Aufgabe 3

busra

7 Juli 2018

```
Pfad <- "C:/BÜSRA/Uni/Master/B Fächer/Data Mining 2/HELM_2018/TRANSFER/SMOTE_1"



# load data sets
hyper <-read.csv('http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/hypothyroid.
names <- read.csv('http://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease/hypothyroid
names <- gsub(pattern =":|[.]",x = names, replacement="")
colnames(hyper) <- names

# fix variables and column headers
colnames(hyper) <-c("target", "age", "sex", "on_thyroxine", "query_on_thyroxine",
                    "on_antithyroid_medication", "thyroid_surgery", "query_hypothyroid",
                    "query_hyperthyroid", "pregnant", "sick", "tumor", "lithium",
                    "goitre", "TSH_measured", "TSH", "T3_measured", "T3", "TT4_measured",
                    "TT4", "T4U_measured", "T4U", "FTI_measured", "FTI", "TBG_measured",
                    "TBG")
```

```
## Start Manuel
write.table(x = hyper, file = paste(Pfad,"hyper_hypothyroid.csv", sep = ""), dec = ",", sep = ";", row.
## End Manuel
```

```
hyper$target <- ifelse(hyper$target=='negative',0,1)
head(hyper,2)
```

```
##   target age sex on_thyroxine query_on_thyroxine on_antithyroid_medication
## 1      1  72   M            f                  f                         f
## 2      1  15   F            t                  f                         f
##   thyroid_surgery query_hypothyroid query_hyperthyroid pregnant sick tumor
## 1               f                 f                  f        f    f     f
## 2               f                 f                  f        f    f     f
##   lithium goitre TSH_measured TSH T3_measured   T3 TT4_measured TT4
## 1       f      f            y  30           y 0.60            y  15
## 2       f      f            y 145          y 1.70            y  19
##   T4U_measured  T4U FTI_measured FTI TBG_measured TBG
## 1            y 1.48            y  10            n   ?
## 2            y 1.13            y  17            n   ?
```

```
## Start Manuel
write.table(x = hyper, file = paste(Pfad,"hyper_01.csv", sep = ""), dec = ",", sep = ";", row.names = F
## End Manuel
```

```
# check balance of outcome variable
print(table(hyper$target))
```

```
##
##    0    1
## 3012  151
```

```r
print(prop.table(table(hyper$target)))
```

```
##
##          0          1
## 0.95226051 0.04773949
```

```r
# binarize all character fields
ind <- sapply(hyper, is.factor)
hyper[ind] <- lapply(hyper[ind], as.character)

hyper[ hyper == "?" ] = NA
hyper[ hyper == "f" ] = 0
hyper[ hyper == "t" ] = 1
hyper[ hyper == "n" ] = 0
hyper[ hyper == "y" ] = 1
hyper[ hyper == "M" ] = 0
hyper[ hyper == "F" ] = 1

hyper[ind] <- lapply(hyper[ind], as.numeric)

## Start Manuel
write.table(x = hyper, file = paste(Pfad,"hyper_01_bin.csv", sep = ""), dec = ",", sep = ";", row.names
## End Manuel
```

```r
replaceNAsWithMean <- function(x) {replace(x, is.na(x), mean(x[!is.na(x)]))}

## Start Manuel

# hyper <- replaceNAsWithMean(hyper)

#repalceNAsWithMode <- function(x) {ux <- unique(x)
#replace(x, is.na(x), ux[!is.na(which.max(tabulate(match(x, ux))))])}

# hyper <- replaceNAsWithMean(hyper)

replaceNAsWithMode <- function(x) {ux <- unique(x)
replace(x, is.na(x), ux[which.max(tabulate(match(x, ux)))])}


#summary(hyper) # achte auf NA's

hyper$age <- replaceNAsWithMean(hyper$age)

hyper$sex <- replaceNAsWithMode(hyper$sex)
#hyper$sex <- replaceNAsWithMean(hyper$sex)

hyper$TSH <- replaceNAsWithMean(hyper$TSH)

hyper$T3 <- replaceNAsWithMean(hyper$T3)

hyper$TT4 <- replaceNAsWithMean(hyper$TT4)

hyper$T4U <- replaceNAsWithMean(hyper$T4U)

hyper$FTI <- replaceNAsWithMean(hyper$FTI)
```

```r
# hier müssten dann alle weiteren Variablen aufgeführt werden, bei denen missing values auftreten


# Test and Check of Preprocessing
mean(hyper$age)
```

```
## [1] 51.15421
```

```r
prop.table(table(hyper$sex))
```

```
##
##         0         1
## 0.2870692 0.7129308
```

```r
mean(hyper$TSH)
```

```
## [1] 5.92318
```

```r
summary(hyper) # achte auf NA's
```

```
##      target              age              sex          on_thyroxine
##  Min.   :0.00000   Min.   : 1.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:38.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :51.15   Median :1.0000   Median :0.0000
##  Mean   :0.04774   Mean   :51.15   Mean   :0.7129   Mean   :0.1457
##  3rd Qu.:0.00000   3rd Qu.:64.00   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :98.00   Max.   :1.0000   Max.   :1.0000
##
##  query_on_thyroxine on_antithyroid_medication thyroid_surgery
##  Min.   :0.00000    Min.   :0.00000           Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000           1st Qu.:0.00000
##  Median :0.00000    Median :0.00000           Median :0.00000
##  Mean   :0.01739    Mean   :0.01328           Mean   :0.03288
##  3rd Qu.:0.00000    3rd Qu.:0.00000           3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000           Max.   :1.00000
##
##  query_hypothyroid query_hyperthyroid    pregnant            sick
##  Min.   :0.00000   Min.   :0.00000    Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000    1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000    Median :0.00000   Median :0.0000
##  Mean   :0.07619   Mean   :0.07683    Mean   :0.01992   Mean   :0.0313
##  3rd Qu.:0.00000   3rd Qu.:0.00000    3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.00000    Max.   :1.00000   Max.   :1.0000
##
##      tumor            lithium              goitre        TSH_measured
##  Min.   :0.00000   Min.   :0.0000000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :0.00000   Median :0.0000000   Median :0.0000   Median :1.000
##  Mean   :0.01265   Mean   :0.0006323   Mean   :0.0313   Mean   :0.852
##  3rd Qu.:0.00000   3rd Qu.:0.0000000   3rd Qu.:0.0000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000000   Max.   :1.0000   Max.   :1.000
##
##       TSH           T3_measured           T3          TT4_measured
##  Min.   :  0.000   Min.   :0.0000   Min.   : 0.00   Min.   :0.0000
```

```
##  1st Qu.:  0.000   1st Qu.:1.0000   1st Qu.: 1.50   1st Qu.:1.0000
##  Median :  1.000   Median :1.0000   Median : 1.94   Median :1.0000
##  Mean   :  5.923   Mean   :0.7803   Mean   : 1.94   Mean   :0.9213
##  3rd Qu.:  5.923   3rd Qu.:1.0000   3rd Qu.: 2.20   3rd Qu.:1.0000
##  Max.   :530.000   Max.   :1.0000   Max.   :10.20   Max.   :1.0000
##
##       TT4          T4U_measured         T4U          FTI_measured
##  Min.   :  2.0   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 85.0   1st Qu.:1.0000   1st Qu.:0.8600   1st Qu.:1.0000
##  Median :107.0   Median :1.0000   Median :0.9782   Median :1.0000
##  Mean   :108.8   Mean   :0.9216   Mean   :0.9782   Mean   :0.9219
##  3rd Qu.:124.0   3rd Qu.:1.0000   3rd Qu.:1.0500   3rd Qu.:1.0000
##  Max.   :450.0   Max.   :1.0000   Max.   :2.2100   Max.   :1.0000
##
##       FTI          TBG_measured         TBG
##  Min.   :  0.0   Min.   :0.0000   Min.   :  0.00
##  1st Qu.: 92.0   1st Qu.:0.0000   1st Qu.: 21.00
##  Median :110.0   Median :0.0000   Median : 28.00
##  Mean   :115.4   Mean   :0.0822   Mean   : 31.28
##  3rd Qu.:126.0   3rd Qu.:0.0000   3rd Qu.: 34.00
##  Max.   :881.0   Max.   :1.0000   Max.   :122.00
##                                   NA's   :2903
```

```
write.table(x = hyper, file = paste(Pfad,"hyper_01_bin_mean.csv", sep = ""), dec = ",", sep = ";", row.
## End Manuel

hyper$TBG <- NULL



write.table(x = hyper, file = paste(Pfad,"hyper_01_bin_mean_0TBG.csv", sep = ""), dec = ",", sep = ";",

summary(hyper) # achte auf NA's    FINAL CHECK
```

```
##      target             age              sex          on_thyroxine
##  Min.   :0.00000   Min.   : 1.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:38.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :51.15   Median :1.0000   Median :0.0000
##  Mean   :0.04774   Mean   :51.15   Mean   :0.7129   Mean   :0.1457
##  3rd Qu.:0.00000   3rd Qu.:64.00   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :98.00   Max.   :1.0000   Max.   :1.0000
##  query_on_thyroxine on_antithyroid_medication thyroid_surgery
##  Min.   :0.00000    Min.   :0.00000           Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000           1st Qu.:0.00000
##  Median :0.00000    Median :0.00000           Median :0.00000
##  Mean   :0.01739    Mean   :0.01328           Mean   :0.03288
##  3rd Qu.:0.00000    3rd Qu.:0.00000           3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000           Max.   :1.00000
##  query_hypothyroid query_hyperthyroid    pregnant            sick
##  Min.   :0.00000   Min.   :0.00000    Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000    1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000    Median :0.00000   Median :0.0000
##  Mean   :0.07619   Mean   :0.07683    Mean   :0.01992   Mean   :0.0313
##  3rd Qu.:0.00000   3rd Qu.:0.00000    3rd Qu.:0.00000   3rd Qu.:0.0000
```

```
##  Max.   :1.00000   Max.   :1.00000    Max.   :1.00000   Max.   :1.0000
##      tumor            lithium              goitre        TSH_measured
##  Min.   :0.00000   Min.   :0.0000000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :0.00000   Median :0.0000000   Median :0.0000   Median :1.000
##  Mean   :0.01265   Mean   :0.0006323   Mean   :0.0313   Mean   :0.852
##  3rd Qu.:0.00000   3rd Qu.:0.0000000   3rd Qu.:0.0000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000000   Max.   :1.0000   Max.   :1.000
##       TSH           T3_measured           T3          TT4_measured
##  Min.   :  0.000   Min.   :0.0000   Min.   : 0.00   Min.   :0.0000
##  1st Qu.:  0.000   1st Qu.:1.0000   1st Qu.: 1.50   1st Qu.:1.0000
##  Median :  1.000   Median :1.0000   Median : 1.94   Median :1.0000
##  Mean   :  5.923   Mean   :0.7803   Mean   : 1.94   Mean   :0.9213
##  3rd Qu.:  5.923   3rd Qu.:1.0000   3rd Qu.: 2.20   3rd Qu.:1.0000
##  Max.   :530.000   Max.   :1.0000   Max.   :10.20   Max.   :1.0000
##       TT4          T4U_measured         T4U           FTI_measured
##  Min.   :  2.0   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 85.0   1st Qu.:1.0000   1st Qu.:0.8600   1st Qu.:1.0000
##  Median :107.0   Median :1.0000   Median :0.9782   Median :1.0000
##  Mean   :108.8   Mean   :0.9216   Mean   :0.9782   Mean   :0.9219
##  3rd Qu.:124.0   3rd Qu.:1.0000   3rd Qu.:1.0500   3rd Qu.:1.0000
##  Max.   :450.0   Max.   :1.0000   Max.   :2.2100   Max.   :1.0000
##       FTI          TBG_measured
##  Min.   :  0.0   Min.   :0.0000
##  1st Qu.: 92.0   1st Qu.:0.0000
##  Median :110.0   Median :0.0000
##  Mean   :115.4   Mean   :0.0822
##  3rd Qu.:126.0   3rd Qu.:0.0000
##  Max.   :881.0   Max.   :1.0000
```

split data into train and test portions (test == valid(ation) in SAS EMiner System)

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.4.4
```

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

```
## Warning: package 'lattice' was built under R version 3.4.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
set.seed(1234)
splitIndex <- createDataPartition(hyper$target, p = .50,
                                  list = FALSE,
                                  times = 1)
trainSplit <- hyper[ splitIndex,]
testSplit  <- hyper[-splitIndex,]
```

```
table(trainSplit$target)
```

```
##
```

```
##    0    1
## 1503   79
```

```
table(testSplit$target)
```

```
##
##    0    1
## 1509   72
```

```
prop.table(table(trainSplit$target))
```

```
##
##          0          1
## 0.95006321 0.04993679
```

```
prop.table(table(testSplit$target))
```

```
##
##         0         1
## 0.9544592 0.0455408
```

```
write.table(x = trainSplit, file = paste(Pfad,"hyper_01_bin_mean_trainSplit.csv", sep = ""), dec = ",",

write.table(x = testSplit, file = paste(Pfad,"hyper_01_bin_mean_testSplit.csv", sep = ""), dec = ",", s
```

```
is.factor(testSplit$target)
```

```
## [1] FALSE
```

```
is.factor(trainSplit$target)
```

```
## [1] FALSE
```

```
levels(testSplit$target)
```

```
## NULL
```

```
levels(trainSplit$target)
```

```
## NULL
```

```
#testSplit$target <- as.factor(testSplit$target)
#trainSplit$target <- as.factor(trainSplit$target)
```

model using treebag

```
ctrl <- trainControl(method = "cv", number = 5)
tbmodel <- train(target ~ ., data = trainSplit, method = "treebag",
                 trControl = ctrl)
```

```
## Warning in train.default(x, y, weights = w, ...): You are trying to do
## regression and your outcome only has two possible values Are you trying to
## do classification? If so, use a 2 level factor as your outcome column.
```

```
summary(tbmodel)
```

```
##             Length Class      Mode
## y           1582   -none-     numeric
## X              0   -none-     NULL
```

```
## mtrees        25   -none-     list
## OOB            1   -none-     logical
## comb           1   -none-     logical
## xNames        24   -none-     character
## problemType    1   -none-     character
## tuneValue      1   data.frame list
## obsLevels      1   -none-     logical
## param          0   -none-     list
```

```
predictors <- names(trainSplit)[names(trainSplit) != 'target']
pred <- predict(tbmodel$finalModel, testSplit[,predictors])


addmargins(table(pred)) # ist binär (0 / 1), wenn testSplit$target binär (factor) ist
```

```
## pred
## 0.00163331139489972 0.00555095037911386  0.0148833270018522
##                1462                   5                   1
##   0.016606211123897  0.0166316388199091  0.0271022360525626
##                   4                   3                   2
##  0.0318641408144673  0.0343569615773087  0.0358641408144673
##                   8                   9                   1
##  0.0398118199443312  0.0496488560991826   0.052982189432516
##                   2                   2                   2
##  0.0569821894325159  0.0937821894325159   0.132756548406875
##                   1                   1                   1
##   0.138300870751197   0.159100870751197   0.175218086868413
##                   1                   1                   1
##    0.22195801360834   0.239151661802112   0.273818328468779
##                   1                   1                   1
##   0.689900741027155   0.846653447413623   0.904758585518761
##                   1                   2                   3
##   0.917422678182854   0.929131538823267   0.945107993061901
##                   3                   3                   2
##   0.946133634087542   0.948551828678339    0.94925901510766
##                   2                   1                   1
##   0.949866697931221   0.950269541423449   0.954431468975766
##                   1                   1                   2
##   0.955441995291555   0.961907993061901   0.961923107771752
##                   2                  34                   2
##   0.962933634087542                 Sum
##                  11                1581
```

```
                      # ist metrisch, wenn testSplit$target nicht binär (factor) ist
```

```
pred # das heißt  0.001633311 wurde 1462 mal als wkeit berrechntet
```

```
##    [1] 0.961907993 0.917422678 0.962933634 0.001633311 0.961907993
##    [6] 0.001633311 0.961907993 0.961907993 0.962933634 0.961907993
##   [11] 0.929131539 0.962933634 0.961907993 0.904758586 0.961907993
##   [16] 0.961907993 0.961907993 0.950269541 0.961907993 0.961907993
##   [21] 0.961907993 0.001633311 0.961907993 0.001633311 0.962933634
##   [26] 0.961907993 0.056982189 0.929131539 0.961907993 0.961907993
##   [31] 0.962933634 0.961907993 0.948551829 0.961907993 0.961907993
##   [36] 0.955441995 0.954431469 0.239151662 0.962933634 0.929131539
##   [41] 0.946133634 0.273818328 0.961907993 0.961907993 0.961907993
```

```
##   [46] 0.917422678 0.949866698 0.946133634 0.961907993 0.001633311
##   [51] 0.961907993 0.961907993 0.689900741 0.001633311 0.961907993
##   [56] 0.962933634 0.962933634 0.961923108 0.846653447 0.962933634
##   [61] 0.904758586 0.846653447 0.961923108 0.955441995 0.917422678
##   [66] 0.962933634 0.961907993 0.961907993 0.954431469 0.961907993
##   [71] 0.961907993 0.961907993 0.001633311 0.001633311 0.001633311
##   [76] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##   [81] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##   [86] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##   [91] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##   [96] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [101] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [106] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [111] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [116] 0.001633311 0.132756548 0.001633311 0.001633311 0.001633311
##  [121] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [126] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [131] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [136] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [141] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [146] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [151] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [156] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [161] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [166] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [171] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [176] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [181] 0.001633311 0.001633311 0.031864141 0.001633311 0.001633311
##  [186] 0.001633311 0.001633311 0.949259015 0.001633311 0.001633311
##  [191] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [196] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [201] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [206] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [211] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [216] 0.016606211 0.001633311 0.016606211 0.001633311 0.001633311
##  [221] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [226] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [231] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [236] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [241] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [246] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [251] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [256] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [261] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [266] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [271] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [276] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [281] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [286] 0.001633311 0.001633311 0.001633311 0.001633311 0.138300871
##  [291] 0.001633311 0.001633311 0.034356962 0.001633311 0.001633311
##  [296] 0.001633311 0.001633311 0.005550950 0.001633311 0.001633311
##  [301] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [306] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [311] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
```

```
##  [316] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [321] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [326] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [331] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [336] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [341] 0.001633311 0.034356962 0.001633311 0.001633311 0.001633311
##  [346] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [351] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [356] 0.001633311 0.001633311 0.001633311 0.961907993 0.001633311
##  [361] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [366] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [371] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [376] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [381] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [386] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [391] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [396] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [401] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [406] 0.001633311 0.159100871 0.001633311 0.001633311 0.001633311
##  [411] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [416] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [421] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [426] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [431] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [436] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [441] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [446] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [451] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [456] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [461] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [466] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [471] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [476] 0.001633311 0.001633311 0.001633311 0.001633311 0.005550950
##  [481] 0.001633311 0.001633311 0.001633311 0.001633311 0.005550950
##  [486] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [491] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [496] 0.001633311 0.034356962 0.001633311 0.001633311 0.001633311
##  [501] 0.001633311 0.001633311 0.001633311 0.031864141 0.005550950
##  [506] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [511] 0.001633311 0.001633311 0.031864141 0.001633311 0.001633311
##  [516] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [521] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [526] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [531] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [536] 0.001633311 0.001633311 0.039811820 0.001633311 0.961907993
##  [541] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [546] 0.001633311 0.945107993 0.001633311 0.001633311 0.001633311
##  [551] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [556] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [561] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [566] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [571] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [576] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [581] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
```

```
##  [586] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [591] 0.001633311 0.001633311 0.001633311 0.049648856 0.001633311
##  [596] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [601] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [606] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [611] 0.001633311 0.001633311 0.016631639 0.001633311 0.001633311
##  [616] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [621] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [626] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [631] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [636] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [641] 0.001633311 0.001633311 0.052982189 0.001633311 0.001633311
##  [646] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [651] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [656] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [661] 0.001633311 0.001633311 0.031864141 0.001633311 0.001633311
##  [666] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [671] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [676] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [681] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [686] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [691] 0.001633311 0.001633311 0.001633311 0.001633311 0.027102236
##  [696] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [701] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [706] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [711] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [716] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [721] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [726] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [731] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [736] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [741] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [746] 0.001633311 0.001633311 0.001633311 0.016631639 0.001633311
##  [751] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [756] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [761] 0.001633311 0.034356962 0.001633311 0.001633311 0.001633311
##  [766] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [771] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [776] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [781] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [786] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [791] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [796] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [801] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [806] 0.027102236 0.001633311 0.001633311 0.001633311 0.001633311
##  [811] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [816] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [821] 0.001633311 0.001633311 0.034356962 0.001633311 0.001633311
##  [826] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [831] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [836] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [841] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [846] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [851] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
```

```
##  [856] 0.001633311 0.035864141 0.001633311 0.001633311 0.001633311
##  [861] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [866] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [871] 0.001633311 0.001633311 0.001633311 0.031864141 0.001633311
##  [876] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [881] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [886] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [891] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [896] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [901] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [906] 0.175218087 0.001633311 0.001633311 0.001633311 0.001633311
##  [911] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [916] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [921] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [926] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [931] 0.001633311 0.039811820 0.001633311 0.001633311 0.001633311
##  [936] 0.093782189 0.001633311 0.001633311 0.001633311 0.001633311
##  [941] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [946] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [951] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [956] 0.001633311 0.961907993 0.001633311 0.001633311 0.001633311
##  [961] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [966] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [971] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [976] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [981] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [986] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
##  [991] 0.001633311 0.001633311 0.001633311 0.005550950 0.001633311
##  [996] 0.001633311 0.001633311 0.001633311 0.001633311 0.034356962
## [1001] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1006] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1011] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1016] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1021] 0.001633311 0.001633311 0.001633311 0.221958014 0.001633311
## [1026] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1031] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1036] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1041] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1046] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1051] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1056] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1061] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1066] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1071] 0.001633311 0.001633311 0.001633311 0.031864141 0.001633311
## [1076] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1081] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1086] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1091] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1096] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1101] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1106] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1111] 0.001633311 0.001633311 0.034356962 0.001633311 0.001633311
## [1116] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1121] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
```

```
## [1126] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1131] 0.031864141 0.001633311 0.001633311 0.001633311 0.001633311
## [1136] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1141] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1146] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1151] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1156] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1161] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1166] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1171] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1176] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1181] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1186] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1191] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1196] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1201] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1206] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1211] 0.001633311 0.001633311 0.001633311 0.016606211 0.001633311
## [1216] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1221] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1226] 0.001633311 0.031864141 0.001633311 0.001633311 0.001633311
## [1231] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1236] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1241] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1246] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1251] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1256] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1261] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1266] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1271] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1276] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1281] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1286] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1291] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1296] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1301] 0.001633311 0.001633311 0.016606211 0.001633311 0.001633311
## [1306] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1311] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1316] 0.001633311 0.001633311 0.001633311 0.001633311 0.049648856
## [1321] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1326] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1331] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1336] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1341] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1346] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1351] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1356] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1361] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1366] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1371] 0.001633311 0.001633311 0.052982189 0.001633311 0.001633311
## [1376] 0.001633311 0.001633311 0.945107993 0.001633311 0.001633311
## [1381] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1386] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1391] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
```

```
## [1396] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1401] 0.001633311 0.001633311 0.001633311 0.001633311 0.016631639
## [1406] 0.001633311 0.001633311 0.001633311 0.014883327 0.001633311
## [1411] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1416] 0.962933634 0.001633311 0.001633311 0.001633311 0.001633311
## [1421] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1426] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1431] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1436] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1441] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1446] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1451] 0.001633311 0.001633311 0.001633311 0.001633311 0.904758586
## [1456] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1461] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1466] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1471] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1476] 0.001633311 0.001633311 0.001633311 0.001633311 0.034356962
## [1481] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1486] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1491] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1496] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1501] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1506] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1511] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1516] 0.001633311 0.034356962 0.001633311 0.001633311 0.001633311
## [1521] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1526] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1531] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1536] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1541] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1546] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1551] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1556] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1561] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1566] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1571] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1576] 0.001633311 0.001633311 0.001633311 0.001633311 0.001633311
## [1581] 0.001633311
```

```
# If prob > 0.5 then 1, else 0. Threshold can be set for better results
pred <- ifelse(pred > 0.5,1,0)

misClasificError <- mean(pred != testSplit$target)
print(paste('Accuracy',1-misClasificError))
```

```
## [1] "Accuracy 0.989247311827957"
```

```
# Confusion matrix
library(caret)
###hier immer data` and `reference` should be factors with the same levels. Fehler falle keine  table()
confusionMatrix(table(data=pred, reference=testSplit$target))
```

```
## Confusion Matrix and Statistics
##
##     reference
## data    0    1
```

```
##    0 1501    9
##    1    8   63
##
##                Accuracy : 0.9892
##                  95% CI : (0.9828, 0.9937)
##     No Information Rate : 0.9545
##     P-Value [Acc > NIR] : 2.733e-15
##
##                   Kappa : 0.8755
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9947
##             Specificity : 0.8750
##          Pos Pred Value : 0.9940
##          Neg Pred Value : 0.8873
##              Prevalence : 0.9545
##          Detection Rate : 0.9494
##    Detection Prevalence : 0.9551
##       Balanced Accuracy : 0.9348
##
##        'Positive' Class : 0
##
```

```
# ROC and AUC
library(ROCR)
```

```
## Warning: package 'ROCR' was built under R version 3.4.3
```

```
## Loading required package: gplots
```

```
## Warning: package 'gplots' was built under R version 3.4.3
```

```
##
## Attaching package: 'gplots'
```

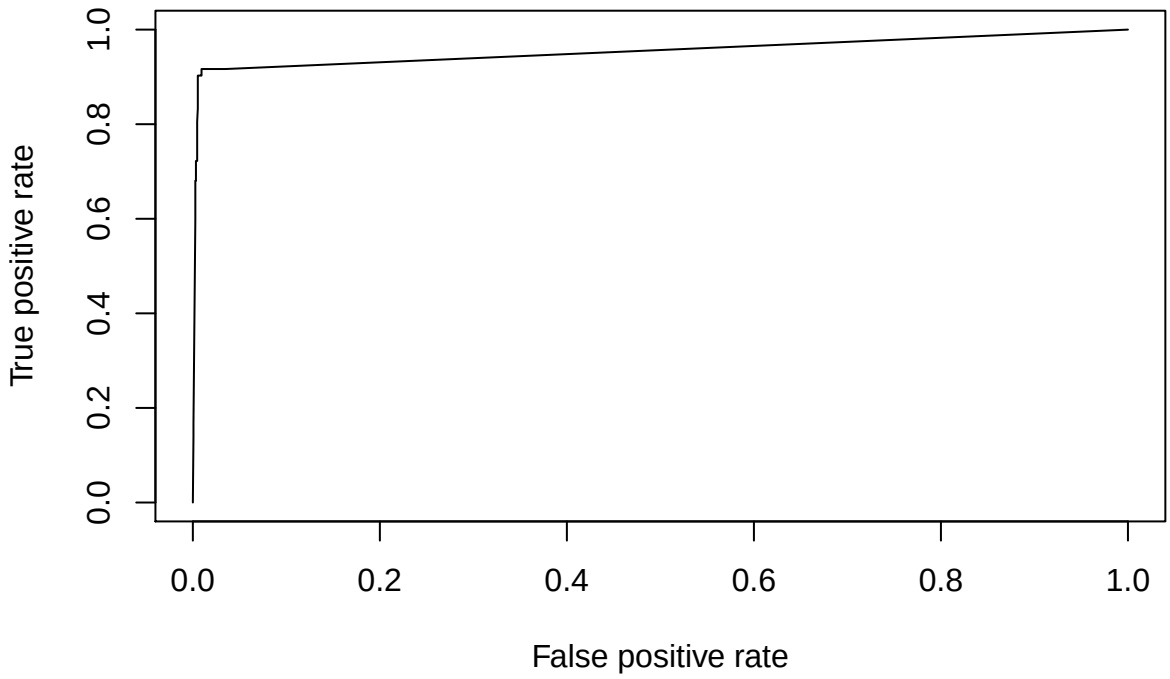
```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
ptree <- predict(tbmodel$finalModel, testSplit[,predictors])
prtree <- prediction(ptree, testSplit$target)
#summary(p)
#p

# neu2
#pr
summary(prtree)
```

```
##     Length      Class       Mode
##          1 prediction         S4
```

```
#summary(testSplit$target)
#test$Survived
#falsch---- muss testSplit heißen
#testSplit$Survived
#falsch - hier NULL -- war ja klar, denn es die abh Var ist hier TARGET!!
#testSplit$target
####Anscheinen die erstn 90 Target = 1 und ab 91 bis 1000 Target = 0
```

```
#### DAHER unten gleich mit PAcket Data Mining with R -DMwR Smote funktion anwenden!!!

# Bitte Inhalt von  Objekt  pr  anschauen :
# pr
# TPR = sensitivity, FPR=specificity
prftree <- performance(prtree, measure = "tpr", x.measure = "fpr")
plot(prftree)
```




```
auctree <- performance(prtree, measure = "auc")
auctree <- auctree@y.values[[1]]
auctree
```

```
## [1] 0.9547115
```

```
library(DMwR)
```

```
## Warning: package 'DMwR' was built under R version 3.4.3
```

```
## Loading required package: grid
```

```
testSplit$target <- as.factor(testSplit$target)
trainSplit$target <- as.factor(trainSplit$target)

#trainSplit$target <- as.factor(trainSplit$target)
trainSplit <- SMOTE(target ~ ., trainSplit, perc.over = 100, perc.under=200)
# trainSplit$target <- as.numeric(trainSplit$target)
```

```
write.table(x = trainSplit, file = paste(Pfad,"hyper_01_bin_mean_trainSplit_SMOTE.csv", sep = ""), dec =
```

```
table(trainSplit$target)
```

```
##
##   0   1
## 158 158
```

```
prop.table(table(trainSplit$target))
```

```
##
##   0   1
## 0.5 0.5
```

```
##die WKEITEN die 1 und 0 aufgeteilt sind hier das idealfall--also 50/50
```

```
# evaluate the SMOTE performance
tbmodel <- train(target ~ ., data = trainSplit, method = "treebag",
                 trControl = ctrl)

predictors <- names(trainSplit)[names(trainSplit) != 'target']
pred <- predict(tbmodel$finalModel, testSplit[,predictors])

addmargins(table(pred))
```

```
## pred
##    0    1  Sum
## 1480  101 1581
```

```
pred
```

```
##    [1] 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1
##   [35] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [69] 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [103] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [137] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [171] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [205] 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [239] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [273] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [307] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [341] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
##  [375] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
##  [409] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [443] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [477] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
##  [511] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
##  [545] 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [579] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [613] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
##  [647] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [681] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [715] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [749] 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [783] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [817] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
##  [851] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [885] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
##  [919] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [953] 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [987] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1021] 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1055] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1089] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## [1123] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1157] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1191] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
## [1225] 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1259] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1293] 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
## [1327] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1361] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1395] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [1429] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
## [1463] 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1497] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1531] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [1565] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## Levels: 0 1
```

```
# If prob > 0.5 then 1, else 0. Threshold can be set for better results
#pred <- ifelse(pred = 1,1,0)

misClasificError <- mean(pred != testSplit$target)
print(paste('Accuracy',1-misClasificError))
```

```
## [1] "Accuracy 0.974067046173308"
```

```
# Confusion matrix
library(caret)
###hier immer data` and `reference` should be factors with the same levels. Fehler falle keine  table()
confusionMatrix(table(data=pred, reference=testSplit$target))
```

```
## Confusion Matrix and Statistics
##
##     reference
## data    0    1
##    0 1474    6
##    1   35   66
##
##                Accuracy : 0.9741
##                  95% CI : (0.965, 0.9813)
##     No Information Rate : 0.9545
##     P-Value [Acc > NIR] : 3.669e-05
##
##                   Kappa : 0.7497
##  Mcnemar's Test P-Value : 1.226e-05
##
##             Sensitivity : 0.9768
##             Specificity : 0.9167
##          Pos Pred Value : 0.9959
##          Neg Pred Value : 0.6535
```

```
##              Prevalence : 0.9545
##          Detection Rate : 0.9323
##    Detection Prevalence : 0.9361
##       Balanced Accuracy : 0.9467
##
##        'Positive' Class : 0
##
```

```
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 3.4.4
```

```
## Type 'citation("pROC")' for a citation.
```

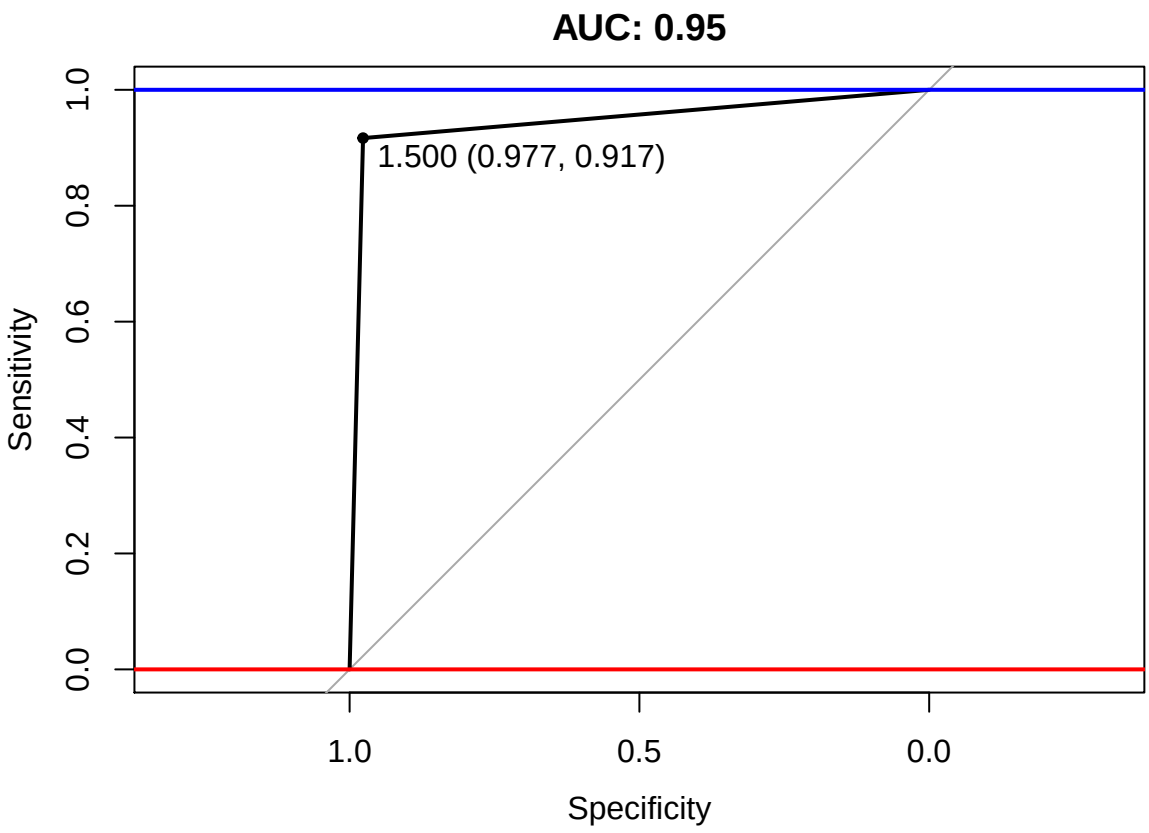
```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
# neu
pred <- as.numeric(pred)

auc <- roc(testSplit$target, pred)
print(auc)
```

```
##
## Call:
## roc.default(response = testSplit$target, predictor = pred)
##
## Data: pred in 1509 controls (testSplit$target 0) < 72 cases (testSplit$target 1).
## Area under the curve: 0.9467
```

```
plot(auc, ylim=c(0,1), print.thres=TRUE, main=paste('AUC:',round(auc$auc[[1]],2)))
abline(h=1,col='blue',lwd=2)
abline(h=0,col='red',lwd=2)
```


AUC: 0.95
1.500 (0.977, 0.917)
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.0
0.5
0.0
Specificity