Nada

Juan Luis Suárez Díaz

6 de marzo de 2019

```
train <- read.csv("train.csv", na.strings = c("?", "NA", "NR", "na", "NaN", "nan"))
train$C <- as.factor(train$C)
test <- read.csv("test.csv", na.strings = c("?", "NA", "NR", "na", "NaN", "nan"))
sample <- read.csv("sampleSubmission.csv")

## BIBLIOTECAS

library(ggplot2)
library(caret)
```

```
## Loading required package: lattice
```

```
library(RKEEL)
# library(rDML) # Por si acaso
library(kknn)
```

```
##
## Attaching package: 'kknn'
```

```
## The following object is masked from 'package:caret':
##
##     contr.dummy
```

```
library(GGally)
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 objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

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

```
## The following object is masked from 'package:GGally':
##
##     nasa
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
library(tidyr)
library(VIM)
```

```
## Loading required package: colorspace
## Loading required package: grid
## Loading required package: data.table
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
## VIM is ready to use.
##  Since version 4.0.0 the GUI is in its own package VIMGUI.
##
##           Please use the package to use the new (and old) GUI.
## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues
##
## Attaching package: 'VIM'
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
library(mice)
```

```
##
## Attaching package: 'mice'
## The following object is masked from 'package:tidyr':
##
##     complete
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library(bmrm)
library(DMwR)
```

```
##
## Attaching package: 'DMwR'
```

```
## The following object is masked from 'package:VIM':
##
##     kNN
```

```
library(NoiseFiltersR)
library(beeswarm)
library(moments)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(FSelector)
library(Gmedian)

has.na <- function(x) apply(x,1,function(z)any(is.na(z)))
indices.nas.train <- which(has.na(train))
outliers.train.por.la.cara <- which(apply(train[,-ncol(train)], MARGIN=1, function(x) any(!is.na(x) & x
outliers.test.por.la.cara <- which(apply(test, MARGIN=1, function(x) any(!is.na(x) & x < -68000)))
```

```
# Variables más correladas
cor(train[-c(indices.nas.train, outliers.train.por.la.cara),-ncol(train)]) %>%
  as.data.frame() %>%
  mutate(var1 = rownames(.)) %>%
  gather(var2, value, -var1) %>%
  arrange(desc(abs(value))) %>%
  filter(var1 < var2) %>%
  head(n=10)
```

```
##    var1 var2      value
## 1   X16  X17  0.9526716
## 2   X26  X44 -0.8867555
## 3   X33  X38  0.8804476
## 4   X16  X23  0.8704657
## 5   X15  X16  0.8560844
## 6   X17  X23  0.8256626
## 7   X12  X34 -0.8143807
## 8   X15  X17  0.7999134
## 9   X40  X46  0.7946815
## 10  X28  X37 -0.7796407
```