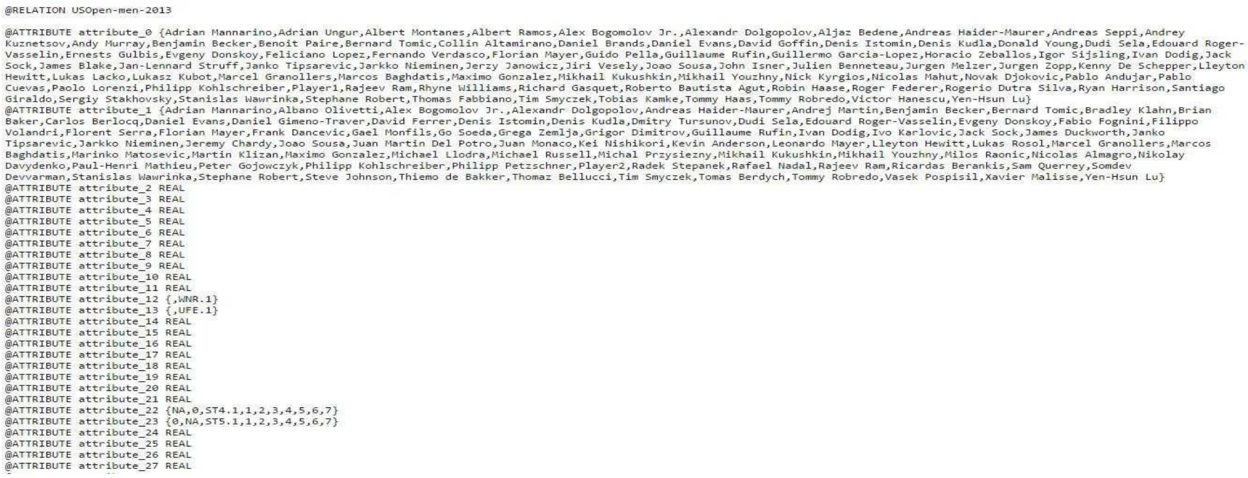
TPW.1 Total Points Won by player 1 (Numeric)
ST1.1 Set 1 result for Player 1 (Numeric-Integer)
ST2.1 Set 2 Result for Player 1 (Numeric-Integer)
ST3.1 Set 3 Result for Player 1 (Numeric-Integer)
ST4.1 Set 4 Result for Player 1 (Numeric-Integer)
ST5.1 Set 5 Result for Player 1 (Numeric-Integer)
FNL.1 Final Number of Games Won by Player 1 (Numeric-Integer)
FSP.2 First Serve Percentage for player 2 (Real Number)
FSW.2 First Serve Won by player 2 (Real Number)
SSP.2 Second Serve Percentage for player 2 (Real Number)
SSW.2 Second Serve Won by player 2 (Real Number)
ACE.2 Aces won by player 2 (Numeric-Integer)
DBF.2 Double Faults committed by player 2 (Numeric-Integer)
WNR.2 Winners earned by player 2 (Numeric)
UFE.2 Unforced Errors committed by player 2 (Numeric)
BPC.2 Break Points Created by player 2 (Numeric)
BPW.2 Break Points Won by player 2 (Numeric)
NPA.2 Net Points Attempted by player 2 (Numeric)
NPW.2 Net Points Won by player 2 (Numeric)
TPW.2 Total Points Won by player 2 (Numeric)
ST1.2 Set 1 result for Player 2 (Numeric-Integer)
ST2.2 Set 2 Result for Player 2 (Numeric-Integer)
ST3.2 Set 3 Result for Player 2 (Numeric-Integer)
ST4.2 Set 4 Result for Player 2 (Numeric-Integer)
ST5.2 Set 5 Result for Player 2 (Numeric-Integer)
FNL.2 Final Number of Games Won by Player 2 (Numeric-Integer)
Round Round of the tournament at which game is played (Numeric-Integer)

DATA PREPARATION

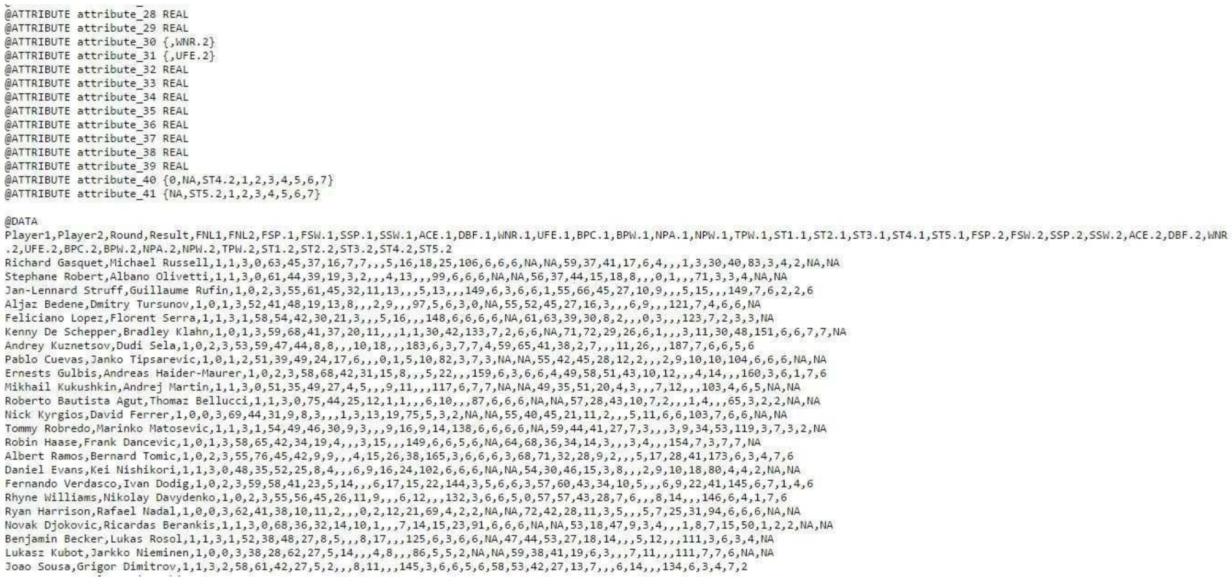
The dataset that is obtained is in the form of a spreadsheet. However the native data storage in Weka's is in ARFF format. For that we have to convert data from spreadsheet to ARFF format. The CSV file is loaded into Weka and viewed using the ARFF Viewer. Then the file is saved in the ARFF format. The ARFF file consists of attribute values for each instance which is separated by commas and list of instances. Now you just need to make an ARFF files; add dataset's name using @relation tag, attribute information using @attribute, and data information using @data line; and save the file as raw text. After loading the ARFF file on to the Weka, we have to perform different classification, regression and clustering techniques to analyze the data set.

From Weka tool open Explorer and then select your ARFF dataset by clicking on the open file button. After opening the file select the classify tab on the top. After going to the classify panel select the classifier which you want to perform to analysis the dataset. In this dataset we are using ZeroR, NaïveBayesMultinomialText, Bagging, ClassificationViaRegression, J48 and Clustering techniques like LVQ, Filtered Clusterers techniques to analyze the data.

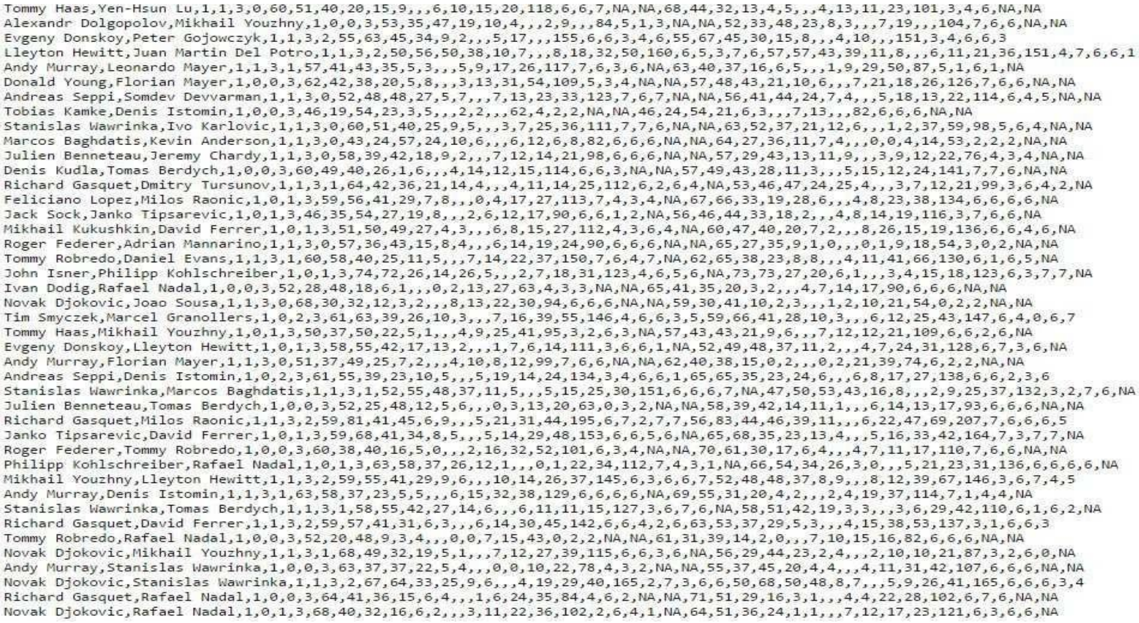
Snapshots of Data:



```
@RELATION USOpen-men-2013

@ATTRIBUTE attribute_0 {Adrian Mannarino,Adrian Ungur,Albert Montanes,Albert Ramos,Alex Bogomolov Jr.,Alexandr Dolgopolov,Aljaz Bedene,Andreas Haider-Maurer,Andreas Seppi,Andrey
Kuznetsov,Andy Murray,Benjamin Becker,Benoit Paire,Bernard Tomic,Collin Altamirano,Daniel Brands,Daniel Evans,David Goffin,Denis Istomin,Denis Kudla,Donald Young,Dudi Sela,Edouard Roger-
Vasselin,Ernests Gulbis,Evgeny Donskoy,Feliciano Lopez,Fernando Verdasco,Florian Mayer,Guido Pella,Guillaume Rufin,Guillermo Garcia-Lopez,Horacio Zeballos,Igor Sijsling,Ivan Dodig,Jack
Sock,James Blake,Jan-Lennard Struff,Janko Tipsarevic,Jarkko Nieminen,Jerzy Janowicz,Jiri Vesely,Joao Sousa,John Isner,Julien Benneteau,Jurgen Melzer,Jurgen Zopp,Kenny De Schepper,Lleyton
Hewitt,Lukas Lacko,Lukasz Kubot,Marcel Granollers,Marcos Baghdatis,Maximo Gonzalez,Mikhail Kukushkin,Mikhail Youzhny,Nick Kyrgios,Nicolas Mahut,Novak Djokovic,Pablo Andujar,Pablo
Cuevas,Paolo Lorenzi,Philipp Kohlschreiber,Player1,Rajeev Ram,Rhyne Williams,Richard Gasquet,Roberto Bautista Agut,Robin Haase,Roger Federer,Rogerio Dutra Silva,Ryan Harrison,Santiago
Giraldo,Sergiy Stakhovsky,Stanislas Wawrinka,Stephane Robert,Thomas Fabbiano,Tim Smyczek,Tobias Kamke,Tommy Haas,Tommy Robredo,Victor Hanescu,Yen-Hsun Lu}
@ATTRIBUTE attribute_1 {Adrian Mannarino,Albano Olivetti,Alex Bogomolov Jr.,Alexandr Dolgopolov,Andreas Haider-Maurer,Andrej Martin,Benjamin Becker,Bernard Tomic,Bradley Klahn,Brian
Baker,Carlos Berlocq,Daniel Evans,Daniel Gimeno-Traver,David Ferrer,Denis Istomin,Denis Kudla,Dmitry Tursunov,Dudi Sela,Edouard Roger-Vasselin,Evgeny Donskoy,Fabio Fognini,Filippo
Volandri,Florent Serra,Florian Mayer,Frank Dancevic,Gael Monfils,Go Soeda,Grega Zemlja,Grigor Dimitrov,Guillaume Rufin,Ivan Dodig,Ivo Karlovic,Jack Sock,James Duckworth,Janko
Tipsarevic,Jarkko Nieminen,Jeremy Chardy,Joao Sousa,Juan Martin Del Potro,Juan Monaco,Kei Nishikori,Kevin Anderson,Leonardo Mayer,Lleyton Hewitt,Lukas Rosol,Marcel Granollers,Marcos
Baghdatis,Marinko Matosevic,Martin Klizan,Maximo Gonzalez,Michael Llodra,Michael Russell,Michal Przysiezny,Mikhail Kukushkin,Mikhail Youzhny,Milos Raonic,Nicolas Almagro,Nikolay
Davydenko,Paul-Henri Mathieu,Peter Gojowczyk,Philipp Kohlschreiber,Philipp Petzschner,Player2,Radek Stepanek,Rafael Nadal,Rajeev Ram,Ricardas Berankis,Sam Querrey,Somdev
Devvarman,Stanislas Wawrinka,Stephane Robert,Steve Johnson,Thiemo de Bakker,Thomaz Bellucci,Tim Smyczek,Tomas Berdych,Tommy Robredo,Vasek Pospisil,Xavier Malisse,Yen-Hsun Lu}
@ATTRIBUTE attribute_2 REAL
@ATTRIBUTE attribute_3 REAL
@ATTRIBUTE attribute_4 REAL
@ATTRIBUTE attribute_5 REAL
@ATTRIBUTE attribute_6 REAL
@ATTRIBUTE attribute_7 REAL
@ATTRIBUTE attribute_8 REAL
@ATTRIBUTE attribute_9 REAL
@ATTRIBUTE attribute_10 REAL
@ATTRIBUTE attribute_11 REAL
@ATTRIBUTE attribute_12 {,WNR.1}
@ATTRIBUTE attribute_13 {,UFE.1}
@ATTRIBUTE attribute_14 REAL
@ATTRIBUTE attribute_15 REAL
@ATTRIBUTE attribute_16 REAL
@ATTRIBUTE attribute_17 REAL
@ATTRIBUTE attribute_18 REAL
@ATTRIBUTE attribute_19 REAL
@ATTRIBUTE attribute_20 REAL
@ATTRIBUTE attribute_21 REAL
@ATTRIBUTE attribute_22 {NA,0,ST4.1,1,2,3,4,5,6,7}
@ATTRIBUTE attribute_23 {0,NA,ST5.1,1,2,3,4,5,6,7}
@ATTRIBUTE attribute_24 REAL
@ATTRIBUTE attribute_25 REAL
@ATTRIBUTE attribute_26 REAL
@ATTRIBUTE attribute_27 REAL
```



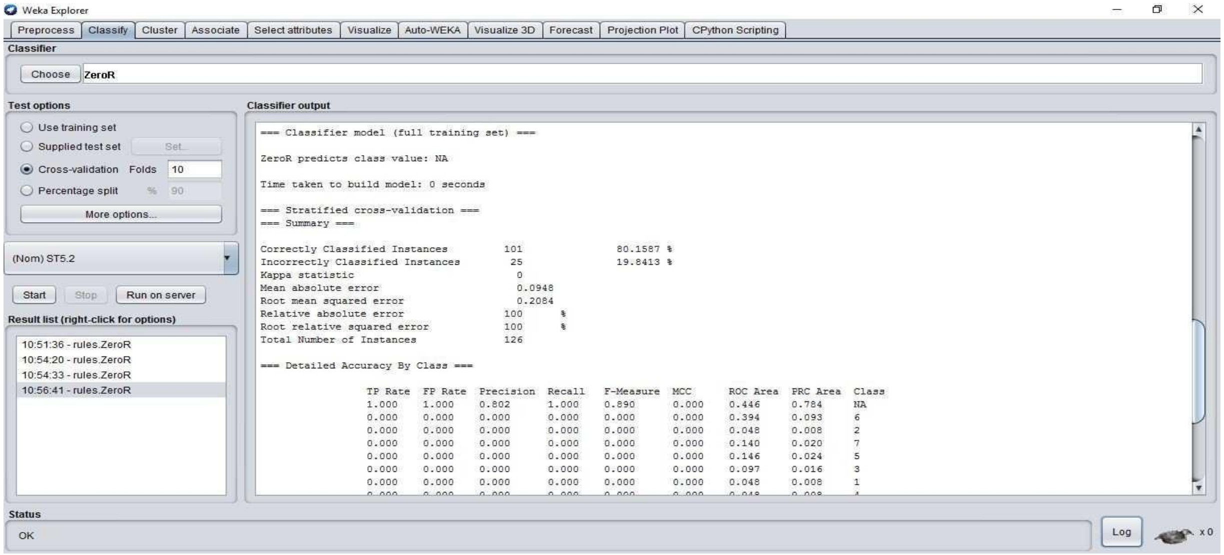
```
@ATTRIBUTE attribute_28 REAL
@ATTRIBUTE attribute_29 REAL
@ATTRIBUTE attribute_30 {,WNR.2}
@ATTRIBUTE attribute_31 {,UFE.2}
@ATTRIBUTE attribute_32 REAL
@ATTRIBUTE attribute_33 REAL
@ATTRIBUTE attribute_34 REAL
@ATTRIBUTE attribute_35 REAL
@ATTRIBUTE attribute_36 REAL
@ATTRIBUTE attribute_37 REAL
@ATTRIBUTE attribute_38 REAL
@ATTRIBUTE attribute_39 REAL
@ATTRIBUTE attribute_40 {0,NA,ST4.2,1,2,3,4,5,6,7}
@ATTRIBUTE attribute_41 {NA,ST5.2,1,2,3,4,5,6,7}

@DATA
Player1,Player2,Round,Result,FNL1,FNL2,FSP.1,FSW.1,SSP.1,SSW.1,ACE.1,DBF.1,WNR.1,UFE.1,BPC.1,BPW.1,NPA.1,NPW.1,TPW.1,ST1.1,ST2.1,ST3.1,ST4.1,ST5.1,FSP.2,FSW.2,SSP.2,SSW.2,ACE.2,DBF.2,WNR
.2,UFE.2,BPC.2,BPW.2,NPA.2,NPW.2,TPW.2,ST1.2,ST2.2,ST3.2,ST4.2,ST5.2
Richard Gasquet,Michael Russell,1,1,3,0,63,45,37,16,7,7,,,5,16,18,25,106,6,6,6,NA,NA,59,37,41,17,6,4,,,1,3,30,40,83,3,4,2,NA,NA
Stephane Robert,Albano Olivetti,1,1,3,0,61,44,39,19,3,2,,,4,13,,,99,6,6,6,NA,NA,56,37,44,15,18,8,,,0,1,,,71,3,3,4,NA,NA
Jan-Lennard Struff,Guillaume Rufin,1,0,2,3,55,61,45,32,11,13,,,5,13,,,149,6,3,6,6,1,55,66,45,27,10,9,,,5,15,,,149,7,6,2,2,6
Aljaz Bedene,Dmitry Tursunov,1,0,1,3,52,41,48,19,13,8,,,2,9,,,97,5,6,3,0,NA,55,52,45,27,16,3,,,6,9,,,121,7,4,6,6,NA
Feliciano Lopez,Florent Serra,1,1,3,1,58,54,42,30,21,3,,,5,16,,,148,6,6,6,6,NA,61,63,39,30,8,2,,,0,3,,,123,7,2,3,3,NA
Kenny De Schepper,Bradley Klahn,1,0,1,3,59,68,41,37,20,11,,,1,1,30,42,133,7,2,6,6,NA,71,72,29,26,6,1,,,3,11,30,48,151,6,6,7,7,NA
Andrey Kuznetsov,Dudi Sela,1,0,2,3,53,59,47,44,8,8,,,10,18,,,183,6,3,7,7,4,59,65,41,38,2,7,,,11,26,,,187,7,6,6,5,6
Pablo Cuevas,Janko Tipsarevic,1,0,1,2,51,39,49,24,17,6,,,0,1,5,10,82,3,7,3,NA,NA,55,42,45,28,12,2,,,2,9,10,10,104,6,6,6,NA,NA
Ernests Gulbis,Andreas Haider-Maurer,1,0,2,3,58,68,42,31,15,8,,,5,22,,,159,6,3,6,6,4,49,58,51,43,10,12,,,4,14,,,160,3,6,1,7,6
Mikhail Kukushkin,Andrej Martin,1,1,3,0,51,35,49,27,4,5,,,9,11,,,117,6,7,7,NA,NA,49,35,51,20,4,3,,,7,12,,,103,4,6,5,NA,NA
Roberto Bautista Agut,Thomaz Bellucci,1,1,3,0,75,44,25,12,1,1,,,6,10,,,87,6,6,6,NA,NA,57,28,43,10,7,2,,,1,4,,,65,3,2,2,NA,NA
Nick Kyrgios,David Ferrer,1,0,0,3,69,44,31,9,8,3,,,1,3,13,19,75,5,3,2,NA,NA,55,40,45,21,11,2,,,5,11,6,6,103,7,6,6,NA,NA
Tommy Robredo,Marinko Matosevic,1,1,3,1,54,49,46,30,9,3,,,9,16,9,14,138,6,6,6,6,NA,59,44,41,27,7,3,,,3,9,34,53,119,3,7,3,2,NA
Robin Haase,Frank Dancevic,1,0,1,3,58,65,42,34,19,4,,,3,15,,,149,6,6,5,6,NA,64,68,36,34,14,3,,,3,4,,,154,7,3,7,7,NA
Albert Ramos,Bernard Tomic,1,0,2,3,55,76,45,42,9,9,,,4,15,26,38,165,3,6,6,6,3,68,71,32,28,9,2,,,5,17,28,41,173,6,3,4,7,6
Daniel Evans,Kei Nishikori,1,1,3,0,48,35,52,25,8,4,,,6,9,16,24,102,6,6,6,NA,NA,54,30,46,15,3,8,,,2,9,10,18,80,4,4,2,NA,NA
Fernando Verdasco,Ivan Dodig,1,0,2,3,59,58,41,23,5,14,,,6,17,15,22,144,3,5,6,6,3,57,60,43,34,10,5,,,6,9,22,41,145,6,7,1,4,6
Rhyne Williams,Nikolay Davydenko,1,0,2,3,55,56,45,26,11,9,,,6,12,,,132,3,6,6,5,0,57,57,43,28,7,6,,,8,14,,,146,6,4,1,7,6
Ryan Harrison,Rafael Nadal,1,0,0,3,62,41,38,10,11,2,,,0,2,12,21,69,4,2,2,NA,NA,72,42,28,11,3,5,,,5,7,25,31,94,6,6,6,NA,NA
Novak Djokovic,Ricardas Berankis,1,1,3,0,68,36,32,14,10,1,,,7,14,15,23,91,6,6,6,NA,NA,53,18,47,9,3,4,,,1,8,7,15,50,1,2,2,NA,NA
Benjamin Becker,Lukas Rosol,1,1,3,1,52,38,48,27,8,5,,,8,17,,,125,6,3,6,6,NA,47,44,53,27,18,14,,,5,12,,,111,3,6,3,4,NA
Lukasz Kubot,Jarkko Nieminen,1,0,0,3,38,28,62,27,5,14,,,4,8,,,86,5,5,2,NA,NA,59,38,41,19,6,3,,,7,11,,,111,7,7,6,NA,NA
Joao Sousa,Grigor Dimitrov,1,1,3,2,58,61,42,27,5,2,,,8,11,,,145,3,6,6,5,6,58,53,42,27,13,7,,,6,14,,,134,6,3,4,7,2
```



```
Tommy Haas,Yen-Hsun Lu,1,1,3,0,60,51,40,20,15,9,,,6,10,15,20,118,6,6,7,NA,NA,68,44,32,13,4,5,,,4,13,11,23,101,3,4,6,NA,NA
Alexandr Dolgopolov,Mikhail Youzhny,1,0,0,3,53,35,47,19,10,4,,,2,9,,,84,5,1,3,NA,NA,52,33,48,23,8,3,,,7,19,,,104,7,6,6,NA,NA
Evgeny Donskoy,Peter Gojowczyk,1,1,3,2,55,63,45,34,9,2,,,5,17,,,155,6,6,3,4,6,55,67,45,30,15,8,,,4,10,,,151,3,4,6,6,3
Lleyton Hewitt,Juan Martin Del Potro,1,1,3,2,50,56,50,38,10,7,,,8,18,32,50,160,6,5,3,7,6,57,57,43,39,11,8,,,6,11,21,36,151,4,7,6,6,1
Andy Murray,Leonardo Mayer,1,1,3,1,57,41,43,35,5,3,,,5,9,17,26,117,7,6,3,6,NA,63,40,37,16,6,5,,,1,9,29,50,87,5,1,6,1,NA
Donald Young,Florian Mayer,1,0,0,3,62,42,38,20,5,8,,,3,13,31,54,109,5,3,4,NA,NA,57,48,43,21,10,6,,,7,21,18,26,126,7,6,6,NA,NA
Andreas Seppi,Somdev Devvarman,1,1,3,0,52,48,48,27,5,7,,,7,13,23,33,123,7,6,7,NA,NA,56,41,44,24,7,4,,,5,18,13,22,114,6,4,5,NA,NA
Tobias Kamke,Denis Istomin,1,0,0,3,46,19,54,23,3,5,,,2,2,,,62,4,2,2,NA,NA,46,24,54,21,6,3,,,7,13,,,82,6,6,6,NA,NA
Stanislas Wawrinka,Ivo Karlovic,1,1,3,0,60,51,40,25,9,5,,,3,7,25,36,111,7,7,6,NA,NA,63,52,37,21,12,6,,,1,2,37,59,98,5,6,4,NA,NA
Marcos Baghdatis,Kevin Anderson,1,1,3,0,43,24,57,24,10,6,,,6,12,6,8,82,6,6,6,NA,NA,64,27,36,11,7,4,,,0,0,4,14,53,2,2,2,NA,NA
Julien Benneteau,Jeremy Chardy,1,1,3,0,58,39,42,18,9,2,,,7,12,14,21,98,6,6,6,NA,NA,57,29,43,13,11,9,,,3,9,12,22,76,4,3,4,NA,NA
Denis Kudla,Tomas Berdych,1,0,0,3,60,49,40,26,1,6,,,4,14,12,15,114,6,6,3,NA,NA,57,49,43,28,11,3,,,5,15,12,24,141,7,7,6,NA,NA
Richard Gasquet,Dmitry Tursunov,1,1,3,1,64,42,36,21,14,4,,,4,11,14,25,112,6,2,6,4,NA,53,46,47,24,25,4,,,3,7,12,21,99,3,6,4,2,NA
Feliciano Lopez,Milos Raonic,1,0,1,3,59,56,41,29,7,8,,,0,4,17,27,113,7,4,3,4,NA,67,66,33,19,28,6,,,4,8,23,38,134,6,6,6,6,NA
Jack Sock,Janko Tipsarevic,1,0,1,3,46,35,54,27,19,8,,,2,6,12,17,90,6,6,1,2,NA,56,46,44,33,18,2,,,4,8,14,19,116,3,7,6,6,NA
Mikhail Kukushkin,David Ferrer,1,0,1,3,51,50,49,27,4,3,,,6,8,15,27,112,4,3,6,4,NA,60,47,40,20,7,2,,,8,26,15,19,136,6,6,4,6,NA
Roger Federer,Adrian Mannarino,1,1,3,0,57,36,43,15,8,4,,,6,14,19,24,90,6,6,6,NA,NA,65,27,35,9,1,0,,,0,1,9,18,54,3,0,2,NA,NA
Tommy Robredo,Daniel Evans,1,1,3,1,60,58,40,25,11,5,,,7,14,22,37,150,7,6,4,7,NA,62,65,38,23,8,8,,,4,11,41,66,130,6,1,6,5,NA
John Isner,Philipp Kohlschreiber,1,0,1,3,74,72,26,14,26,5,,,2,7,18,31,123,4,6,5,6,NA,73,73,27,20,6,1,,,3,4,15,18,123,6,3,7,7,NA
Ivan Dodig,Rafael Nadal,1,0,0,3,52,28,48,18,6,1,,,0,2,13,27,63,4,3,3,NA,NA,65,41,35,20,3,2,,,4,7,14,17,90,6,6,6,NA,NA
Novak Djokovic,Joao Sousa,1,1,3,0,68,30,32,12,3,2,,,8,13,22,30,94,6,6,6,NA,NA,59,30,41,10,2,3,,,1,2,10,21,54,0,2,2,NA,NA
Tim Smyczek,Marcel Granollers,1,0,2,3,61,63,39,26,10,3,,,7,16,39,55,146,4,6,6,3,5,59,66,41,28,10,3,,,6,12,25,43,147,6,4,0,6,7
Tommy Haas,Mikhail Youzhny,1,0,1,3,50,37,50,22,5,1,,,4,9,25,41,95,3,2,6,3,NA,57,43,43,21,9,6,,,7,12,12,21,109,6,6,2,6,NA
Evgeny Donskoy,Lleyton Hewitt,1,0,1,3,58,55,42,17,13,2,,,1,7,6,14,111,3,6,6,1,NA,52,49,48,37,11,2,,,4,7,24,31,128,6,7,3,6,NA
Andy Murray,Florian Mayer,1,1,3,0,51,37,49,25,7,2,,,4,10,8,12,99,7,6,6,NA,NA,62,40,38,15,0,2,,,0,2,21,39,74,6,2,2,NA,NA
Andreas Seppi,Denis Istomin,1,0,2,3,61,55,39,23,10,5,,,5,19,14,24,134,3,4,6,6,1,65,65,35,23,24,6,,,6,8,17,27,138,6,6,2,3,6
Stanislas Wawrinka,Marcos Baghdatis,1,1,3,1,52,55,48,37,11,5,,,5,15,25,30,151,6,6,6,7,NA,47,50,53,43,16,8,,,2,9,25,37,132,3,2,7,6,NA
Julien Benneteau,Tomas Berdych,1,0,0,3,52,25,48,12,5,6,,,0,3,13,20,63,0,3,2,NA,NA,58,39,42,14,11,1,,,6,14,13,17,93,6,6,6,NA,NA
Richard Gasquet,Milos Raonic,1,1,3,2,59,81,41,45,6,9,,,5,21,31,44,195,6,7,2,7,7,56,83,44,46,39,11,,,6,22,47,69,207,7,6,6,6,5
Janko Tipsarevic,David Ferrer,1,0,1,3,59,68,41,34,8,5,,,5,14,29,48,153,6,6,5,6,NA,65,68,35,23,13,4,,,5,16,33,42,164,7,3,7,7,NA
Roger Federer,Tommy Robredo,1,0,0,3,60,38,40,16,5,0,,,2,16,32,52,101,6,3,4,NA,NA,70,61,30,17,6,4,,,4,7,11,17,110,7,6,6,NA,NA
Philipp Kohlschreiber,Rafael Nadal,1,0,1,3,63,58,37,26,12,1,,,0,1,22,34,112,7,4,3,1,NA,66,54,34,26,3,0,,,5,21,23,31,136,6,6,6,6,NA
Mikhail Youzhny,Lleyton Hewitt,1,1,3,2,59,55,41,29,9,6,,,10,14,26,37,145,6,3,6,6,7,52,48,48,37,8,9,,,8,12,39,67,146,3,6,7,4,5
Andy Murray,Denis Istomin,1,1,3,1,63,58,37,23,5,5,,,6,15,32,38,129,6,6,6,6,NA,69,55,31,20,4,2,,,2,4,19,37,114,7,1,4,4,NA
Stanislas Wawrinka,Tomas Berdych,1,1,3,1,58,55,42,27,14,6,,,6,11,11,15,127,3,6,7,6,NA,58,51,42,19,3,3,,,3,6,29,42,110,6,1,6,2,NA
Richard Gasquet,David Ferrer,1,1,3,2,59,57,41,31,6,3,,,6,14,30,45,142,6,6,4,2,6,63,53,37,29,5,3,,,4,15,38,53,137,3,1,6,6,3
Tommy Robredo,Rafael Nadal,1,0,0,3,52,20,48,9,3,4,,,0,0,7,15,43,0,2,2,NA,NA,61,31,39,14,2,0,,,7,10,15,16,82,6,6,6,NA,NA
Novak Djokovic,Mikhail Youzhny,1,1,3,1,68,49,32,19,5,1,,,7,12,27,39,115,6,6,3,6,NA,56,29,44,23,2,4,,,2,10,10,21,87,3,2,6,0,NA
Andy Murray,Stanislas Wawrinka,1,0,0,3,63,37,37,22,5,4,,,0,0,10,22,78,4,3,2,NA,NA,55,37,45,20,4,4,,,4,11,31,42,107,6,6,6,NA,NA
Novak Djokovic,Stanislas Wawrinka,1,1,3,2,67,64,33,25,9,6,,,4,19,29,40,165,2,7,3,6,6,50,68,50,48,8,7,,,5,9,26,41,165,6,6,6,3,4
Richard Gasquet,Rafael Nadal,1,0,0,3,64,41,36,15,6,4,,,1,6,24,35,84,4,6,2,NA,NA,71,51,29,16,3,1,,,4,4,22,28,102,6,7,6,NA,NA
Novak Djokovic,Rafael Nadal,1,0,1,3,68,40,32,16,6,2,,,3,11,22,36,102,2,6,4,1,NA,64,51,36,24,1,1,,,7,12,17,23,121,6,3,6,6,NA
```

CLASSIFICATION & REGRESSION MODELS RESULT:

ZeroR result



Classifier: ZeroR

Test options:
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 90

(Nom) ST5.2

Result list (right-click for options):
10:51:36 - rules.ZeroR
10:54:20 - rules.ZeroR
10:54:33 - rules.ZeroR
10:56:41 - rules.ZeroR

```
=== Classifier model (full training set) ===

ZeroR predicts class value: NA

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         101               80.1587 %
Incorrectly Classified Instances        25               19.8413 %
Kappa statistic                          0
Mean absolute error                      0.0948
Root mean squared error                  0.2084
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.802      1.000    0.890      0.000    0.446     0.784     NA
                 0.000    0.000    0.000      0.000    0.000      0.000    0.394     0.093     6
                 0.000    0.000    0.000      0.000    0.000      0.000    0.048     0.008     2
                 0.000    0.000    0.000      0.000    0.000      0.000    0.140     0.020     7
                 0.000    0.000    0.000      0.000    0.000      0.000    0.146     0.024     5
                 0.000    0.000    0.000      0.000    0.000      0.000    0.097     0.016     3
                 0.000    0.000    0.000      0.000    0.000      0.000    0.048     0.008     1
                 0.000    0.000    0.000      0.000    0.000      0.000    0.048     0.008     4
```

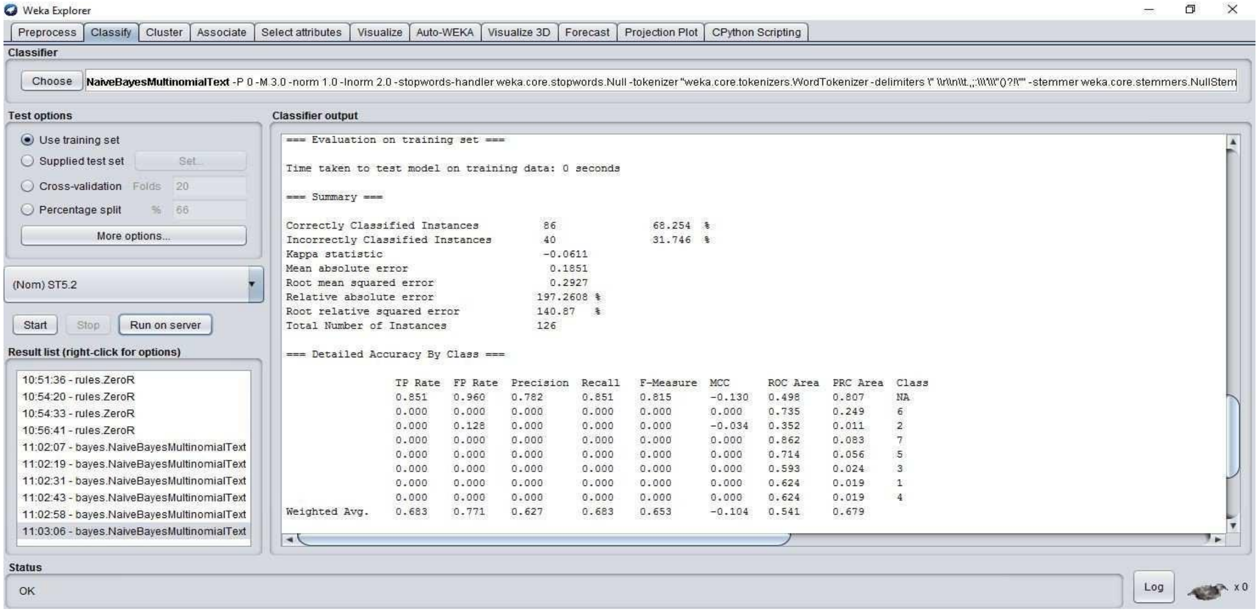
Status: OK

The default method to calibrate the accuracy.

=== Classifier model (full training set) ===
ZeroR predicts class value: NA

```
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances         101               80.1587 %
Incorrectly Classified Instances        25               19.8413 %
Kappa statistic                          0
Mean absolute error                      0.0948
Root mean squared error                  0.2084
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              126
=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area PRC Area
Class
               1.000    1.000    0.802      1.000    0.890      0.000    0.446    0.784     NA
               0.000    0.000    0.000      0.000    0.000      0.000    0.394    0.093     6
               0.000    0.000    0.000      0.000    0.000      0.000    0.048    0.008     2
               0.000    0.000    0.000      0.000    0.000      0.000    0.140    0.020     7
               0.000    0.000    0.000      0.000    0.000      0.000    0.146    0.024     5
               0.000    0.000    0.000      0.000    0.000      0.000    0.097    0.016     3
               0.000    0.000    0.000      0.000    0.000      0.000    0.048    0.008     1
               0.000    0.000    0.000      0.000    0.000      0.000    0.048    0.008     4
Weighted Avg.    0.802    0.802    0.643      0.802    0.713      0.000    0.411    0.640
```

NaiveBayesMultinomialText Result



Classifier: NaiveBayesMultinomialText -P 0 -M 3.0 -norm 1.0 -lnorm 2.0 -stopwords-handler weka.core.stopwords.Null -tokenizer "weka.core.tokenizers.WordTokenizer -delimiters \" \\r\\n\\t.,;:\\\'\\\"()?!\"" -stemmer weka.core.stemmers.NullStem

Result list (right-click for options)

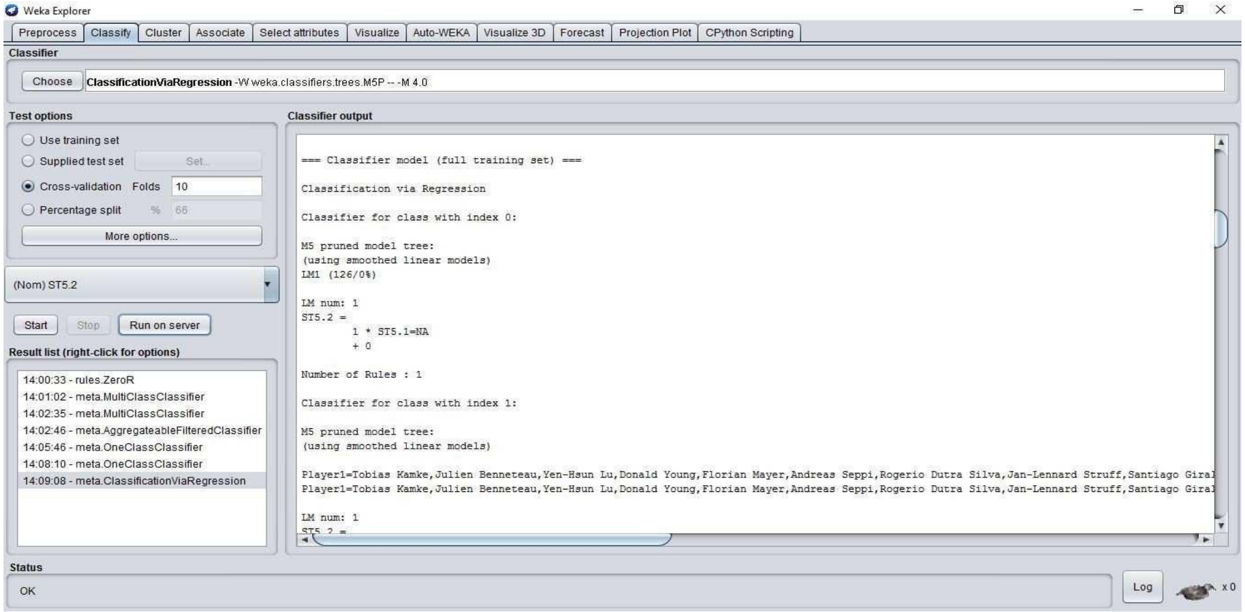
```
10:51:36 - rules.ZeroR
10:54:20 - rules.ZeroR
10:54:33 - rules.ZeroR
10:56:41 - rules.ZeroR
11:02:07 - bayes.NaiveBayesMultinomialText
11:02:19 - bayes.NaiveBayesMultinomialText
11:02:31 - bayes.NaiveBayesMultinomialText
11:02:43 - bayes.NaiveBayesMultinomialText
11:02:58 - bayes.NaiveBayesMultinomialText
11:03:06 - bayes.NaiveBayesMultinomialText
```

```
=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances          86               68.254  %
Incorrectly Classified Instances        40               31.746  %
Kappa statistic                         -0.0611
Mean absolute error                      0.1851
Root mean squared error                  0.2927
Relative absolute error                197.2608 %
Root relative squared error            140.87   %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.851    0.960    0.782      0.851    0.815      -0.130   0.498     0.807     NA
                 0.000    0.000    0.000      0.000    0.000      0.000    0.735     0.249     6
                 0.000    0.128    0.000      0.000    0.000      -0.034   0.352     0.011     2
                 0.000    0.000    0.000      0.000    0.000      0.000    0.862     0.083     7
                 0.000    0.000    0.000      0.000    0.000      0.000    0.714     0.056     5
                 0.000    0.000    0.000      0.000    0.000      0.000    0.593     0.024     3
                 0.000    0.000    0.000      0.000    0.000      0.000    0.624     0.019     1
                 0.000    0.000    0.000      0.000    0.000      0.000    0.624     0.019     4
Weighted Avg.    0.683    0.771    0.627      0.683    0.653      -0.104   0.541     0.679
```

```
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances          101               80.1587 %
Incorrectly Classified Instances         25               19.8413 %
Kappa statistic                          0
Mean absolute error                      0.0948
Root mean squared error                  0.2084
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              126

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area PRC Area
Class
               1.000    1.000    0.802      1.000    0.890      0.000    0.446    0.784    NA
               0.000    0.000    0.000      0.000    0.000      0.000    0.394    0.093    6
               0.000    0.000    0.000      0.000    0.000      0.000    0.048    0.008    2
               0.000    0.000    0.000      0.000    0.000      0.000    0.140    0.020    7
               0.000    0.000    0.000      0.000    0.000      0.000    0.146    0.019    5
               0.000    0.000    0.000      0.000    0.000      0.000    0.097    0.016    3
               0.000    0.000    0.000      0.000    0.000      0.000    0.048    0.008    1
```

0.000	0.000	0.000	0.000	0.000	0.000	0.048	0.008	4

ClassificationViaRegression



Classifier: ClassificationViaRegression -W weka.classifiers.trees.M5P -- -M 4.0

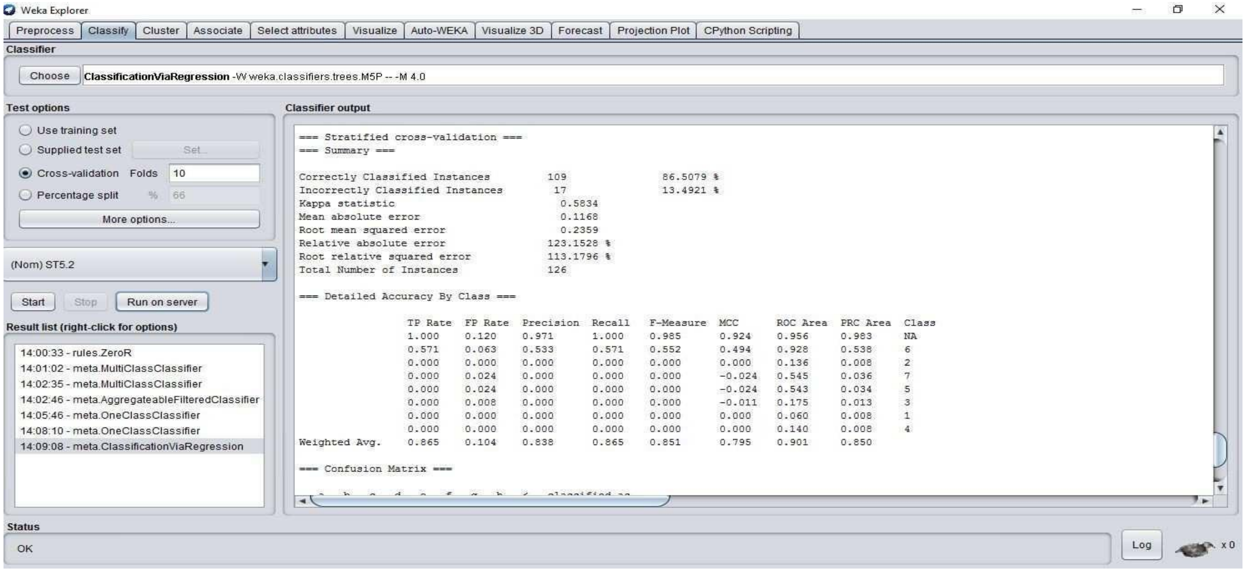
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) ST5.2

```
=== Classifier model (full training set) ===

Classification via Regression

Classifier for class with index 0:

M5 pruned model tree:
(using smoothed linear models)
LM1 (126/0%)

LM num: 1
ST5.2 =
        1 * ST5.1=NA
        + 0

Number of Rules : 1

Classifier for class with index 1:

M5 pruned model tree:
(using smoothed linear models)

Player1=Tobias Kamke,Julien Benneteau,Yen-Hsun Lu,Donald Young,Florian Mayer,Andreas Seppi,Rogerio Dutra Silva,Jan-Lennard Struff,Santiago Giral
Player1=Tobias Kamke,Julien Benneteau,Yen-Hsun Lu,Donald Young,Florian Mayer,Andreas Seppi,Rogerio Dutra Silva,Jan-Lennard Struff,Santiago Giral

LM num: 1
ST5.2 =
```

```
14:00:33 - rules.ZeroR
14:01:02 - meta.MultiClassClassifier
14:02:35 - meta.MultiClassClassifier
14:02:46 - meta.AggregateableFilteredClassifier
14:05:46 - meta.OneClassClassifier
14:08:10 - meta.OneClassClassifier
14:09:08 - meta.ClassificationViaRegression
```

Status: OK



Classifier: ClassificationViaRegression -W weka.classifiers.trees.M5P -- -M 4.0

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) ST5.2

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         109               86.5079 %
Incorrectly Classified Instances        17               13.4921 %
Kappa statistic                          0.5834
Mean absolute error                      0.1168
Root mean squared error                  0.2359
Relative absolute error                123.1528 %
Root relative squared error            113.1796 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.120    0.971      1.000    0.985      0.924    0.956     0.983     NA
                 0.571    0.063    0.533      0.571    0.552      0.494    0.928     0.538     6
                 0.000    0.000    0.000      0.000    0.000      0.000    0.136     0.008     2
                 0.000    0.024    0.000      0.000    0.000      -0.024   0.545     0.036     7
                 0.000    0.024    0.000      0.000    0.000      -0.024   0.543     0.034     5
                 0.000    0.008    0.000      0.000    0.000      -0.011   0.175     0.013     3
                 0.000    0.000    0.000      0.000    0.000      0.000    0.060     0.008     1
                 0.000    0.000    0.000      0.000    0.000      0.000    0.140     0.008     4
Weighted Avg.    0.865    0.104    0.838      0.865    0.851      0.795    0.901     0.850

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
```

```
14:00:33 - rules.ZeroR
14:01:02 - meta.MultiClassClassifier
14:02:35 - meta.MultiClassClassifier
14:02:46 - meta.AggregateableFilteredClassifier
14:05:46 - meta.OneClassClassifier
14:08:10 - meta.OneClassClassifier
14:09:08 - meta.ClassificationViaRegression
```

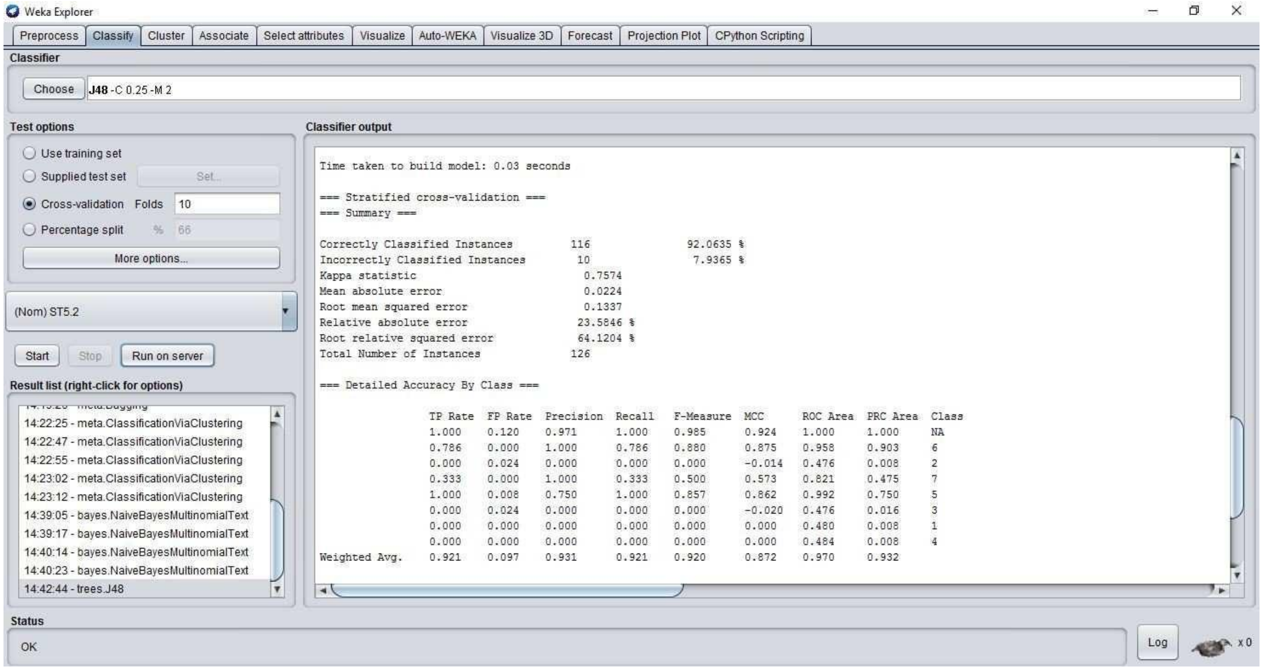
Status: OK

Classificationviaregression cannot handle string attributes of this data set. Hence had to apply remove filter on the dataset during the preprocess stage and then classify using this method. The results are as follows:

Time taken to build model: 2.9 seconds

```
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances         109               86.5079 %
Incorrectly Classified Instances        17               13.4921 %
Kappa statistic                          0.5834
Mean absolute error                      0.1168
Root mean squared error                  0.2359
Relative absolute error                123.1528 %
Root relative squared error            113.1796 %
Total Number of Instances              126
=== Detailed Accuracy By Class ===
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area PRC Area
Class
                 1.000    0.120    0.971      1.000    0.985      0.924    0.956    0.983     NA
                 0.571    0.063    0.533      0.571    0.552      0.494    0.928    0.538     6
                 0.000    0.000    0.000      0.000    0.000      0.000    0.136    0.008     2
                 0.000    0.024    0.000      0.000    0.000      -0.024   0.545    0.036     7
                 0.000    0.024    0.000      0.000    0.000      -0.024   0.543    0.034     5
                 0.000    0.008    0.000      0.000    0.000      -0.011   0.175    0.013     3
                 0.000    0.000    0.000      0.000    0.000      0.000    0.060    0.008     1
                 0.000    0.000    0.000      0.000    0.000      0.000    0.140    0.008     4
Weighted Avg.    0.865    0.104    0.838      0.865    0.851      0.795    0.901    0.850
=== Confusion Matrix ===
 a  b  c  d  e  f  g  h   <-- classified as
 101 0 0 0 0 0 0 0 |  a = NA
   2  8  0  3  1  0  0  0 |   b = 6
   0  0  0  0  1  0  0  0 |   c = 2
   1  2  0  0  0  0  0  0 |   d = 7
   0  2  0  0  0  1  0  0 |   e = 5
   0  1  0  0  1  0  0  0 |   f = 3
   0  1  0  0  0  0  0  0 |   g = 1
   0  1  0  0  0  0  0  0 |   h = 4
```

J48



Classifier: J48 -C 0.25 -M 2

Test options:
- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) ST5.2

Result list (right-click for options)

```
14:15:20 - meta.Bagging
14:22:25 - meta.ClassificationViaClustering
14:22:47 - meta.ClassificationViaClustering
14:22:55 - meta.ClassificationViaClustering
14:23:02 - meta.ClassificationViaClustering
14:23:12 - meta.ClassificationViaClustering
14:39:05 - bayes.NaiveBayesMultinomialText
14:39:17 - bayes.NaiveBayesMultinomialText
14:40:14 - bayes.NaiveBayesMultinomialText
14:40:23 - bayes.NaiveBayesMultinomialText
14:42:44 - trees.J48
```

Classifier output

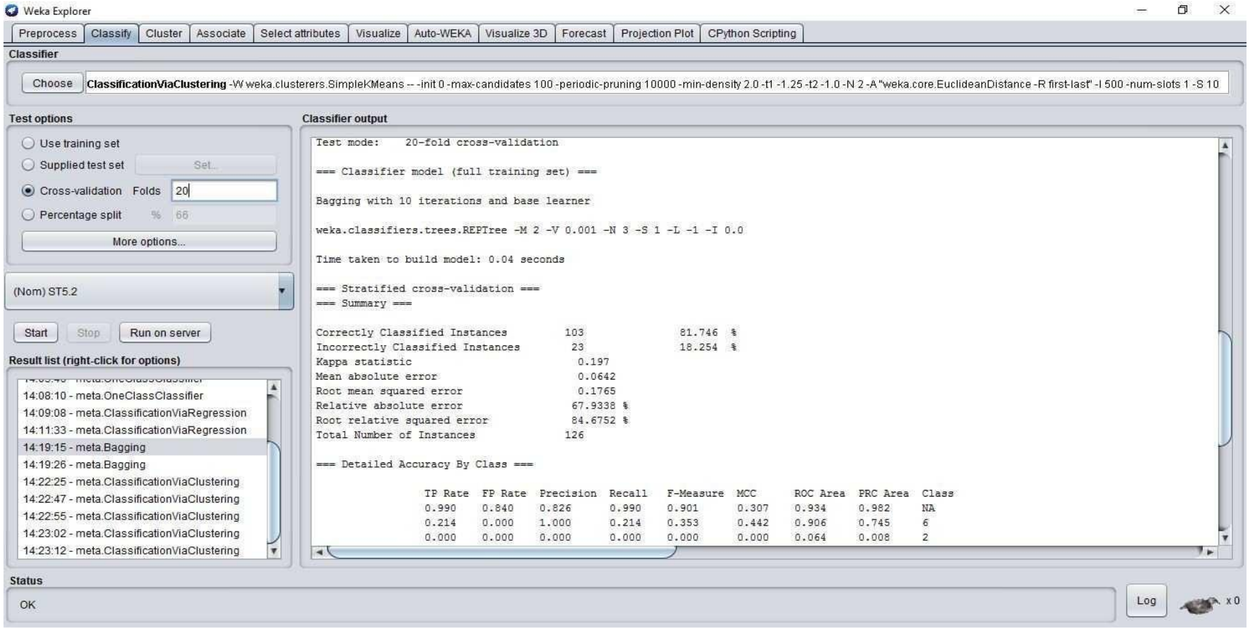
```
Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         116               92.0635 %
Incorrectly Classified Instances        10                7.9365 %
Kappa statistic                          0.7574
Mean absolute error                      0.0224
Root mean squared error                  0.1337
Relative absolute error                 23.5846 %
Root relative squared error             64.1204 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.120    0.971      1.000    0.985      0.924    1.000     1.000     NA
                 0.786    0.000    1.000      0.786    0.880      0.875    0.958     0.903     6
                 0.000    0.024    0.000      0.000    0.000      -0.014   0.476     0.008     2
                 0.333    0.000    1.000      0.333    0.500      0.573    0.821     0.475     7
                 1.000    0.008    0.750      1.000    0.857      0.862    0.992     0.750     5
                 0.000    0.024    0.000      0.000    0.000      -0.020   0.476     0.016     3
                 0.000    0.000    0.000      0.000    0.000      0.000    0.480     0.008     1
                 0.000    0.000    0.000      0.000    0.000      0.000    0.484     0.008     4
Weighted Avg.    0.921    0.097    0.931      0.921    0.920      0.872    0.970     0.932
```

Status
OK

```
Time taken to build model: 0.03 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances         116              92.0635 %
Incorrectly Classified Instances        10               7.9365 %
Kappa statistic                    0.7574
Mean absolute error                    0.0224
Root mean squared error                  0.1337
Relative absolute error               23.5846 %
Root relative squared error              64.1204 %
Total Number of Instances               126
=== Detailed Accuracy By Class ===
             TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area PRC Area
Class
             1.000    0.120    0.971      1.000    0.985      0.924    1.000    1.000    NA
             0.786    0.000    1.000      0.786    0.880      0.875    0.958    0.903    6
             0.000    0.024    0.000      0.000    0.000      -0.014   0.476    0.008    2
             0.333    0.000    1.000      0.333    0.500      0.573    0.821    0.475    7
             1.000    0.008    0.750      1.000    0.857      0.862    0.992    0.750    5
             0.000    0.024    0.000      0.000    0.000      -0.020   0.476    0.016    3
             0.000    0.000    0.000      0.000    0.000      0.000    0.480    0.008    1
```

	0.000	0.000	0.000	0.000	0.000	0.000	0.484	0.008	4
Weighted Avg.	0.921	0.097	0.931	0.921	0.920	0.872	0.970	0.932	

Bagging Result



Classifier: ClassificationViaClustering -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

○ Use training set
○ Supplied test set
● Cross-validation Folds 20
○ Percentage split % 66

(Nom) ST5.2

```
14:08:10 - meta.OneClassClassifier
14:09:08 - meta.ClassificationViaRegression
14:11:33 - meta.ClassificationViaRegression
14:19:15 - meta.Bagging
14:19:26 - meta.Bagging
14:22:25 - meta.ClassificationViaClustering
14:22:47 - meta.ClassificationViaClustering
14:22:55 - meta.ClassificationViaClustering
14:23:02 - meta.ClassificationViaClustering
14:23:12 - meta.ClassificationViaClustering
```

```
Test mode:    20-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 10 iterations and base learner

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         103               81.746  %
Incorrectly Classified Instances        23               18.254  %
Kappa statistic                          0.197
Mean absolute error                      0.0642
Root mean squared error                  0.1765
Relative absolute error                 67.9338 %
Root relative squared error             84.6752 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.990    0.840    0.826      0.990    0.901      0.307    0.934     0.982     NA
                 0.214    0.000    1.000      0.214    0.353      0.442    0.906     0.745     6
                 0.000    0.000    0.000      0.000    0.000      0.000    0.064     0.008     2
```

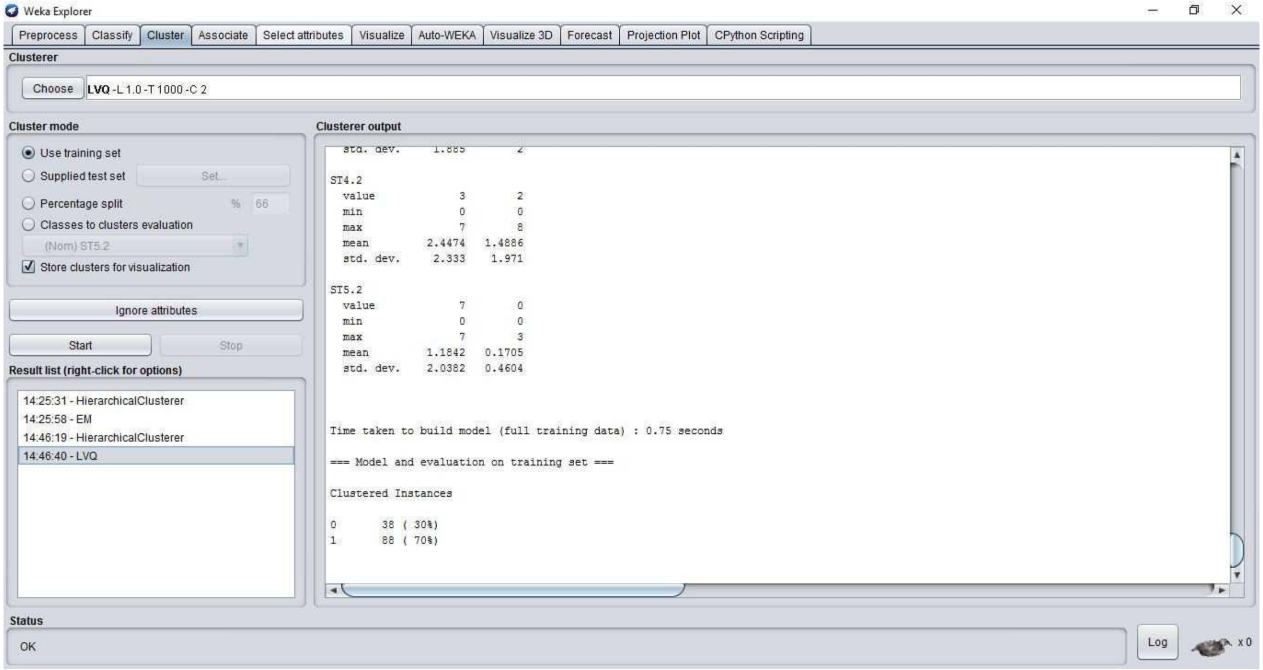
Status: OK

Test mode: 20-fold cross-validation
=== Classifier model (full training set) ===
Bagging with 10 iterations and base learner
weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Time taken to build model: 0.04 seconds
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	103	81.746 %
Incorrectly Classified Instances	23	18.254 %
Kappa statistic	0.197	
Mean absolute error	0.0642	
Root mean squared error	0.1765	
Relative absolute error	67.9338 %	
Root relative squared error	84.6752 %	
Total Number of Instances	126	

CLUSTERING

LVQ Clustering



LVQ -L 1.0 -T 1000 -C 2

● Use training set
○ Supplied test set
○ Percentage split % 66
○ Classes to clusters evaluation
(Nom) ST5.2
☑ Store clusters for visualization

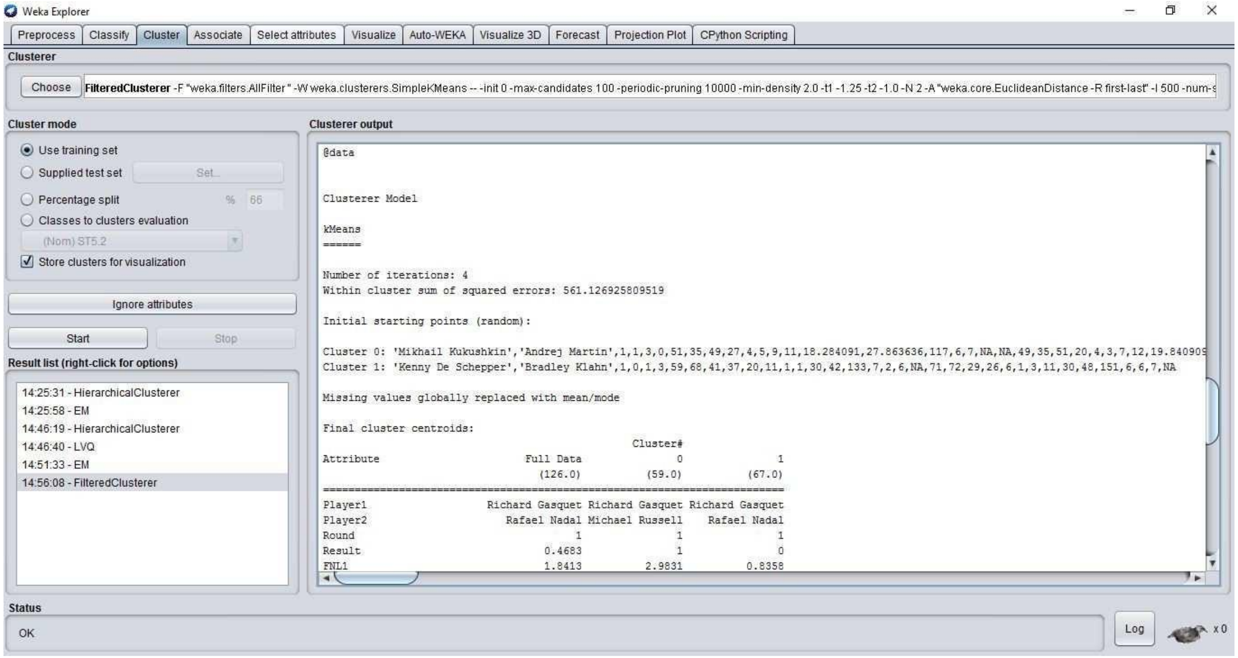
14:25:31 - HierarchicalClusterer
14:25:58 - EM
14:46:19 - HierarchicalClusterer
14:46:40 - LVQ

```
  std. dev.     1.885        2

ST4.2
  value             3        2
  min               0        0
  max               7        8
  mean         2.4474   1.4886
  std. dev.     2.333    1.971

ST5.2
  value             7        0
  min               0        0
  max               7        3
  mean         1.1842   0.1705
  std. dev.    2.0382   0.4604



Time taken to build model (full training data) : 0.75 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       38 ( 30%)
1       88 ( 70%)
```

OK

```
=== Clustering model (full training set) ===
LVQ
==================
Number of clusters: 2
            Cluster
Attribute         0        1
               (38)     (88)

===============================
Player1
  value          19       19
  min             0        0
  max            80       79
  mean      33.7895  38.7159
  std. dev. 26.0967  21.8858
Player2
  value          78       18
  min             0        2
  max            78       78
  mean        36.5526  39.0114
  std. dev.   23.9737  22.3722

FNL1
  value            3        1
  min              2        0
  max              3        3
  mean        2.8947   1.3864
  std. dev.    0.311   1.1884

FNL2
  value            2        3
  min              0        0
  max              3        3
  mean        1.1053   2.2159
  std. dev.   0.9238   1.2635
Time taken to build model (full training data) : 0.75 seconds
=== Model and evaluation on training set ===
Clustered Instances
0       38 ( 30%)
1       88 ( 70%)
```

Filtered Cluster



Clusterer: FilteredClusterer -F "weka.filters.AllFilter " -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-s

● Use training set
○ Supplied test set
○ Percentage split % 66
○ Classes to clusters evaluation
(Nom) ST5.2
☑ Store clusters for visualization

- 14:25:31 - HierarchicalClusterer
- 14:25:58 - EM
- 14:46:19 - HierarchicalClusterer
- 14:46:40 - LVQ
- 14:51:33 - EM
- 14:56:08 - FilteredClusterer

```
@data


Clusterer Model

kMeans
======

Number of iterations: 4
Within cluster sum of squared errors: 561.126925809519

Initial starting points (random):

Cluster 0: 'Mikhail Kukushkin','Andrej Martin',1,1,3,0,51,35,49,27,4,5,9,11,18.284091,27.863636,117,6,7,NA,NA,49,35,51,20,4,3,7,12,19.840909
Cluster 1: 'Kenny De Schepper','Bradley Klahn',1,0,1,3,59,68,41,37,20,11,1,1,30,42,133,7,2,6,NA,71,72,29,26,6,1,3,11,30,48,151,6,6,7,NA

Missing values globally replaced with mean/mode

Final cluster centroids:
                                              Cluster#
Attribute                        Full Data           0               1
                                   (126.0)      (59.0)          (67.0)
==========================================================================
Player1                    Richard Gasquet Richard Gasquet Richard Gasquet
Player2                       Rafael Nadal Michael Russell    Rafael Nadal
Round                                    1               1               1
Result                              0.4683               1               0
FNL1                                1.8413          2.9831          0.8358
```

OK

```
kMeans
======
Number of iterations: 4
Within cluster sum of squared errors: 561.126925809519
Initial starting points (random):
Cluster 0: 'Mikhail Kukushkin','Andrej
Martin',1,1,3,0,51,35,49,27,4,5,9,11,18.284091,27.863636,117,6,7,NA,NA,49,35,51,20,4,3,7,12,
19.840909,31.170455,103,4,6,NA,NA
Cluster 1: 'Kenny De Schepper','Bradley
Klahn',1,0,1,3,59,68,41,37,20,11,1,1,30,42,133,7,2,6,NA,71,72,29,26,6,1,3,11,30,48,151,6,6,7,N
A
Missing values globally replaced with mean/mode
Final cluster centroids:
                                     Cluster#
Attribute                   Full Data             0              1
                          (126.0)          (59.0)          (67.0)
=================================================================
====
Player1               Richard Gasquet Richard Gasquet Richard Gasquet
Player2                 Rafael Nadal Michael Russell    Rafael Nadal
Round                          1              1              1
Result                    0.4683            1              0
FNL1                       1.8413         2.9831         0.8358
FNL2                        1.881         0.6441         2.9701
FSP.1                     58.6508        59.7966        57.6418
FSW.1                      47.4365        47.3051        47.5522
SSP.1                     41.3492        40.2034        42.3582
SSW.1                       23.381        23.7627        23.0448
ACE.1                       8.5079         9.1017         7.9851
DBF.1                      4.9524          3.9661         5.8209
BPC.1                      4.1984          5.8814         2.7164
BPW.1                     10.2619          12.339          8.4328
NPA.1                     18.2841         18.6654         17.9483
NPW.1                      27.8636         27.7958         27.9233
TPW.1                    112.9365       118.8305       107.7463
ST1.1                     4.9683          5.8644         4.1791
ST2.1                     4.8889          5.7119         4.1642
ST4.1                          NA             NA             NA
ST5.1                          NA             NA             NA
```

```
FSP.2                 58.9206     57.7627     59.9403
FSW.2                  46.9365     42.4068     50.9254
SSP.2                 41.0794     42.2373     40.0597
SSW.2                   23.127     21.5085     24.5522
ACE.2                   9.2619      8.6949      9.7612
DBF.2                   4.5952      5.4068      3.8806
BPC.2                   4.0873      2.5254      5.4627
BPW.2                    10.246      7.4068     12.7463
NPA.2                  19.8409     19.7562     19.9155
NPW.2                   31.1705     33.0603     29.5063
TPW.2                  113.1825     98.8136    125.8358
ST1.2                  5.0159      3.9661      5.9403
ST2.2                  4.5159      3.6102      5.3134
ST4.2                      NA          NA          NA
ST5.2                      NA          NA          NA
Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===
Clustered Instances
0       59 ( 47%)
1       67 ( 53%)
```

CONCLUSION

Based on the analysis of the data set of the US open men tennis tournament 2013 and the application of different ML methods and classification methods on the selected dataset, to use J48 method and then to perform the clustering is the best way to predict the finalists of the tournament. Secondarily, we can use classification via regression. The results are convincing and elaborated as necessary.