# R Notebook

## Parametros:

**Measure =** G-mean
**Columns =** sampling, weight_space, underbagging, learner
**Performance =** holdout_measure_residual
**Filter keys =** NULL
**Filter values =** NULL

```
library("scmamp")
library(dplyr)
```

## Tratamento dos dados

Carregando data set compilado

```
ds = read.csv("/home/rodrigo/Dropbox/UNICAMP/IC/estudo_cost_learning/SummaryResults/summary_compilation_

ds = filter(ds, learner != "classif.rusboost")
summary(ds)
```

```
##                  learner      weight_space
##  classif.ksvm        :17100   Mode :logical
##  classif.randomForest:17100   FALSE:41040
##  classif.rusboost    :    0   TRUE :10260
##  classif.xgboost     :17100   NA's :0
##
##
##
##                             measure        sampling     underbagging
##  Accuracy                       :10260   ADASYN:10260   Mode :logical
##  Area under the curve           :10260   FALSE :30780   FALSE:41040
##  F1 measure                     :10260   SMOTE :10260   TRUE :10260
##  G-mean                         :10260                  NA's :0
##  Matthews correlation coefficient:10260
##
##
##  tuning_measure    holdout_measure   holdout_measure_residual
##  Min.   :-0.1277   Min.   :-0.2120   Min.   :-0.4658
##  1st Qu.: 0.6911   1st Qu.: 0.4001   1st Qu.: 0.1994
##  Median : 0.9700   Median : 0.8571   Median : 0.5581
##  Mean   : 0.7903   Mean   : 0.6718   Mean   : 0.5298
##  3rd Qu.: 0.9975   3rd Qu.: 0.9900   3rd Qu.: 0.8755
##  Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.0000
##  NA's   :1077      NA's   :1077      NA's   :1077
##  iteration_count                      dataset        imba.rate
##  Min.   :1       abalone                 :  900   Min.   :0.0010
##  1st Qu.:1       adult                   :  900   1st Qu.:0.0100
##  Median :2       bank                    :  900   Median :0.0300
##  Mean   :2       car                     :  900   Mean   :0.0286
```

```
##  3rd Qu.:3       cardiotocography-10clases:  900   3rd Qu.:0.0500
##  Max.   :3       cardiotocography-3clases :  900   Max.   :0.0500
##  NA's   :1077    (Other)                  :45900
```

Filtrando pela metrica

```
ds = filter(ds, measure == params$measure)
```

Filtrando o data set

```
if(params$filter_keys != 'NULL' && !is.null(params$filter_keys)){
  dots = paste0(params$filter_keys," == '",params$filter_values,"'")
  ds = filter_(ds, .dots = dots)
}

summary(ds)
```

```
##                  learner     weight_space
##  classif.ksvm        :3420   Mode :logical
##  classif.randomForest:3420   FALSE:8208
##  classif.rusboost    :   0   TRUE :2052
##  classif.xgboost     :3420   NA's :0
##
##
##
##                             measure        sampling    underbagging
##  Accuracy                       :    0   ADASYN:2052   Mode :logical
##  Area under the curve           :    0   FALSE :6156   FALSE:8208
##  F1 measure                     :    0   SMOTE :2052   TRUE :2052
##  G-mean                         :10260                 NA's :0
##  Matthews correlation coefficient:    0
##
##
##  tuning_measure   holdout_measure  holdout_measure_residual
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.6205   1st Qu.:0.0000   1st Qu.:0.1683
##  Median :0.9426   Median :0.7071   Median :0.4879
##  Mean   :0.7570   Mean   :0.5918   Mean   :0.4829
##  3rd Qu.:0.9950   3rd Qu.:0.9547   3rd Qu.:0.7996
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :189      NA's   :189      NA's   :189
##  iteration_count                      dataset       imba.rate
##  Min.   :1       abalone                  : 180   Min.   :0.0010
##  1st Qu.:1       adult                    : 180   1st Qu.:0.0100
##  Median :2       bank                     : 180   Median :0.0300
##  Mean   :2       car                      : 180   Mean   :0.0286
##  3rd Qu.:3       cardiotocography-10clases: 180   3rd Qu.:0.0500
##  Max.   :3       cardiotocography-3clases : 180   Max.   :0.0500
##  NA's   :189     (Other)                  :9180
```

Computando as médias das iteracoes

```
ds = group_by(ds, learner , weight_space , measure , sampling , underbagging , dataset , imba.rate)
ds = summarise(ds, tuning_measure = mean(tuning_measure), holdout_measure = mean(holdout_measure),
               holdout_measure_residual = mean(holdout_measure_residual))

ds = as.data.frame(ds)
```

Criando dataframe

```
# Dividindo o ds em n, um para cada técnica
splited_df = ds %>% group_by_at(.vars = params$columns) %>% do(vals = as.data.frame(.)) %>% select(vals)

# Juntando cada uma das partes horizontalmente em um data set
df_tec_wide = do.call("cbind", splited_df)

# Renomeando duplicacao de nomes
colnames(df_tec_wide) = make.unique(colnames(df_tec_wide))

# Selecionando apenas as medidas da performance escolhida
df_tec_wide_residual = select(df_tec_wide, matches(paste("^", params$performance, "$|", params$performa

# Renomeando colunas
new_names = NULL
for(i in (1:length(splited_df))){
  id = toString(sapply(splited_df[[i]][1, params$columns], as.character))
  new_names = c(new_names, id)
}
colnames(df_tec_wide_residual) = new_names

# Verificando a dimensao do df
dim(df_tec_wide_residual)
```

```
## [1] 228  15
```

```
# Renomeando a variavel
df = df_tec_wide_residual

head(df)
```

```
##   ADASYN, FALSE, FALSE, classif.ksvm
## 1                          0.2298821
## 2                          0.2298821
## 3                          0.3345384
## 4                          0.3591805
## 5                          0.2087155
## 6                          0.2087155
##   ADASYN, FALSE, FALSE, classif.randomForest
## 1                                  0.1970417
## 2                                  0.1970417
## 3                                  0.3302646
## 4                                  0.3628625
## 5                                  0.3227803
## 6                                         NA
##   ADASYN, FALSE, FALSE, classif.xgboost FALSE, FALSE, FALSE, classif.ksvm
## 1                            0.1383185                          0.1284832
## 2                            0.1383185                          0.1284832
## 3                            0.2179843                          0.2796001
## 4                            0.2834615                          0.3217932
## 5                            0.4670462                          0.1892378
## 6                            0.4670462                          0.1892378
##   FALSE, FALSE, FALSE, classif.randomForest
## 1                               0.009245003
## 2                               0.009245003
```

```
## 3                                0.000000000
## 4                                0.046686536
## 5                                0.485410784
## 6                                         NA
##   FALSE, FALSE, FALSE, classif.xgboost FALSE, FALSE, TRUE, classif.ksvm
## 1                           0.02231941                          0.5605192
## 2                           0.02231941                          0.5605192
## 3                           0.05945454                          0.6303596
## 4                           0.09187478                          0.6140664
## 5                           0.45120516                          0.6578023
## 6                           0.45120516                          0.6578023
##   FALSE, FALSE, TRUE, classif.randomForest
## 1                                0.5870339
## 2                                0.5870339
## 3                                0.6425348
## 4                                0.6408856
## 5                                0.8082926
## 6                                0.8082926
##   FALSE, FALSE, TRUE, classif.xgboost FALSE, TRUE, FALSE, classif.ksvm
## 1                           0.5874821                          0.1254198
## 2                           0.5874821                          0.1254198
## 3                           0.6355059                          0.2506877
## 4                           0.6484918                          0.3179478
## 5                           0.8136582                          0.1993466
## 6                           0.8136582                          0.1993466
##   FALSE, TRUE, FALSE, classif.randomForest
## 1                              0.009245003
## 2                              0.009245003
## 3                              0.016316883
## 4                              0.046686536
## 5                              0.470026646
## 6                                       NA
##   FALSE, TRUE, FALSE, classif.xgboost SMOTE, FALSE, FALSE, classif.ksvm
## 1                          0.009245003                          0.2239917
## 2                          0.009245003                          0.2239917
## 3                          0.016302738                          0.3256847
## 4                          0.072048065                          0.3728492
## 5                          0.454617465                          0.2172850
## 6                          0.454617465                          0.2172850
##   SMOTE, FALSE, FALSE, classif.randomForest
## 1                                 0.1956080
## 2                                 0.1956080
## 3                                 0.3399340
## 4                                 0.3630242
## 5                                 0.3552308
## 6                                        NA
##   SMOTE, FALSE, FALSE, classif.xgboost
## 1                            0.1399832
## 2                            0.1399832
## 3                            0.2325349
## 4                            0.2731270
## 5                            0.4490729
## 6                            0.4490729
```

```r
summary(df)
```

```
##  ADASYN, FALSE, FALSE, classif.ksvm
##  Min.   :0.00000
##  1st Qu.:0.01912
##  Median :0.20029
##  Mean   :0.27684
##  3rd Qu.:0.44320
##  Max.   :0.98958
##  NA's   :7
##  ADASYN, FALSE, FALSE, classif.randomForest
##  Min.   :0.0000
##  1st Qu.:0.2772
##  Median :0.5287
##  Mean   :0.5249
##  3rd Qu.:0.7949
##  Max.   :0.9999
##  NA's   :22
##  ADASYN, FALSE, FALSE, classif.xgboost FALSE, FALSE, FALSE, classif.ksvm
##  Min.   :0.0000                        Min.   :0.0000
##  1st Qu.:0.3282                        1st Qu.:0.0000
##  Median :0.5941                        Median :0.2039
##  Mean   :0.5824                        Mean   :0.2754
##  3rd Qu.:0.8492                        3rd Qu.:0.4020
##  Max.   :1.0000                        Max.   :1.0000
##
##  FALSE, FALSE, FALSE, classif.randomForest
##  Min.   :0.0000
##  1st Qu.:0.1502
##  Median :0.4231
##  Mean   :0.4525
##  3rd Qu.:0.7398
##  Max.   :1.0000
##  NA's   :5
##  FALSE, FALSE, FALSE, classif.xgboost FALSE, FALSE, TRUE, classif.ksvm
##  Min.   :0.0000                       Min.   :0.02295
##  1st Qu.:0.1799                       1st Qu.:0.42430
##  Median :0.4512                       Median :0.60860
##  Mean   :0.4761                       Mean   :0.58945
##  3rd Qu.:0.7605                       3rd Qu.:0.77733
##  Max.   :1.0000                       Max.   :0.99115
##
##  FALSE, FALSE, TRUE, classif.randomForest
##  Min.   :0.1064
##  1st Qu.:0.6343
##  Median :0.8094
##  Mean   :0.7556
##  3rd Qu.:0.9393
##  Max.   :0.9999
##  NA's   :5
##  FALSE, FALSE, TRUE, classif.xgboost FALSE, TRUE, FALSE, classif.ksvm
##  Min.   :0.1499                      Min.   :0.0000
##  1st Qu.:0.6056                      1st Qu.:0.0000
##  Median :0.8053                      Median :0.1974
```

```
##  Mean   :0.7440                     Mean   :0.2665
##  3rd Qu.:0.9285                     3rd Qu.:0.3978
##  Max.   :0.9999                     Max.   :1.0000
##
##  FALSE, TRUE, FALSE, classif.randomForest
##  Min.   :0.0000
##  1st Qu.:0.1633
##  Median :0.4199
##  Mean   :0.4510
##  3rd Qu.:0.7357
##  Max.   :1.0000
##  NA's   :7
##  FALSE, TRUE, FALSE, classif.xgboost SMOTE, FALSE, FALSE, classif.ksvm
##  Min.   :0.0000                     Min.   :0.00000
##  1st Qu.:0.1905                     1st Qu.:0.03675
##  Median :0.4782                     Median :0.20460
##  Mean   :0.4758                     Mean   :0.26819
##  3rd Qu.:0.7378                     3rd Qu.:0.42079
##  Max.   :1.0000                     Max.   :0.98106
##
##  SMOTE, FALSE, FALSE, classif.randomForest
##  Min.   :0.0000
##  1st Qu.:0.2431
##  Median :0.5293
##  Mean   :0.5289
##  3rd Qu.:0.8125
##  Max.   :1.0000
##  NA's   :17
##  SMOTE, FALSE, FALSE, classif.xgboost
##  Min.   :0.0000
##  1st Qu.:0.3218
##  Median :0.5848
##  Mean   :0.5809
##  3rd Qu.:0.8507
##  Max.   :1.0000
##
```

## Verificando a média de cada coluna selecionada

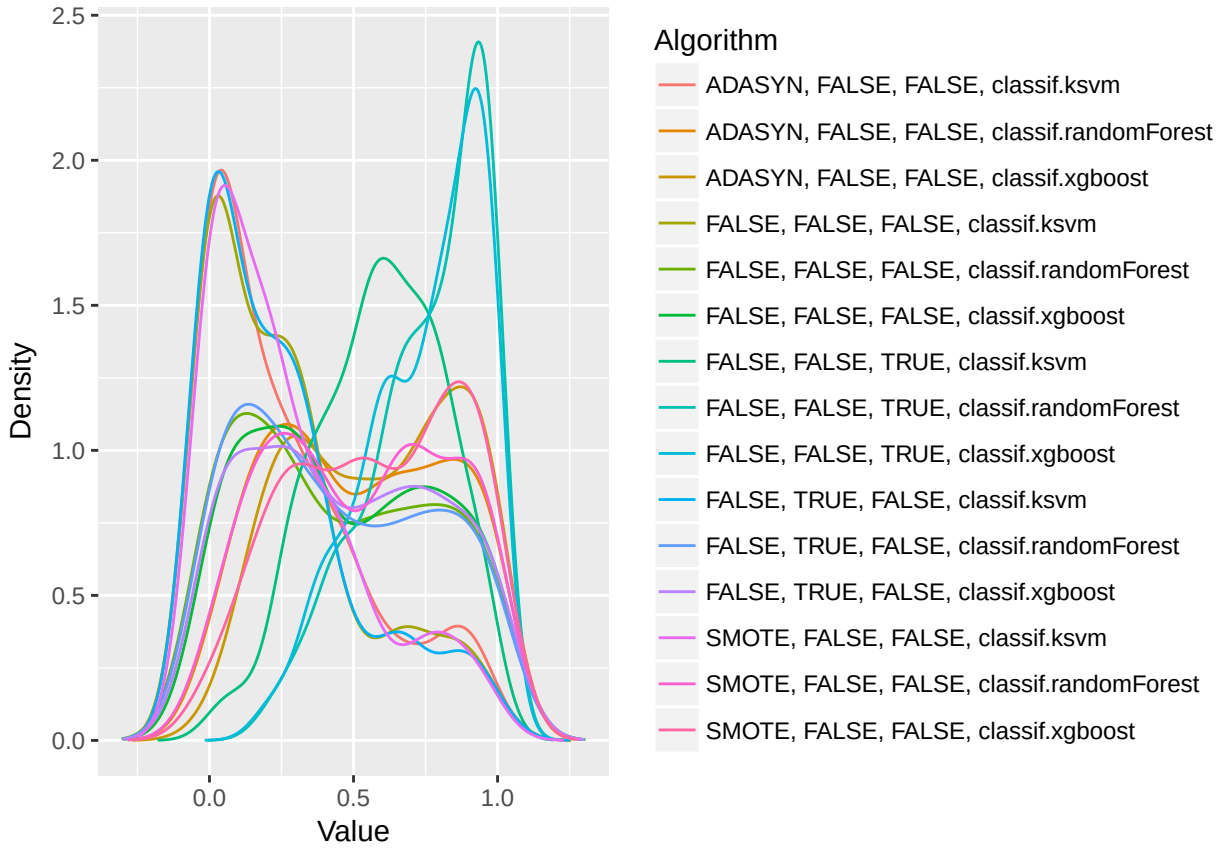
```
for(i in (1:dim(df)[2])){
  print(paste("Media da coluna ", colnames(df)[i], " = ", mean(df[,i], na.rm = TRUE), sep=""))
}
```

```
## [1] "Media da coluna ADASYN, FALSE, FALSE, classif.ksvm = 0.276838229370951"
## [1] "Media da coluna ADASYN, FALSE, FALSE, classif.randomForest = 0.524937996398"
## [1] "Media da coluna ADASYN, FALSE, FALSE, classif.xgboost = 0.582357727140875"
## [1] "Media da coluna FALSE, FALSE, FALSE, classif.ksvm = 0.275375750958183"
## [1] "Media da coluna FALSE, FALSE, FALSE, classif.randomForest = 0.452489507225035"
## [1] "Media da coluna FALSE, FALSE, FALSE, classif.xgboost = 0.476073481566564"
## [1] "Media da coluna FALSE, FALSE, TRUE, classif.ksvm = 0.589452452823026"
## [1] "Media da coluna FALSE, FALSE, TRUE, classif.randomForest = 0.7556113504064"
## [1] "Media da coluna FALSE, FALSE, TRUE, classif.xgboost = 0.744028573937436"
```

```
## [1] "Media da coluna FALSE, TRUE, FALSE, classif.ksvm = 0.26651524671154"
## [1] "Media da coluna FALSE, TRUE, FALSE, classif.randomForest = 0.450963486440501"
## [1] "Media da coluna FALSE, TRUE, FALSE, classif.xgboost = 0.475841724104327"
## [1] "Media da coluna SMOTE, FALSE, FALSE, classif.ksvm = 0.268191541263612"
## [1] "Media da coluna SMOTE, FALSE, FALSE, classif.randomForest = 0.528862349290817"
## [1] "Media da coluna SMOTE, FALSE, FALSE, classif.xgboost = 0.580920962340804"
```

## Fazendo teste de normalidade

```
plotDensities(data = na.omit(df))
```




## Testando as diferencas

```
friedmanTest(df)
```

```
##
##  Friedman's rank sum test
##
## data:  df
## Friedman's chi-squared = 1547.7, df = 14, p-value < 2.2e-16
```

## Testando as diferencas par a par

```
test <- nemenyiTest (df, alpha=0.05)
abs(test$diff.matrix) > test$statistic
```

```
##       ADASYN, FALSE, FALSE, classif.ksvm
##  [1,]                              FALSE
##  [2,]                               TRUE
##  [3,]                               TRUE
##  [4,]                              FALSE
##  [5,]                               TRUE
##  [6,]                               TRUE
##  [7,]                               TRUE
##  [8,]                               TRUE
##  [9,]                               TRUE
## [10,]                              FALSE
## [11,]                               TRUE
## [12,]                               TRUE
## [13,]                              FALSE
## [14,]                               TRUE
## [15,]                               TRUE
##       ADASYN, FALSE, FALSE, classif.randomForest
##  [1,]                                       TRUE
##  [2,]                                      FALSE
##  [3,]                                       TRUE
##  [4,]                                       TRUE
##  [5,]                                      FALSE
##  [6,]                                      FALSE
##  [7,]                                       TRUE
##  [8,]                                       TRUE
##  [9,]                                       TRUE
## [10,]                                       TRUE
## [11,]                                       TRUE
## [12,]                                      FALSE
## [13,]                                       TRUE
## [14,]                                      FALSE
## [15,]                                       TRUE
##       ADASYN, FALSE, FALSE, classif.xgboost
##  [1,]                                  TRUE
##  [2,]                                  TRUE
##  [3,]                                 FALSE
##  [4,]                                  TRUE
##  [5,]                                  TRUE
##  [6,]                                  TRUE
##  [7,]                                 FALSE
##  [8,]                                  TRUE
##  [9,]                                  TRUE
## [10,]                                  TRUE
## [11,]                                  TRUE
## [12,]                                  TRUE
## [13,]                                  TRUE
## [14,]                                  TRUE
## [15,]                                 FALSE
```

```
##       FALSE, FALSE, FALSE, classif.ksvm
##  [1,]                              FALSE
##  [2,]                               TRUE
##  [3,]                               TRUE
##  [4,]                              FALSE
##  [5,]                               TRUE
##  [6,]                               TRUE
##  [7,]                               TRUE
##  [8,]                               TRUE
##  [9,]                               TRUE
## [10,]                              FALSE
## [11,]                               TRUE
## [12,]                               TRUE
## [13,]                              FALSE
## [14,]                               TRUE
## [15,]                               TRUE
##       FALSE, FALSE, FALSE, classif.randomForest
##  [1,]                                       TRUE
##  [2,]                                      FALSE
##  [3,]                                       TRUE
##  [4,]                                       TRUE
##  [5,]                                      FALSE
##  [6,]                                      FALSE
##  [7,]                                       TRUE
##  [8,]                                       TRUE
##  [9,]                                       TRUE
## [10,]                                       TRUE
## [11,]                                      FALSE
## [12,]                                      FALSE
## [13,]                                       TRUE
## [14,]                                       TRUE
## [15,]                                       TRUE
##       FALSE, FALSE, FALSE, classif.xgboost
##  [1,]                                 TRUE
##  [2,]                                FALSE
##  [3,]                                 TRUE
##  [4,]                                 TRUE
##  [5,]                                FALSE
##  [6,]                                FALSE
##  [7,]                                 TRUE
##  [8,]                                 TRUE
##  [9,]                                 TRUE
## [10,]                                 TRUE
## [11,]                                FALSE
## [12,]                                FALSE
## [13,]                                 TRUE
## [14,]                                FALSE
## [15,]                                 TRUE
##       FALSE, FALSE, TRUE, classif.ksvm
##  [1,]                              TRUE
##  [2,]                              TRUE
##  [3,]                             FALSE
##  [4,]                              TRUE
##  [5,]                              TRUE
```

```
##  [6,]                            TRUE
##  [7,]                           FALSE
##  [8,]                            TRUE
##  [9,]                            TRUE
## [10,]                            TRUE
## [11,]                            TRUE
## [12,]                            TRUE
## [13,]                            TRUE
## [14,]                            TRUE
## [15,]                           FALSE
##       FALSE, FALSE, TRUE, classif.randomForest
##  [1,]                                     TRUE
##  [2,]                                     TRUE
##  [3,]                                     TRUE
##  [4,]                                     TRUE
##  [5,]                                     TRUE
##  [6,]                                     TRUE
##  [7,]                                     TRUE
##  [8,]                                    FALSE
##  [9,]                                    FALSE
## [10,]                                     TRUE
## [11,]                                     TRUE
## [12,]                                     TRUE
## [13,]                                     TRUE
## [14,]                                     TRUE
## [15,]                                     TRUE
##       FALSE, FALSE, TRUE, classif.xgboost FALSE, TRUE, FALSE, classif.ksvm
##  [1,]                                TRUE                              FALSE
##  [2,]                                TRUE                               TRUE
##  [3,]                                TRUE                               TRUE
##  [4,]                                TRUE                              FALSE
##  [5,]                                TRUE                               TRUE
##  [6,]                                TRUE                               TRUE
##  [7,]                                TRUE                               TRUE
##  [8,]                               FALSE                               TRUE
##  [9,]                               FALSE                               TRUE
## [10,]                                TRUE                              FALSE
## [11,]                                TRUE                               TRUE
## [12,]                                TRUE                               TRUE
## [13,]                                TRUE                              FALSE
## [14,]                                TRUE                               TRUE
## [15,]                                TRUE                               TRUE
##       FALSE, TRUE, FALSE, classif.randomForest
##  [1,]                                     TRUE
##  [2,]                                     TRUE
##  [3,]                                     TRUE
##  [4,]                                     TRUE
##  [5,]                                    FALSE
##  [6,]                                    FALSE
##  [7,]                                     TRUE
##  [8,]                                     TRUE
##  [9,]                                     TRUE
## [10,]                                     TRUE
## [11,]                                    FALSE
```

```
## [12,]                                    FALSE
## [13,]                                     TRUE
## [14,]                                     TRUE
## [15,]                                     TRUE
##       FALSE, TRUE, FALSE, classif.xgboost
##  [1,]                                TRUE
##  [2,]                               FALSE
##  [3,]                                TRUE
##  [4,]                                TRUE
##  [5,]                               FALSE
##  [6,]                               FALSE
##  [7,]                                TRUE
##  [8,]                                TRUE
##  [9,]                                TRUE
## [10,]                                TRUE
## [11,]                               FALSE
## [12,]                               FALSE
## [13,]                                TRUE
## [14,]                               FALSE
## [15,]                                TRUE
##       SMOTE, FALSE, FALSE, classif.ksvm
##  [1,]                              FALSE
##  [2,]                               TRUE
##  [3,]                               TRUE
##  [4,]                              FALSE
##  [5,]                               TRUE
##  [6,]                               TRUE
##  [7,]                               TRUE
##  [8,]                               TRUE
##  [9,]                               TRUE
## [10,]                              FALSE
## [11,]                               TRUE
## [12,]                               TRUE
## [13,]                              FALSE
## [14,]                               TRUE
## [15,]                               TRUE
##       SMOTE, FALSE, FALSE, classif.randomForest
##  [1,]                                      TRUE
##  [2,]                                     FALSE
##  [3,]                                      TRUE
##  [4,]                                      TRUE
##  [5,]                                      TRUE
##  [6,]                                     FALSE
##  [7,]                                      TRUE
##  [8,]                                      TRUE
##  [9,]                                      TRUE
## [10,]                                      TRUE
## [11,]                                      TRUE
## [12,]                                     FALSE
## [13,]                                      TRUE
## [14,]                                     FALSE
## [15,]                                      TRUE
##       SMOTE, FALSE, FALSE, classif.xgboost
##  [1,]                                TRUE
```

```
##  [2,]                                TRUE
##  [3,]                               FALSE
##  [4,]                                TRUE
##  [5,]                                TRUE
##  [6,]                                TRUE
##  [7,]                               FALSE
##  [8,]                                TRUE
##  [9,]                                TRUE
## [10,]                                TRUE
## [11,]                                TRUE
## [12,]                                TRUE
## [13,]                                TRUE
## [14,]                                TRUE
## [15,]                               FALSE
```

## Plotando os ranks

```
print(colMeans(rankMatrix(df)))
```

```
##         ADASYN, FALSE, FALSE, classif.ksvm
##                                   11.835526
## ADASYN, FALSE, FALSE, classif.randomForest
##                                    7.776316
##      ADASYN, FALSE, FALSE, classif.xgboost
##                                    5.298246
##          FALSE, FALSE, FALSE, classif.ksvm
##                                   11.673246
##  FALSE, FALSE, FALSE, classif.randomForest
##                                    9.149123
##       FALSE, FALSE, FALSE, classif.xgboost
##                                    8.335526
##           FALSE, FALSE, TRUE, classif.ksvm
##                                    6.149123
##   FALSE, FALSE, TRUE, classif.randomForest
##                                    2.660088
##        FALSE, FALSE, TRUE, classif.xgboost
##                                    2.708333
##           FALSE, TRUE, FALSE, classif.ksvm
##                                   11.861842
##   FALSE, TRUE, FALSE, classif.randomForest
##                                    9.425439
##        FALSE, TRUE, FALSE, classif.xgboost
##                                    8.438596
##          SMOTE, FALSE, FALSE, classif.ksvm
##                                   11.778509
##  SMOTE, FALSE, FALSE, classif.randomForest
##                                    7.605263
##       SMOTE, FALSE, FALSE, classif.xgboost
##                                    5.304825
```

# Plotando grafico de Critical Diference

```
result = tryCatch({
    plotCD(df, alpha=0.05, cex = 0.35)
}, error = function(e) {})
```