AIBL

## R Markdown

How the Machine Learning help doctors as part of Dementia Diagnosis with the highest degree of accuracy and reliability from medical record and neuropsychological assessment of AIBL. What are the important predictors? Which part of the Computational Resources, methods of Data Mining activities are the most important and why. Algorithms, Accuracy, Pre-processing, ethical issue, etc. Hint: AI Ethical issue has a fair complicated cases. Is there any validation model framework available to practically align Data Mining with the needs of all their stakeholders.

x\_testdata <- read.csv("c:/users/lenovo/desktop/test.csv")  
  
x\_traindata <- read.csv("c:/users/lenovo/desktop/train.csv")  
  
  
  
  
# Code for installation of the required Libraries  
  
#install.packages("earth", lib="C:/Users/YourUser/Documents/R/win-library/3.3")  
  
library(mlbench)

## Warning: package 'mlbench' was built under R version 4.1.3

library(party)

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x stringr::boundary() masks strucchange::boundary()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(ranger)  
library(e1071)  
library(ggplot2)  
library(dplyr)  
library(Hmisc)

## Loading required package: survival

##   
## Attaching package: 'survival'

## The following object is masked from 'package:caret':  
##   
## cluster

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following object is masked from 'package:e1071':  
##   
## impute

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

library(readxl)  
library(plotrix)  
library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 4.1.3

library(dplyr)  
library(GGally)

## Warning: package 'GGally' was built under R version 4.1.3

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(PerformanceAnalytics)

## Warning: package 'PerformanceAnalytics' was built under R version 4.1.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 4.1.3

##   
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

##   
## Attaching package: 'PerformanceAnalytics'

## The following objects are masked from 'package:e1071':  
##   
## kurtosis, skewness

## The following object is masked from 'package:graphics':  
##   
## legend

library(cowplot)  
library(caret)  
library(rpart)  
library(rpart.plot)  
library(e1071)  
library(randomForest)

## Warning: package 'randomForest' was built under R version 4.1.3

## randomForest 4.7-1

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:ranger':  
##   
## importance

## The following object is masked from 'package:dplyr':  
##   
## combine

## The following object is masked from 'package:ggplot2':  
##   
## margin

library(gbm)

## Warning: package 'gbm' was built under R version 4.1.3

## Loaded gbm 2.1.8

library(Metrics)

## Warning: package 'Metrics' was built under R version 4.1.3

##   
## Attaching package: 'Metrics'

## The following objects are masked from 'package:caret':  
##   
## precision, recall

library(varImp)

## Warning: package 'varImp' was built under R version 4.1.3

## Loading required package: measures

## Warning: package 'measures' was built under R version 4.1.3

##   
## Attaching package: 'measures'

## The following objects are masked from 'package:caret':  
##   
## MAE, RMSE

##   
## Attaching package: 'varImp'

## The following object is masked from 'package:caret':  
##   
## varImp

library(vtreat)

## Warning: package 'vtreat' was built under R version 4.1.3

## Loading required package: wrapr

## Warning: package 'wrapr' was built under R version 4.1.3

##   
## Attaching package: 'wrapr'

## The following object is masked from 'package:dplyr':  
##   
## coalesce

## The following objects are masked from 'package:tidyr':  
##   
## pack, unpack

## The following object is masked from 'package:tibble':  
##   
## view

##   
## Attaching package: 'vtreat'

## The following object is masked from 'package:party':  
##   
## fit

## The following object is masked from 'package:modeltools':  
##   
## fit

library(AUC)

## Warning: package 'AUC' was built under R version 4.1.3

## AUC 0.3.2

## Type AUCNews() to see the change log and ?AUC to get an overview.

##   
## Attaching package: 'AUC'

## The following objects are masked from 'package:Metrics':  
##   
## accuracy, auc

## The following objects are masked from 'package:caret':  
##   
## sensitivity, specificity

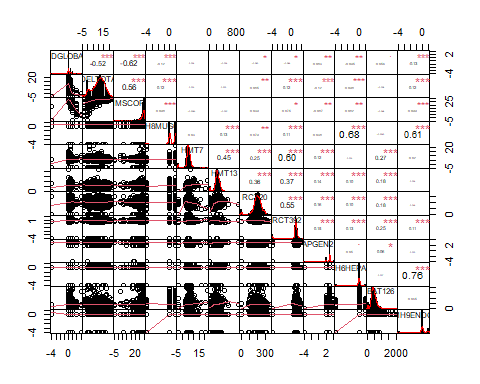
library(Boruta)

## Warning: package 'Boruta' was built under R version 4.1.3

#code for Pre-processing, feature engineering and variable evaluation  
  
set.seed(123)  
describe(x\_traindata)

## x\_traindata   
##   
## 34 Variables 1179 Observations  
## --------------------------------------------------------------------------------  
## RID   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 485 1 384.3 308.1 22.0 45.0   
## .25 .50 .75 .90 .95   
## 119.0 365.0 620.0 771.2 819.8   
##   
## lowest : 2 3 4 5 10, highest: 890 891 893 895 897  
## --------------------------------------------------------------------------------  
## SITEID   
## n missing distinct Info Mean Gmd   
## 1179 0 2 0.677 1.344 0.4514   
##   
## Value 1 2  
## Frequency 774 405  
## Proportion 0.656 0.344  
## --------------------------------------------------------------------------------  
## VISCODE   
## n missing distinct   
## 1179 0 5   
##   
## lowest : bl m18 m36 m54 m72, highest: bl m18 m36 m54 m72  
##   
## Value bl m18 m36 m54 m72  
## Frequency 485 314 163 127 90  
## Proportion 0.411 0.266 0.138 0.108 0.076  
## --------------------------------------------------------------------------------  
## PTGENDER   
## n missing distinct Info Mean Gmd   
## 1179 0 2 0.749 1.517 0.4998   
##   
## Value 1 2  
## Frequency 569 610  
## Proportion 0.483 0.517  
## --------------------------------------------------------------------------------  
## PTDOB   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 33 0.998 1936 7.967 1924 1926   
## .25 .50 .75 .90 .95   
## 1931 1937 1942 1945 1946   
##   
## lowest : 1915 1916 1917 1918 1919, highest: 1943 1944 1945 1946 1947  
## --------------------------------------------------------------------------------  
## MHPSYCH   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.547 0.09754 0.5493   
##   
## Value -4 0 1  
## Frequency 33 899 247  
## Proportion 0.028 0.763 0.209  
## --------------------------------------------------------------------------------  
## MH2NEURL   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.349 0.04919 0.3404   
##   
## Value -4 0 1  
## Frequency 20 1021 138  
## Proportion 0.017 0.866 0.117  
## --------------------------------------------------------------------------------  
## MH4CARD   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.751 0.3384 0.6472   
##   
## Value -4 0 1  
## Frequency 24 660 495  
## Proportion 0.02 0.56 0.42  
## --------------------------------------------------------------------------------  
## MH6HEPAT   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.177 -0.02629 0.226   
##   
## Value -4 0 1  
## Frequency 21 1105 53  
## Proportion 0.018 0.937 0.045  
## --------------------------------------------------------------------------------  
## MH8MUSCL   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.765 0.4122 0.6666   
##   
## Value -4 0 1  
## Frequency 25 568 586  
## Proportion 0.021 0.482 0.497  
## --------------------------------------------------------------------------------  
## MH9ENDO   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.424 0.06785 0.4139   
##   
## Value -4 0 1  
## Frequency 24 979 176  
## Proportion 0.020 0.830 0.149  
## --------------------------------------------------------------------------------  
## MH10GAST   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.647 0.2129 0.5661   
##   
## Value -4 0 1  
## Frequency 23 813 343  
## Proportion 0.020 0.690 0.291  
## --------------------------------------------------------------------------------  
## MH12RENA   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.218 -0.0229 0.27   
##   
## Value -4 0 1  
## Frequency 24 1086 69  
## Proportion 0.020 0.921 0.059  
## --------------------------------------------------------------------------------  
## MH16SMOK   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.887 -1.171 2.301   
##   
## Value -4 0 1  
## Frequency 438 370 371  
## Proportion 0.372 0.314 0.315  
## --------------------------------------------------------------------------------  
## MH17MALI   
## n missing distinct Info Mean Gmd   
## 1179 0 3 0.467 0.1069 0.4222   
##   
## Value -4 0 1  
## Frequency 20 953 206  
## Proportion 0.017 0.808 0.175  
## --------------------------------------------------------------------------------  
## APGEN1   
## n missing distinct Info Mean Gmd   
## 1179 0 4 0.691 3.338 0.4876   
##   
## Value -4 2 3 4  
## Frequency 2 4 757 416  
## Proportion 0.002 0.003 0.642 0.353  
## --------------------------------------------------------------------------------  
## APGEN2   
## n missing distinct Info Mean Gmd   
## 1179 0 4 0.512 2.869 0.3906   
##   
## Value -4 2 3 4  
## Frequency 2 196 925 56  
## Proportion 0.002 0.166 0.785 0.047  
## --------------------------------------------------------------------------------  
## MMSCORE   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 24 0.945 27.75 2.949 21 25   
## .25 .50 .75 .90 .95   
## 27 29 30 30 30   
##   
## lowest : -4 2 3 6 7, highest: 26 27 28 29 30  
## --------------------------------------------------------------------------------  
## CDGLOBAL   
## n missing distinct Info Mean Gmd   
## 1179 0 6 0.548 0.1633 0.2944   
##   
## lowest : -4.0 0.0 0.5 1.0 2.0, highest: 0.0 0.5 1.0 2.0 3.0  
##   
## Value -4.0 0.0 0.5 1.0 2.0 3.0  
## Frequency 2 902 195 61 15 4  
## Proportion 0.002 0.765 0.165 0.052 0.013 0.003  
## --------------------------------------------------------------------------------  
## LIMMTOTAL   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 25 0.996 11.49 5.439 2 5   
## .25 .50 .75 .90 .95   
## 8 12 15 17 19   
##   
## lowest : -4 0 1 2 3, highest: 19 20 21 22 23  
## --------------------------------------------------------------------------------  
## LDELTOTAL   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 25 0.996 10.05 6.069 0 2   
## .25 .50 .75 .90 .95   
## 7 11 14 16 18   
##   
## lowest : -4 0 1 2 3, highest: 19 20 21 22 23  
## --------------------------------------------------------------------------------  
## AXT117   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 347 1 1.684 1.481 0.258 0.600   
## .25 .50 .75 .90 .95   
## 1.075 1.650 2.350 3.200 3.901   
##   
## lowest : -4.000 0.000 0.009 0.020 0.030, highest: 6.760 6.860 6.990 8.490 12.660  
## --------------------------------------------------------------------------------  
## BAT126   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 394 1 442.2 240.3 176.2 230.4   
## .25 .50 .75 .90 .95   
## 305.6 402.5 524.5 677.7 831.4   
##   
## lowest : -4.000 116.562 120.628 126.049 135.537  
## highest: 1863.634 1897.518 1897.519 1931.402 2033.056  
## --------------------------------------------------------------------------------  
## HMT3   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 210 1 4.279 0.8887 3.659 3.940   
## .25 .50 .75 .90 .95   
## 4.200 4.480 4.760 5.030 5.180   
##   
## lowest : -4.00 3.15 3.24 3.27 3.32, highest: 5.84 5.89 5.99 6.05 6.06  
## --------------------------------------------------------------------------------  
## HMT7   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 91 0.999 5.617 2.065 3.6 4.0   
## .25 .50 .75 .90 .95   
## 4.8 5.6 6.6 7.6 8.3   
##   
## lowest : -4.0 2.2 2.6 2.8 2.9, highest: 14.4 15.7 17.2 20.8 22.1  
## --------------------------------------------------------------------------------  
## HMT13   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 253 1 222.2 76.68 117.7 154.0   
## .25 .50 .75 .90 .95   
## 187.0 221.0 264.0 303.0 327.0   
##   
## lowest : -4 16 51 67 74, highest: 469 487 520 556 875  
## --------------------------------------------------------------------------------  
## HMT40   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 71 0.999 13.59 2.206 11.70 12.40   
## .25 .50 .75 .90 .95   
## 13.20 13.90 14.80 15.60 16.11   
##   
## lowest : -4.0 10.1 10.4 10.6 10.9, highest: 17.3 17.5 17.9 18.0 18.1  
## --------------------------------------------------------------------------------  
## HMT100   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 95 0.999 30.37 3.425 27.8 29.0   
## .25 .50 .75 .90 .95   
## 30.4 31.3 32.2 33.2 33.7   
##   
## lowest : -4.0 21.7 23.0 24.9 25.9, highest: 35.3 35.6 36.0 36.4 36.6  
## --------------------------------------------------------------------------------  
## HMT102   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 41 0.998 33.04 2.637 32.7 33.1   
## .25 .50 .75 .90 .95   
## 33.6 34.0 34.5 34.8 35.1   
##   
## lowest : -4.0 31.6 31.9 32.2 32.3, highest: 35.5 35.7 35.8 35.9 36.0  
## --------------------------------------------------------------------------------  
## RCT6   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 94 1 37.09 12.39 22.83 26.43   
## .25 .50 .75 .90 .95   
## 30.64 36.04 42.65 51.06 55.93   
##   
## lowest : -4.000 16.820 18.021 19.222 19.823  
## highest: 82.296 83.497 90.706 109.928 115.334  
## --------------------------------------------------------------------------------  
## RCT11   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 55 0.997 90.03 18.03 73.87 79.27   
## .25 .50 .75 .90 .95   
## 82.87 90.08 97.29 106.29 117.10   
##   
## lowest : -4.000 63.056 66.659 68.461 70.262  
## highest: 163.946 196.374 210.787 212.589 234.208  
## --------------------------------------------------------------------------------  
## RCT20   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 67 0.999 202.2 55.94 119.9 150.8   
## .25 .50 .75 .90 .95   
## 177.9 204.9 232.0 262.9 286.1   
##   
## lowest : -4.000 88.929 92.796 96.662 100.529  
## highest: 328.652 332.519 351.851 359.584 367.317  
## --------------------------------------------------------------------------------  
## RCT392   
## n missing distinct Info Mean Gmd .05 .10   
## 1179 0 97 0.999 0.7612 0.4363 0.577 0.633   
## .25 .50 .75 .90 .95   
## 0.724 0.837 0.995 1.131 1.244   
##   
## lowest : -4.000 0.351 0.419 0.441 0.452, highest: 1.810 1.866 1.923 2.002 2.093  
## --------------------------------------------------------------------------------  
## class   
## n missing distinct   
## 1179 0 2   
##   
## Value HEALTHY CONTROL MCI or AD  
## Frequency 932 247  
## Proportion 0.791 0.209  
## --------------------------------------------------------------------------------

chart.Correlation(select(x\_traindata, CDGLOBAL,LDELTOTAL, MMSCORE, MH8MUSCL, HMT7,HMT13,RCT20, RCT392, APGEN2, MH6HEPAT, BAT126, MH9ENDO), histogram = TRUE, main = "Correlation between Variables")



str(x\_traindata)

## 'data.frame': 1179 obs. of 34 variables:  
## $ RID : int 2 2 3 3 3 3 3 4 4 4 ...  
## $ SITEID : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ VISCODE : chr "bl" "m18" "bl" "m18" ...  
## $ PTGENDER : int 1 1 2 2 2 2 2 2 2 2 ...  
## $ PTDOB : int 1941 1941 1939 1939 1939 1939 1939 1922 1922 1922 ...  
## $ MHPSYCH : int 0 0 1 1 1 1 1 0 0 0 ...  
## $ MH2NEURL : int 1 1 0 0 0 0 0 0 0 0 ...  
## $ MH4CARD : int 0 0 0 0 0 0 0 1 1 1 ...  
## $ MH6HEPAT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH8MUSCL : int 0 0 0 0 0 0 0 1 1 1 ...  
## $ MH9ENDO : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH10GAST : int 0 0 1 1 1 1 1 0 0 0 ...  
## $ MH12RENA : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH16SMOK : int -4 -4 0 0 0 0 0 1 1 1 ...  
## $ MH17MALI : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ APGEN1 : int 3 3 4 4 4 4 4 3 3 3 ...  
## $ APGEN2 : int 3 3 3 3 3 3 3 3 3 3 ...  
## $ MMSCORE : int 30 30 30 28 27 26 27 27 28 27 ...  
## $ CDGLOBAL : num 0 0 0 0 0 0.5 0.5 0 0 0.5 ...  
## $ LIMMTOTAL: int 16 14 9 14 9 8 1 9 7 11 ...  
## $ LDELTOTAL: int 14 15 11 13 11 5 4 2 2 11 ...  
## $ AXT117 : num 1.26 1.01 1.31 2.15 1.94 1.49 1.81 1.33 1.25 2.09 ...  
## $ BAT126 : num 484 397 403 435 350 ...  
## $ HMT3 : num 4.39 4.56 3.87 3.9 3.75 3.72 3.94 4.13 3.67 3.54 ...  
## $ HMT7 : num 5.5 4.6 5.2 5.6 4 4 4.2 6.8 7.1 7.1 ...  
## $ HMT13 : int 220 190 254 311 226 217 202 327 309 383 ...  
## $ HMT40 : num 14.4 14.9 12.6 12.4 12.3 12.4 12.7 13.1 11.6 11.4 ...  
## $ HMT100 : num 32.8 32.7 32.5 31.8 32.7 33.3 32.2 31.7 31.7 32.3 ...  
## $ HMT102 : num 34.3 34.4 34.2 33.8 33.9 34.8 33.5 34.2 34.7 35.9 ...  
## $ RCT6 : num 37.8 51.1 18 23.4 25.8 ...  
## $ RCT11 : num 90.1 88.3 129.7 97.3 90.1 ...  
## $ RCT20 : num 174 182 251 267 302 ...  
## $ RCT392 : num 0.916 0.928 0.792 0.792 0.735 ...  
## $ class : chr "HEALTHY CONTROL" "HEALTHY CONTROL" "HEALTHY CONTROL" "HEALTHY CONTROL" ...

str(x\_testdata)

## 'data.frame': 507 obs. of 33 variables:  
## $ RID : int 899 900 902 904 904 904 904 906 906 906 ...  
## $ SITEID : int 2 2 1 2 2 2 2 2 2 2 ...  
## $ VISCODE : chr "bl" "bl" "bl" "bl" ...  
## $ PTGENDER : int 2 1 1 2 2 2 2 1 1 1 ...  
## $ PTDOB : int 1943 1937 1931 1945 1945 1945 1945 1932 1932 1932 ...  
## $ MHPSYCH : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH2NEURL : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH4CARD : int 0 1 1 1 1 1 1 0 0 0 ...  
## $ MH6HEPAT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH8MUSCL : int 1 0 1 0 0 0 0 0 0 0 ...  
## $ MH9ENDO : int 0 0 1 0 0 0 0 0 0 0 ...  
## $ MH10GAST : int 1 0 1 0 0 0 0 1 1 1 ...  
## $ MH12RENA : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ MH16SMOK : int -4 -4 -4 0 0 0 0 1 1 1 ...  
## $ MH17MALI : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ APGEN1 : int 3 3 3 4 4 4 4 3 3 3 ...  
## $ APGEN2 : int 2 3 3 3 3 3 3 2 2 2 ...  
## $ MMSCORE : int 30 28 29 29 28 30 29 29 29 29 ...  
## $ CDGLOBAL : num 0 0 0.5 0 0 0 0 0 0 0 ...  
## $ LIMMTOTAL: int 16 14 14 15 12 13 16 16 21 17 ...  
## $ LDELTOTAL: int 14 10 11 17 12 16 16 14 19 18 ...  
## $ AXT117 : num 1.3 2 0.85 3 2.5 3.6 3.1 1.6 1.8 1.9 ...  
## $ BAT126 : num 786 420 485 501 366 ...  
## $ HMT3 : num 4.05 4.95 4.13 4.43 4.6 4.34 4.29 5.24 5.35 5.41 ...  
## $ HMT7 : num 7 7.8 5.8 6.9 6.6 5.3 8.7 6.9 6.7 5.8 ...  
## $ HMT13 : int 344 199 252 262 221 222 199 217 195 178 ...  
## $ HMT40 : num 12.6 14.9 11.5 14.5 14.6 14.1 14.1 16.5 16.7 17.2 ...  
## $ HMT100 : num 31.2 30.2 27.8 32.7 31.6 32.5 32.9 31.5 31.2 31.9 ...  
## $ HMT102 : num 33.6 33.4 33.1 34.1 33.3 34.3 34.3 34 34 34.6 ...  
## $ RCT6 : num 33 27 37.8 35.4 34.8 ...  
## $ RCT11 : num 82.9 90.1 120.7 88.3 84.7 ...  
## $ RCT20 : num 220 174 159 228 143 ...  
## $ RCT392 : num 0.622 1.018 0.837 1.041 1.052 ...

x\_traindata$class <-factor(x\_traindata$class)  
  
  
#boruta <- Boruta(x\_traindata$class ~ ., data = x\_traindata, doTrace = 2, maxRuns = 100)  
#boruta\_signif <- names(boruta$finalDecision[boruta$finalDecision %in% c("Confirmed", "Tentative")])  
#print(boruta\_signif) # significant variables  
#plot(boruta, cex.axis=1.1, las=2, xlab="", main="Variable Importance") # plot variable importance  
  
library(earth)

## Warning: package 'earth' was built under R version 4.1.3

## Loading required package: plotmo

## Warning: package 'plotmo' was built under R version 4.1.3

## Loading required package: TeachingDemos

## Warning: package 'TeachingDemos' was built under R version 4.1.3

##   
## Attaching package: 'TeachingDemos'

## The following objects are masked from 'package:Hmisc':  
##   
## cnvrt.coords, subplot

marsModel <- earth(class ~ ., data=x\_traindata) # build model  
ev <- evimp (marsModel) # estimate variable importance  
print(ev)

## nsubsets gcv rss  
## CDGLOBAL 18 100.0 100.0  
## LDELTOTAL 15 15.3 18.3  
## MMSCORE 15 15.1 18.1  
## MH8MUSCL 13 8.9 13.0  
## HMT7 9 7.4 10.7  
## MH6HEPAT 8 6.6 9.9  
## BAT126 6 4.9 8.0  
## MH9ENDO 4 3.3 6.2  
## RID 1 2.0 3.2

df = subset(x\_traindata, select = c(CDGLOBAL,LDELTOTAL, MMSCORE, MH8MUSCL, HMT7,HMT13,RCT20, RCT392, APGEN2, MH6HEPAT, BAT126, MH9ENDO, class) )

## Including modelling with new variables and testing and training the data

The model are then evaluated by the comparison with the test data, with the model selected, on the Result evaluation with the help of environmental variables in Support Vector Classifier, Naïve Bayes, Random Forest

## + Fold01: mtry=3   
## - Fold01: mtry=3   
## + Fold02: mtry=3   
## - Fold02: mtry=3   
## + Fold03: mtry=3   
## - Fold03: mtry=3   
## + Fold04: mtry=3   
## - Fold04: mtry=3   
## + Fold05: mtry=3   
## - Fold05: mtry=3   
## + Fold06: mtry=3   
## - Fold06: mtry=3   
## + Fold07: mtry=3   
## - Fold07: mtry=3   
## + Fold08: mtry=3   
## - Fold08: mtry=3   
## + Fold09: mtry=3   
## - Fold09: mtry=3   
## + Fold10: mtry=3   
## - Fold10: mtry=3   
## Aggregating results  
## Fitting final model on full training set

nb\_predict

| x |
| --- |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |

## user system elapsed   
## 17.71 8.62 30.35

## user system elapsed   
## 18.17 8.62 30.82

svm\_predict

| x |
| --- |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| MCI or AD |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |
| HEALTHY CONTROL |

## [1] 96.26802

## Random Forest   
##   
## 1179 samples  
## 12 predictor  
## 2 classes: 'HEALTHY CONTROL', 'MCI or AD'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 1060, 1061, 1062, 1062, 1061, 1061, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.969498 0.9099623  
##   
## Tuning parameter 'mtry' was held constant at a value of 3

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = df, y = df$class)  
##   
## A-priori probabilities:  
## df$class  
## HEALTHY CONTROL MCI or AD   
## 0.7905004 0.2094996   
##   
## Conditional probabilities:  
## CDGLOBAL  
## df$class [,1] [,2]  
## HEALTHY CONTROL 0.01555794 0.1639484  
## MCI or AD 0.72064777 0.5851297  
##   
## LDELTOTAL  
## df$class [,1] [,2]  
## HEALTHY CONTROL 11.850858 4.128738  
## MCI or AD 3.275304 3.950548  
##   
## MMSCORE  
## df$class [,1] [,2]  
## HEALTHY CONTROL 28.85837 1.608685  
## MCI or AD 23.55061 5.857771  
##   
## MH8MUSCL  
## df$class [,1] [,2]  
## HEALTHY CONTROL 0.4817597 0.6657897  
## MCI or AD 0.1497976 1.1952778  
##   
## HMT7  
## df$class [,1] [,2]  
## HEALTHY CONTROL 5.644099 1.966895  
## MCI or AD 5.517004 3.156931  
##   
## HMT13  
## df$class [,1] [,2]  
## HEALTHY CONTROL 223.8552 69.34482  
## MCI or AD 215.7895 90.56846  
##   
## RCT20  
## df$class [,1] [,2]  
## HEALTHY CONTROL 205.1053 48.03329  
## MCI or AD 191.0298 70.34670  
##   
## RCT392  
## df$class [,1] [,2]  
## HEALTHY CONTROL 0.8133315 0.5662246  
## MCI or AD 0.5643117 1.2552843  
##   
## APGEN2  
## df$class [,1] [,2]  
## HEALTHY CONTROL 2.836910 0.4237958  
## MCI or AD 2.991903 0.8013804  
##   
## MH6HEPAT  
## df$class [,1] [,2]  
## HEALTHY CONTROL 0.0000000 0.4397050  
## MCI or AD -0.1255061 0.9131143  
##   
## BAT126  
## df$class [,1] [,2]  
## HEALTHY CONTROL 441.1252 236.1028  
## MCI or AD 446.4596 302.8418  
##   
## MH9ENDO  
## df$class [,1] [,2]  
## HEALTHY CONTROL 0.1201717 0.5849351  
## MCI or AD -0.1295547 0.9539270  
##   
## class  
## df$class HEALTHY CONTROL MCI or AD  
## HEALTHY CONTROL 1 0  
## MCI or AD 0 1

## AIBL.nb\_model.prediction  
## HEALTHY CONTROL MCI or AD  
## HEALTHY CONTROL 907 25  
## MCI or AD 7 240

##   
## Call:  
## svm(formula = class ~ ., data = df, cost = 10, cross = 10, type = "C-classification",   
## kernel = "radial", na.action = na.omit)  
##   
##   
## Parameters:  
## SVM-Type: C-classification   
## SVM-Kernel: radial   
## cost: 10   
##   
## Number of Support Vectors: 201

##   
## Call:  
## svm(formula = class ~ ., data = df, gamma = 0.1)  
##   
##   
## Parameters:  
## SVM-Type: C-classification   
## SVM-Kernel: radial   
## cost: 1   
##   
## Number of Support Vectors: 247