

GA Classification

Marwa ZAHOUANI

12/10/2021

```
# Install packages if missing
list.of.packages <- c("parallel", "doParallel", "caret", "randomForest", "funModeling", "tidyverse", "G
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)

# Load libraries
library(caret)

## Le chargement a nécessité le package : lattice
## Le chargement a nécessité le package : ggplot2
library(randomForest)

## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.
##
## Attachement du package : 'randomForest'
## L'objet suivant est masqué depuis 'package:ggplot2':
##
##      margin
library(funModeling)

## Le chargement a nécessité le package : Hmisc
## Le chargement a nécessité le package : survival
##
## Attachement du package : 'survival'
## L'objet suivant est masqué depuis 'package:caret':
##
##      cluster
## Le chargement a nécessité le package : Formula
##
## Attachement du package : 'Hmisc'
## Les objets suivants sont masqués depuis 'package:base':
##
##      format.pval, units
## funModeling v.1.9.4 :)
## Examples and tutorials at livebook.datascienceheroes.com
```

```

## / Now in Spanish: librovivodecienciadedatos.ai
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v tibble 3.1.4      v dplyr 1.0.7
## v tidyr 1.1.3      v stringr 1.4.0
## v readr 2.0.1      v forcats 0.5.1
## v purrr 0.3.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::combine()      masks randomForest::combine()
## x dplyr::filter()       masks stats::filter()
## x dplyr::lag()          masks stats::lag()
## x purrr::lift()         masks caret::lift()
## x randomForest::margin() masks ggplot2::margin()
## x dplyr::src()          masks Hmisc::src()
## x dplyr::summarize()    masks Hmisc::summarize()
library(GA)

## Le chargement a nécessité le package : foreach
##
## Attachement du package : 'foreach'
## Les objets suivants sont masqués depuis 'package:purrr':
##
##      accumulate, when
## Le chargement a nécessité le package : iterators
## Package 'GA' version 3.2.1
## Type 'citation("GA")' for citing this R package in publications.
##
## Attachement du package : 'GA'
## L'objet suivant est masqué depuis 'package:utils':
##
##      de
##library(rsample)
library(caTools)
source("lib_ga.R")

data=read_delim("marwa13 (2).csv", delim = ",")

## Rows: 1128 Columns: 25
## -- Column specification -----
## Delimiter: ","
## chr (13): output, Liquid1, Liquid2, Liquid3, Activ6, Activ9, Activ10, debt11...
## dbl (12): year, Liquid4, Liquid5, Activ7, Activ8, debt12, debt13, debt15, de...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```

```

# Data preparation

data2=na.omit(data) # <- use with care...

## Warning: One or more parsing issues, see `problems()` for details
sum(is.na(data2))

## [1] 0

data_y=as.factor(data2$output)
data_x=select(data2, -output, -year)

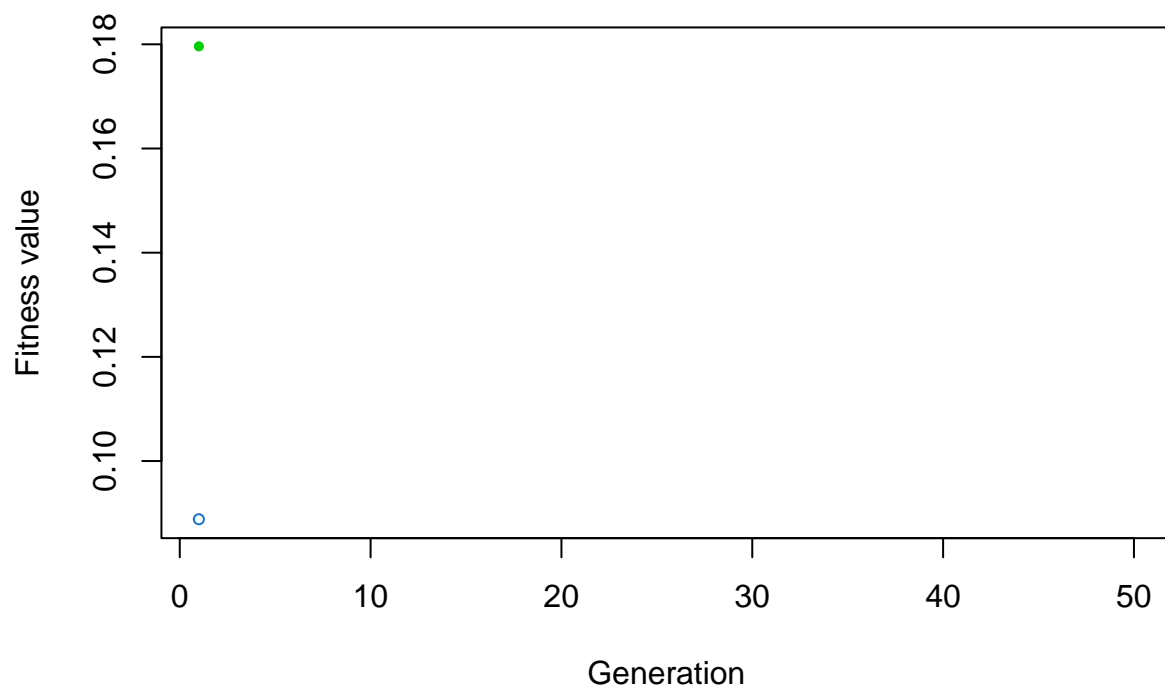
# GA parameters
param_nBits=ncol(data_x)
col_names=colnames(data_x)

# Executing the GA

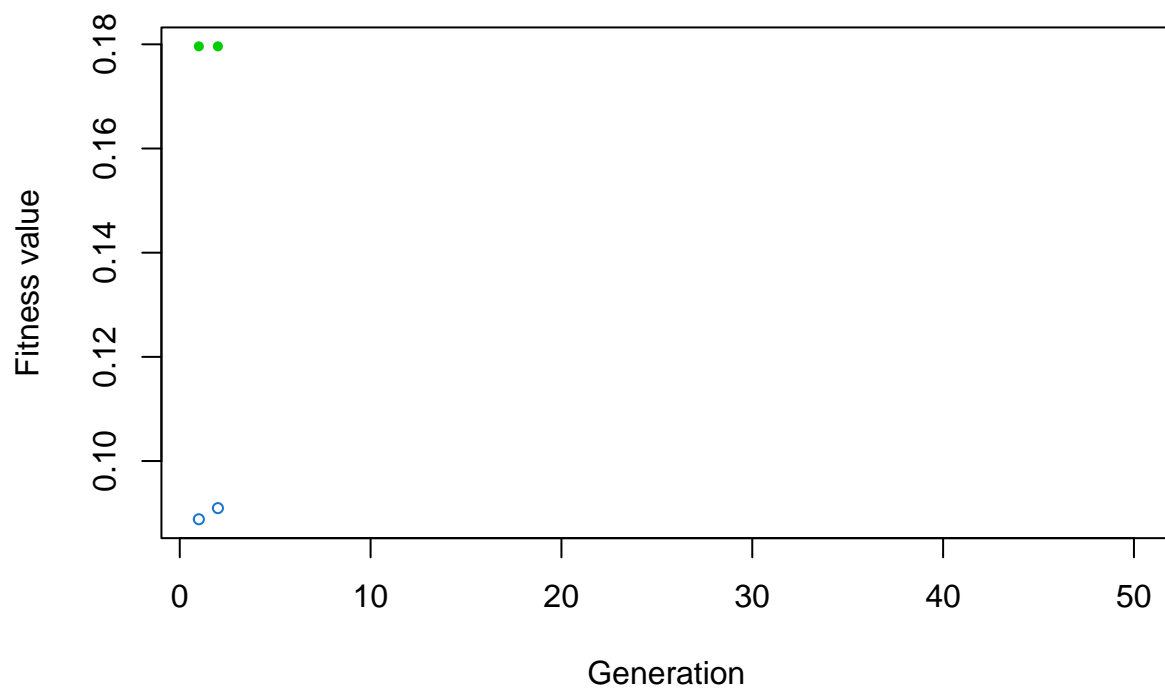
ga_GA_1 = ga(fitness = function(vars) custom_fitness(vars = vars,
                                                    data_x = data_x,
                                                    data_y = data_y,
                                                    p_sampling = 0.7), # custom fitness function
            type = "binary", # optimization data type
            crossover=gabin_uCrossover, # cross-over method
            elitism = 3, # number of best ind. to pass to next iteration
            pmutation = 0.03, # mutation rate prob
            popSize = 100, # the number of individuals/solutions
            nBits = param_nBits, # total number of variables
            names=col_names, # variable name
            run=5, # max iter without improvement (stopping criteria)
            maxiter = 50, # total runs or generations
            monitor=plot, # plot the result at each iteration
            keepBest = TRUE, # keep the best solution at the end
            parallel = T, # allow parallel procesing
            seed=84211 # for reproducibility purposes
)

```

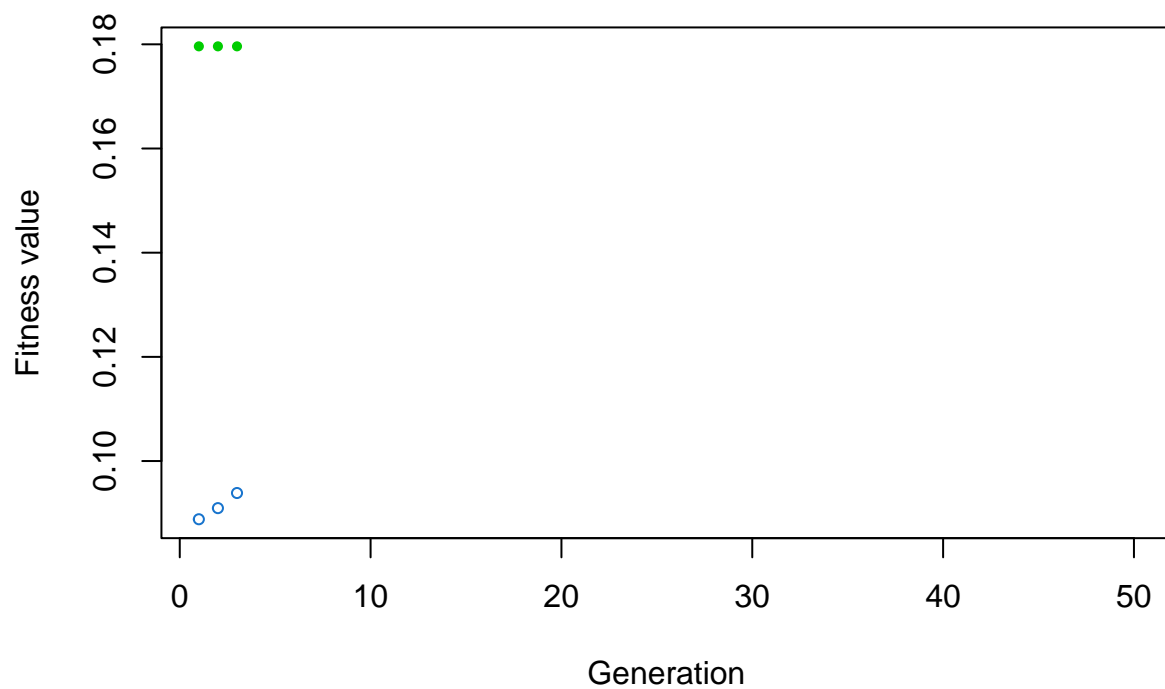
Iteration 1



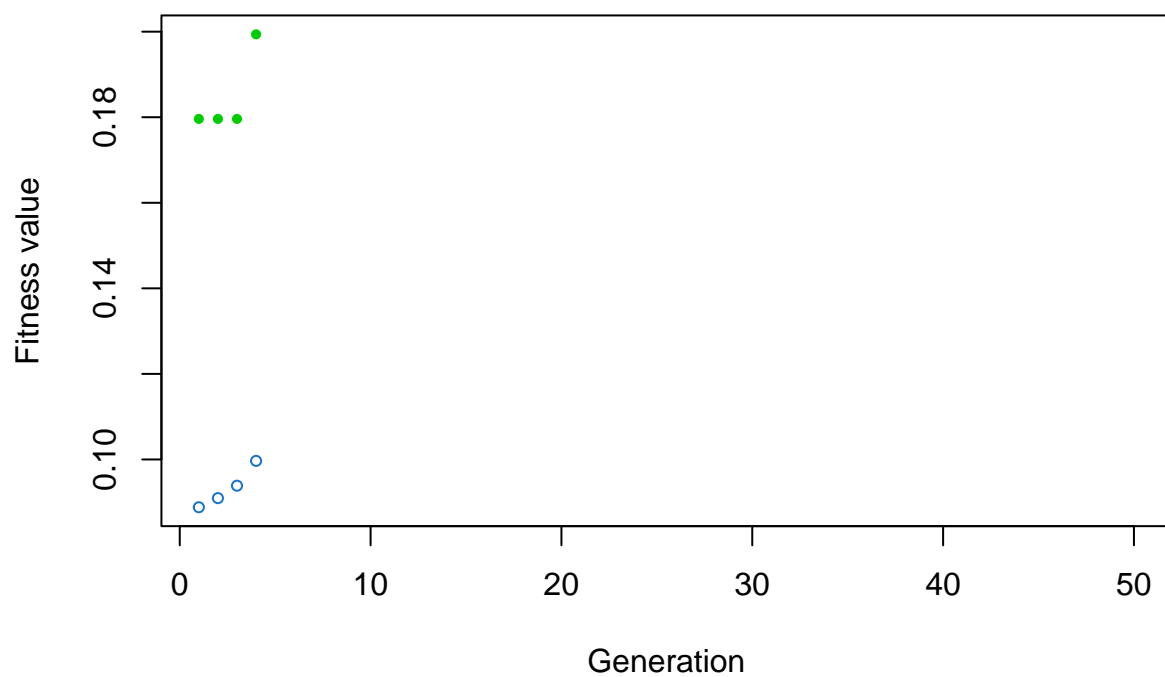
Iteration 2



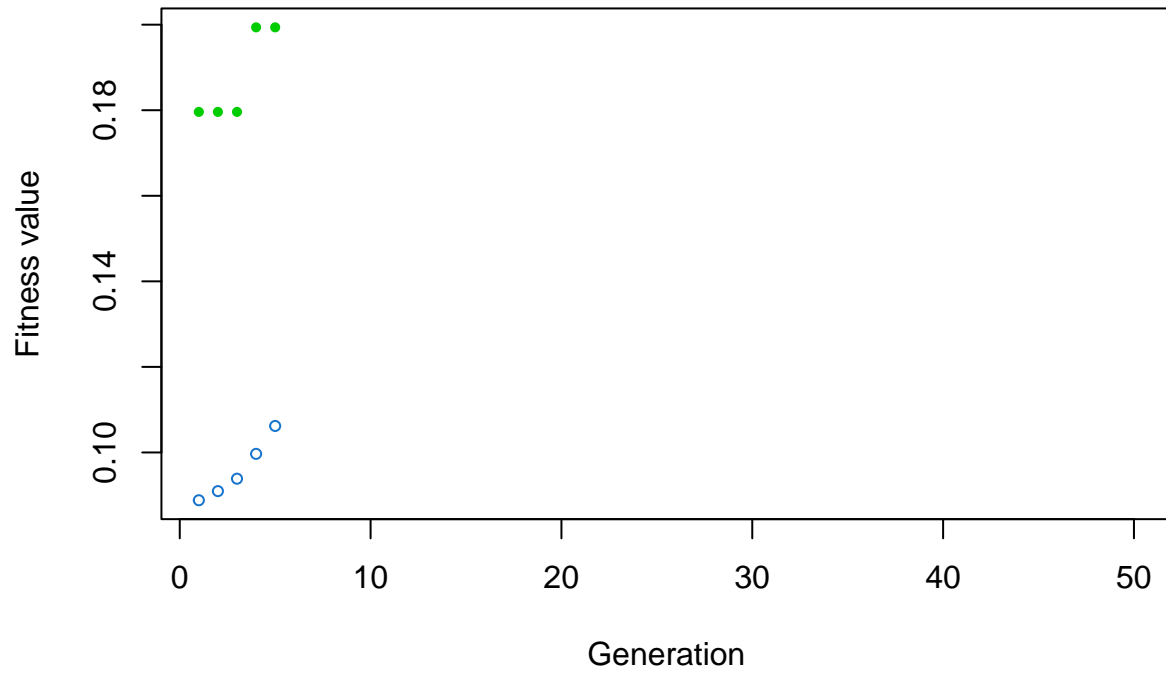
Iteration 3



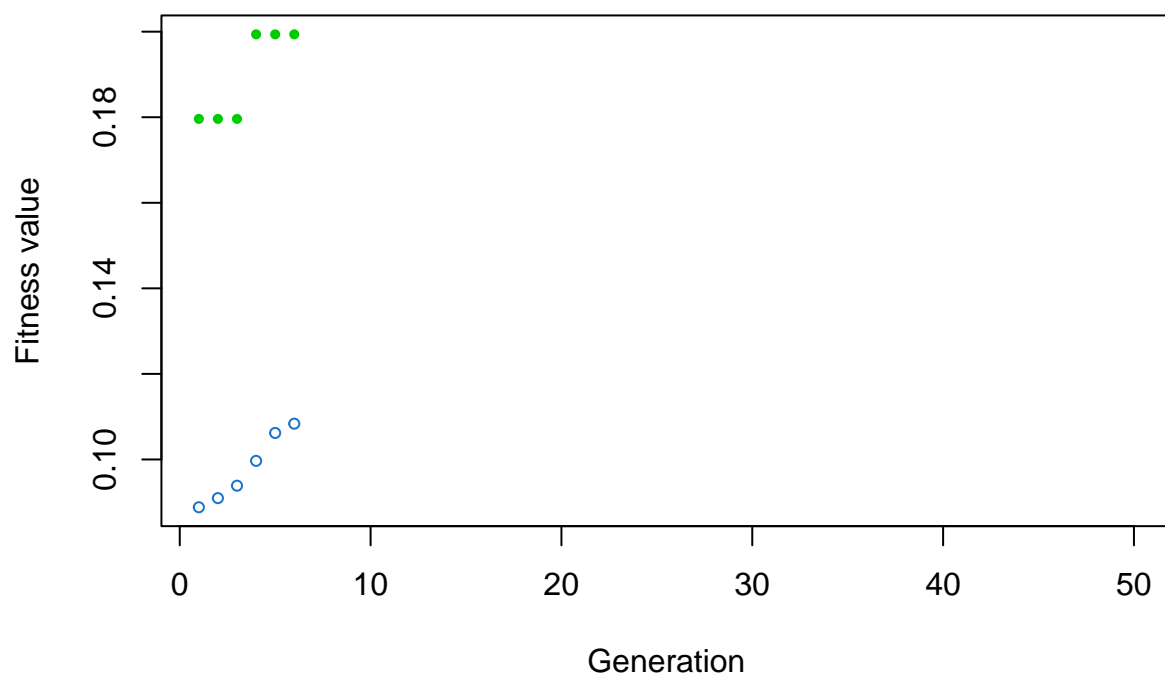
Iteration 4



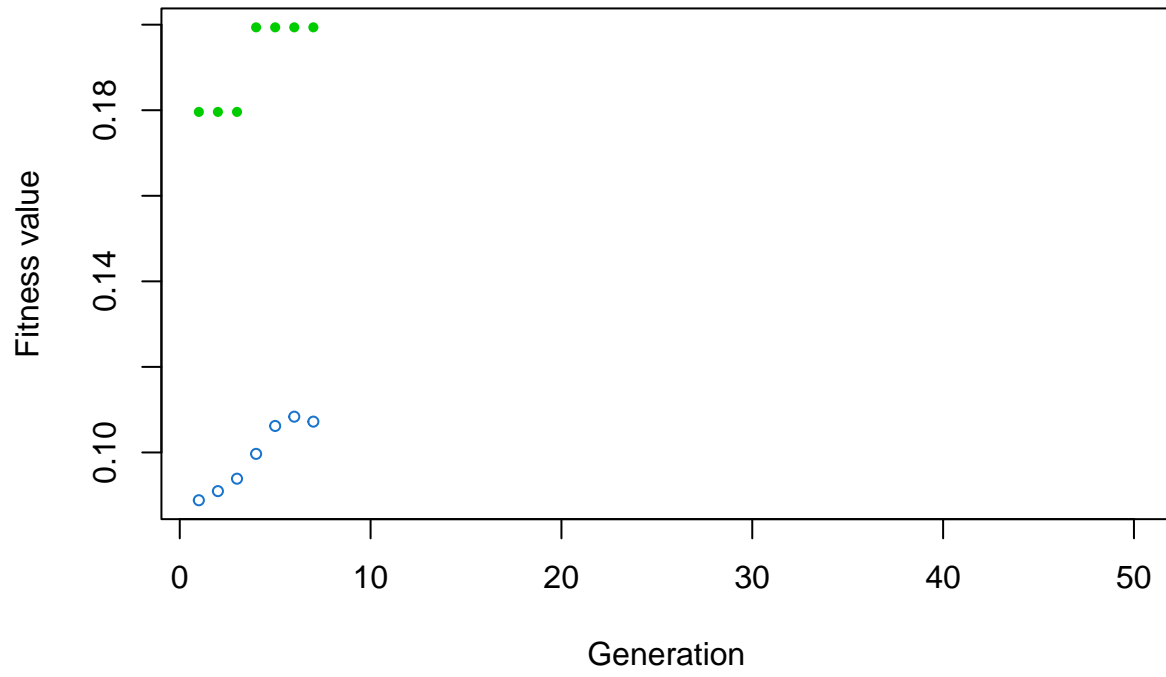
Iteration 5



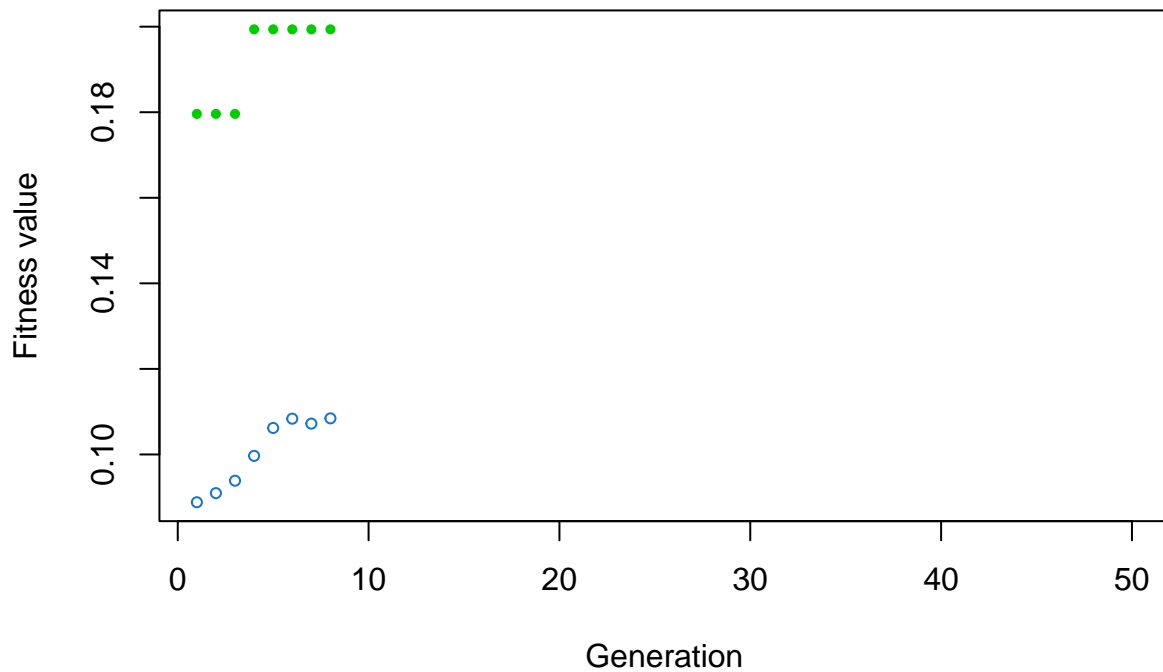
Iteration 6



Iteration 7



Iteration 8



Checking the results

```
summary(ga_GA_1)
```

```
## -- Genetic Algorithm -----
```

```
##
```

```
## GA settings:
```

```
## Type = binary
```

```
## Population size = 100
```

```
## Number of generations = 50
```

```
## Elitism = 3
```

```
## Crossover probability = 0.8
```

```
## Mutation probability = 0.03
```

```
##
```

```
## GA results:
```

```
## Iterations = 8
```

```
## Fitness function value = 0.1993749
```

```
## Solution =
```

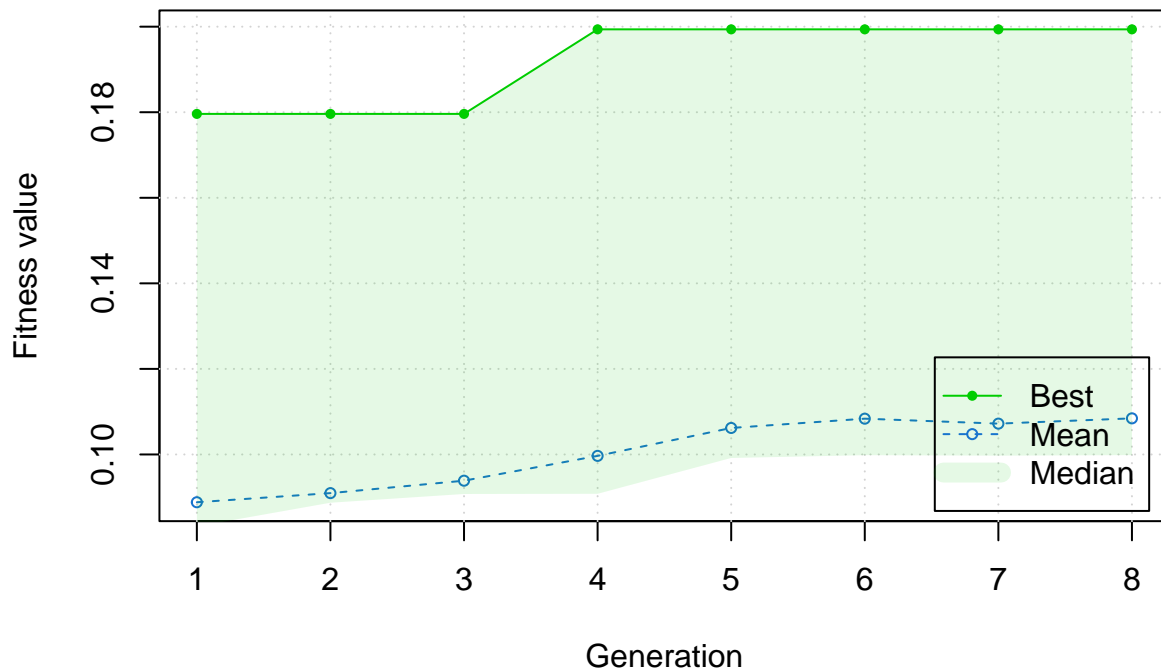
```
## Liquid1 Liquid2 Liquid3 Liquid4 Liquid5 Activ6 Activ7 Activ8 Activ9
```

```
## [1,] 0 0 0 0 0 1 0 0 0
```

```
## Activ10 ... Pofit22 Pofit23
```

```
## [1,] 0 0 0
```

```
plot.ga(ga_GA_1)
```



```
# Following line will return the variable names of the final and best solution
set.seed(1)
best_vars_ga=col_names[ga_GA_1@solution[1,]==1]
```

```
# Checking the variables of the best solution...
set.seed(21)
best_vars_ga
```

```
## [1] "Activ6" "debt11" "debt13" "debt17" "Pofit21"
```

```
# Checking the accuracy
set.seed(123)
get_accuracy_metric(data_tr_sample = data_x, target = data_y, best_vars_ga)
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```

## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## Warning: Setting row names on a tibble is deprecated.
## [1] 0.9773225
###confusionMatrix###
##### end of GA #####
#####The end of GA

```