

Customer_Churn.R

vitor

2021-01-10

```
#Package's
# install.packages('data.table')
# install.packages("C50")
# install.packages("ggplot2")
# install.packages("caret")
# install.packages("ROSE")
# install.packages("corrgram")
# install.packages("gmodels")
```

```
#Library's
library(gmodels)
library(data.table)
library(C50)
library(ggplot2)
library(caret)
library(ROSE)
library(corrgram)
library(randomForest)
```

```
# Loading datasets
df_train <- fread("df_train.csv")
df_test <- fread("df_test.csv")
```

```
# Data Visualization
str(df_train)
```

```
## Classes 'data.table' and 'data.frame':  1667 obs. of  21 variables:
##  $ V1                      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ state                   : chr  "HI" "MT" "OH" "NM" ...
##  $ account_length         : int  101 137 103 99 108 117 63 94 138 128 ...
##  $ area_code               : chr  "area_code_510" "area_code_510" "area_code_408" "area_code_41
##  $ international_plan     : chr  "no" "no" "no" "no" ...
##  $ voice_mail_plan        : chr  "no" "no" "yes" "no" ...
##  $ number_vmail_messages  : int  0 0 29 0 0 0 32 0 0 43 ...
##  $ total_day_minutes       : num  70.9 223.6 294.7 216.8 197.4 ...
##  $ total_day_calls         : int  123 86 95 123 78 85 124 97 117 100 ...
##  $ total_day_charge        : num  12.1 38 50.1 36.9 33.6 ...
##  $ total_eve_minutes       : num  212 245 237 126 124 ...
##  $ total_eve_calls         : int  73 139 105 88 101 68 125 112 46 89 ...
##  $ total_eve_charge        : num  18 20.8 20.2 10.7 10.5 ...
```

```
## $ total_night_minutes      : num  236 94.2 300.3 220.6 204.5 ...
## $ total_night_calls       : int   73 81 127 82 107 90 120 106 71 92 ...
## $ total_night_charge      : num   10.62 4.24 13.51 9.93 9.2 ...
## $ total_intl_minutes      : num   10.6 9.5 13.7 15.7 7.7 6.9 12.9 11.1 9.9 11.9 ...
## $ total_intl_calls        : int    3 7 6 2 4 5 3 6 4 1 ...
## $ total_intl_charge       : num    2.86 2.57 3.7 4.24 2.08 1.86 3.48 3 2.67 3.21 ...
## $ number_customer_service_calls: int   3 0 1 1 2 1 1 0 2 0 ...
## $ churn                   : chr   "no" "no" "no" "no" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
str(df_test)
```

```
## Classes 'data.table' and 'data.frame':  3333 obs. of  21 variables:
## $ V1                      : int   1 2 3 4 5 6 7 8 9 10 ...
## $ state                   : chr   "KS" "OH" "NJ" "OH" ...
## $ account_length         : int  128 107 137 84 75 118 121 147 117 141 ...
## $ area_code              : chr   "area_code_415" "area_code_415" "area_code_415" "area_code_403" ...
## $ international_plan     : chr   "no" "no" "no" "yes" ...
## $ voice_mail_plan        : chr   "yes" "yes" "no" "no" ...
## $ number_vmail_messages  : int   25 26 0 0 0 0 24 0 0 37 ...
## $ total_day_minutes      : num   265 162 243 299 167 ...
## $ total_day_calls        : int  110 123 114 71 113 98 88 79 97 84 ...
## $ total_day_charge       : num   45.1 27.5 41.4 50.9 28.3 ...
## $ total_eve_minutes      : num  197.4 195.5 121.2 61.9 148.3 ...
## $ total_eve_calls        : int   99 103 110 88 122 101 108 94 80 111 ...
## $ total_eve_charge       : num   16.78 16.62 10.3 5.26 12.61 ...
## $ total_night_minutes    : num   245 254 163 197 187 ...
## $ total_night_calls      : int   91 103 104 89 121 118 118 96 90 97 ...
## $ total_night_charge     : num   11.01 11.45 7.32 8.86 8.41 ...
## $ total_intl_minutes     : num    10 13.7 12.2 6.6 10.1 6.3 7.5 7.1 8.7 11.2 ...
## $ total_intl_calls       : int    3 3 5 7 3 6 7 6 4 5 ...
## $ total_intl_charge      : num    2.7 3.7 3.29 1.78 2.73 1.7 2.03 1.92 2.35 3.02 ...
## $ number_customer_service_calls: int   1 1 0 2 3 0 3 0 1 0 ...
## $ churn                  : chr   "no" "no" "no" "no" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
View(df_train)
```

```
View(df_test)
```

```
# Data Pre-processing
```

```
df_train$churn <- sapply(df_train$churn, function(x){ifelse(x == 'no',0,1)})
```

```
df_test$churn <- sapply(df_test$churn, function(x){ifelse(x == 'no',0,1)})
```

```
#Counting Churn
```

```
prop.table(table(df_train$churn))
```

```
##
##          0          1
## 0.8656269 0.1343731
```

```
# Applying Over-Undersampling
```

```
train <- ovun.sample(churn ~., data = df_train, method = 'both', N = 37000)$data  
str(train)
```

```
## 'data.frame': 37000 obs. of 21 variables:  
## $ V1 : int 524 710 1289 876 1148 1003 593 97 836 635 ...  
## $ state : chr "CO" "KS" "WY" "UT" ...  
## $ account_length : int 49 114 117 82 114 136 70 103 39 91 ...  
## $ area_code : chr "area_code_415" "area_code_415" "area_code_408" "area_code_511" ...  
## $ international_plan : chr "no" "no" "no" "no" ...  
## $ voice_mail_plan : chr "no" "no" "no" "no" ...  
## $ number_vmail_messages : int 0 0 0 0 0 0 0 0 38 0 ...  
## $ total_day_minutes : num 172 209 150 192 140 ...  
## $ total_day_calls : int 142 127 121 95 152 92 82 81 66 90 ...  
## $ total_day_charge : num 29.2 35.6 25.6 32.6 23.8 ...  
## $ total_eve_minutes : num 224 165 345 117 216 ...  
## $ total_eve_calls : int 100 93 132 105 76 109 112 88 87 102 ...  
## $ total_eve_charge : num 19.04 14.05 29.33 9.92 18.35 ...  
## $ total_night_minutes : num 177.3 196.8 183.9 202.6 96.9 ...  
## $ total_night_calls : int 86 98 97 113 111 67 108 103 112 113 ...  
## $ total_night_charge : num 7.98 8.86 8.28 9.12 4.36 8.46 8.09 8.72 7.82 10 ...  
## $ total_intl_minutes : num 7.4 10.6 8.3 8.2 7.9 12.3 12.6 11 9.5 14.8 ...  
## $ total_intl_calls : int 1 8 2 5 10 4 1 3 3 3 ...  
## $ total_intl_charge : num 2 2.86 2.24 2.21 2.13 3.32 3.4 2.97 2.57 4 ...  
## $ number_customer_service_calls: int 0 3 0 0 2 2 1 2 0 0 ...  
## $ churn : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
test <- ovun.sample(churn ~., data = df_test, method = 'both', N = 20000)$data  
str(test)
```

```
## 'data.frame': 20000 obs. of 21 variables:  
## $ V1 : int 706 2315 461 2957 129 2244 2532 1519 488 1855 ...  
## $ state : chr "HI" "VA" "UT" "NM" ...  
## $ account_length : int 99 43 90 112 92 70 180 104 76 80 ...  
## $ area_code : chr "area_code_415" "area_code_408" "area_code_415" "area_code_411" ...  
## $ international_plan : chr "no" "no" "no" "no" ...  
## $ voice_mail_plan : chr "no" "yes" "no" "no" ...  
## $ number_vmail_messages : int 0 35 0 0 0 0 0 0 0 0 ...  
## $ total_day_minutes : num 62.9 200.2 261.8 81.6 91.7 ...  
## $ total_day_calls : int 81 105 128 94 90 108 134 78 100 138 ...  
## $ total_day_charge : num 10.7 34 44.5 13.9 15.6 ...  
## $ total_eve_minutes : num 231 244 221 268 194 ...  
## $ total_eve_calls : int 64 88 104 112 123 116 113 119 139 91 ...  
## $ total_eve_charge : num 19.6 20.8 18.8 22.8 16.5 ...  
## $ total_night_minutes : num 169 207 137 141 175 ...  
## $ total_night_calls : int 121 97 91 75 86 96 87 102 105 94 ...  
## $ total_night_charge : num 7.6 9.32 6.15 6.34 7.88 ...  
## $ total_intl_minutes : num 8.5 11.6 9.6 8.6 9.2 8.6 10.1 11.3 10.5 5.5 ...  
## $ total_intl_calls : int 5 4 5 18 4 4 4 5 2 4 ...  
## $ total_intl_charge : num 2.3 3.13 2.59 2.32 2.48 2.32 2.73 3.05 2.84 1.49 ...  
## $ number_customer_service_calls: int 1 3 1 1 2 2 1 2 0 1 ...  
## $ churn : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
#Removing V1 from dataset
```

```
train$V1 <- NULL
```

```
test$V1 <- NULL
```

```
#Veryfing NA values
```

```
table(is.na(train))
```

```
##
```

```
## FALSE
```

```
## 740000
```

```
table(is.na(test))
```

```
##
```

```
## FALSE
```

```
## 400000
```

```
#Categorization Function
```

```
catfun <- function(dataset, features){
```

```
  for (feature in features){
```

```
    dataset[[feature]] <- as.factor(dataset[[feature]])
```

```
  }
```

```
  return(dataset)
```

```
}
```

```
#Categorization
```

```
str(train)
```

```
## 'data.frame': 37000 obs. of 20 variables:
```

```
## $ state : chr "CO" "KS" "WY" "UT" ...
```

```
## $ account_length : int 49 114 117 82 114 136 70 103 39 91 ...
```

```
## $ area_code : chr "area_code_415" "area_code_415" "area_code_408" "area_code_51
```

```
## $ international_plan : chr "no" "no" "no" "no" ...
```

```
## $ voice_mail_plan : chr "no" "no" "no" "no" ...
```

```
## $ number_vmail_messages : int 0 0 0 0 0 0 0 0 38 0 ...
```

```
## $ total_day_minutes : num 172 209 150 192 140 ...
```

```
## $ total_day_calls : int 142 127 121 95 152 92 82 81 66 90 ...
```

```
## $ total_day_charge : num 29.2 35.6 25.6 32.6 23.8 ...
```

```
## $ total_eve_minutes : num 224 165 345 117 216 ...
```

```
## $ total_eve_calls : int 100 93 132 105 76 109 112 88 87 102 ...
```

```
## $ total_eve_charge : num 19.04 14.05 29.33 9.92 18.35 ...
```

```
## $ total_night_minutes : num 177.3 196.8 183.9 202.6 96.9 ...
```

```
## $ total_night_calls : int 86 98 97 113 111 67 108 103 112 113 ...
```

```
## $ total_night_charge : num 7.98 8.86 8.28 9.12 4.36 8.46 8.09 8.72 7.82 10 ...
```

```
## $ total_intl_minutes : num 7.4 10.6 8.3 8.2 7.9 12.3 12.6 11 9.5 14.8 ...
```

```
## $ total_intl_calls : int 1 8 2 5 10 4 1 3 3 3 ...
```

```
## $ total_intl_charge : num 2 2.86 2.24 2.21 2.13 3.32 3.4 2.97 2.57 4 ...
```

```
## $ number_customer_service_calls: int 0 3 0 0 2 2 1 2 0 0 ...
```

```
## $ churn : num 0 0 0 0 0 0 0 0 0 0 ...
```

```

cat<- c('state', 'area_code', 'international_plan','voice_mail_plan', 'churn' )

train <- catfun(train, cat)
test  <- catfun(test, cat)

#Feature Selection before real train model
View(train)
randf <- C5.0(train[,-20], train$churn)
varImp(randf)

```

```

##                                Overall
## total_day_minutes             100.00
## state                         86.57
## number_customer_service_calls 79.76
## international_plan            65.74
## total_eve_charge              52.45
## voice_mail_plan               30.86
## total_day_calls               25.18
## account_length               25.06
## total_night_minutes           22.78
## total_night_charge            14.18
## total_intl_minutes            12.19
## total_intl_calls              10.85
## total_night_calls             10.35
## total_eve_calls               3.67
## total_eve_minutes             2.00
## area_code                     1.67
## total_intl_charge              0.40
## number_vmail_messages         0.00
## total_day_charge              0.00

```

```

#Model 1 with all features

```

```

modelv1 <- C5.0(train[,-20], train$churn)

summary(modelv1)

```

```

##
## Call:
## C5.0.default(x = train[, -20], y = train$churn)
##
##
## C5.0 [Release 2.07 GPL Edition]      Sun Jan 10 15:31:22 2021
## -----
##
## Class specified by attribute 'outcome'
##
## Read 37000 cases (20 attributes) from undefined.data
##
## Decision tree:
##
## total_day_minutes > 251.1:

```

```

## :...voice_mail_plan = yes:
## :   :...state in {AK,AL,AR,AZ,CA,CO,CT,DC,DE,FL,GA,HI,IA,ID,IL,IN,KS,KY,LA,MA,
## :   :   :       MD,ME,MI,MN,MO,MS,MT,NC,ND,NE,NH,NJ,NM,NY,OH,OK,OR,PA,RI,SC,
## :   :   :       SD,TN,TX,UT,VA,VT,WA,WI,WV,WY}: 0 (380)
## :   :   state = NV: 1 (92)
## :   voice_mail_plan = no:
## :   :...total_eve_charge > 16.56:
## :       :...total_night_minutes > 143.7: 1 (4262)
## :       :   total_night_minutes <= 143.7:
## :       :       :...total_day_minutes <= 275.5: 0 (37)
## :       :       :   total_day_minutes > 275.5: 1 (350)
## :       total_eve_charge <= 16.56:
## :       :...state in {AL,CA,CO,CT,DC,DE,GA,IA,ID,IL,LA,ME,MI,MN,MO,MS,MT,NC,ND,
## :       :       :       NE,NM,OH,OR,PA,RI,SC,TN,VT,WI,WV,WY}: 0 (250)
## :       :       state in {AK,AR,AZ,FL,HI,IN,KS,KY,MA,MD,NH,NJ,NV,NY,OK,SD,TX,UT,VA,
## :       :       :       WA}:
## :       :       :...total_day_minutes > 288.2: 1 (1292)
## :       :       :   total_day_minutes <= 288.2:
## :       :       :       :...state in {AK,AZ,FL,HI,IN,KS,KY,MA,NH,NJ,NV,OK,SD,VA,
## :       :       :       :       :       WA}: 0 (177)
## :       :       :       state in {AR,MD,NY,TX,UT}:
## :       :       :       :...total_day_minutes <= 257.9: 0 (56)
## :       :       :       :   total_day_minutes > 257.9:
## :       :       :       :       :...total_day_calls <= 106: 1 (486)
## :       :       :       :       :   total_day_calls > 106:
## :       :       :       :       :       :...total_night_charge <= 12.04: 0 (29)
## :       :       :       :       :       :   total_night_charge > 12.04: 1 (78)
## total_day_minutes <= 251.1:
## :...number_customer_service_calls > 3:
## :   :...total_day_minutes <= 162.6:
## :       :   :...state in {AR,CA,CT,DE,FL,ID,KY,LA,MA,ME,MI,MN,MO,MS,MT,NE,NH,NM,NY,
## :       :       :       OH,OR,SD,TN,TX,UT,WI,WV}: 1 (3057)
## :       :       :   state in {AK,AL,AZ,CO,DC,GA,HI,IA,IL,IN,KS,NC,ND,NV,OK,PA,RI,SC,VA,
## :       :       :       :       VT,WA,WY}: 0 (31)
## :       :       :   state in {MD,NJ}:
## :       :       :       :...total_day_minutes <= 130.5: 1 (256)
## :       :       :       :   total_day_minutes > 130.5: 0 (27)
## :       total_day_minutes > 162.6:
## :       :...total_intl_minutes <= 6: 0 (95)
## :       :   total_intl_minutes > 6:
## :       :       :...account_length <= 62:
## :       :       :       :...international_plan = no: 0 (146)
## :       :       :       :   international_plan = yes: 1 (79)
## :       :       :       :   account_length > 62:
## :       :       :       :       :...state in {DC,GA,IA,KS,LA,MA,MT,TX}: 1 (721)
## :       :       :       :       :   state in {AK,AL,AR,AZ,CA,CO,CT,DE,FL,HI,IL,IN,KY,MD,MI,MN,
## :       :       :       :       :       :       MO,MS,NC,ND,NE,NH,NM,NV,NY,OH,OK,OR,PA,RI,SC,SD,
## :       :       :       :       :       :       :       TN,UT,VA,WI,WY}: 0 (376)
## :       :       :       :       :   state in {ID,ME,NJ,VT,WA,WV}:
## :       :       :       :       :       :...total_day_minutes > 201.1: 0 (63)
## :       :       :       :       :       :   total_day_minutes <= 201.1:
## :       :       :       :       :       :       :...total_day_minutes > 195: 1 (355)
## :       :       :       :       :       :       :   total_day_minutes <= 195:
## :       :       :       :       :       :       :       :...total_eve_calls <= 92: 1 (153)

```



```

##                                     :...account_length > 165:
##                                     :...state in {DC,ID,LA,TX}: 1 (0)
##                                     :   state in {AL,AZ,CA,IL,KY,OH,TN,UT,
##                                     :   :           WV}: 0 (139)
##                                     :   state in {IN,MN,OR,WA}:
##                                     :   :...account_length > 186: 0 (27)
##                                     :       account_length <= 186:
##                                     :       :...total_eve_calls <= 86: 0 (26)
##                                     :       total_eve_calls > 86: 1 (361)
##                                     account_length <= 165:
##                                     :...total_night_charge <= 4.73:
##                                     :...state in {AL,AZ,CA,DC,ID,IL,KY,MN,OR,
##                                     :   :           TN,TX,WA}: 1 (0)
##                                     :   state in {IN,LA,OH}: 0 (37)
##                                     :   state in {UT,WV}:
##                                     :   :...voice_mail_plan = no: 1 (168)
##                                     :       voice_mail_plan = yes: 0 (10)
##                                     total_night_charge > 4.73:
##                                     :...state in {MN,OR,WV}: 0 (773)
##                                     :       state in {AL,AZ,CA,DC,ID,IL,IN,KY,LA,
##                                     :       :           OH,TN,TX,UT,WA}:
##                                     :       :...account_length > 149: 0 (180)
##                                     :       account_length <= 149: [S1]
##
## SubTree [S1]
##
## total_night_charge <= 5.82: 0 (191)
## total_night_charge > 5.82:
## :...total_night_minutes <= 131.5: 1 (76)
##     total_night_minutes > 131.5:
##     :...account_length > 148:
##     :...state = CA: 0 (10)
##     :   state in {AL,AZ,DC,ID,IL,IN,KY,LA,OH,TN,TX,UT,WA}: 1 (85)
##     account_length <= 148:
##     :...total_day_calls > 111:
##     :...state = UT: 1 (0)
##     :   state in {AL,DC,IL,IN,KY,OH,TX,WA}: 0 (215)
##     :   state in {AZ,CA,ID,LA,TN}:
##     :   :...voice_mail_plan = yes: 0 (22)
##     :       voice_mail_plan = no:
##     :       :...total_night_calls <= 111: 1 (415)
##     :           total_night_calls > 111:
##     :           :...state in {AZ,CA,ID,TN}: 1 (73)
##     :               state = LA: 0 (14)
##     total_day_calls <= 111:
##     :...number_customer_service_calls > 2:
##     :...state in {DC,UT}: 1 (0)
##     :   state in {AZ,CA,ID,IN,KY,LA,TN,WA}: 0 (214)
##     :   state in {AL,IL,OH,TX}:
##     :   :...total_night_calls <= 81: 0 (31)
##     :       total_night_calls > 81:
##     :       :...total_eve_calls <= 94: 0 (27)
##     :           total_eve_calls > 94:
##     :           :...voice_mail_plan = no: 1 (409)

```



```

##          :          voice_mail_plan = yes: 0 (4)
## number_customer_service_calls <= 2:
##      :...state in {AL,AZ,CA,DC,ID,IL,LA,OH,TN,TX}: 0 (1447)
##          state in {IN,KY,UT,WA}:
##      :...total_eve_minutes > 185.6: 0 (243)
##          total_eve_minutes <= 185.6:
##      :...total_night_calls > 119: 0 (70)
##          total_night_calls <= 119:
##      :...area_code = area_code_510: 0 (65)
##          area_code in {area_code_408,area_code_415}:
##      :...total_night_charge <= 7.98: 0 (37)
##          total_night_charge > 7.98:
##      :...total_eve_calls <= 89: 1 (147)
##          total_eve_calls > 89:
##      :...total_night_calls > 98: 0 (28)
##          total_night_calls <= 98:
##      :...total_intl_charge <= 2.62: 0 (10)
##          total_intl_charge > 2.62: 1 (139)
##
##
## Evaluation on training data (37000 cases):
##
##      Decision Tree
##      -----
##      Size      Errors
##
##      78      0( 0.0%)   <<
##
##
##      (a)   (b)   <-classified as
##      ----  ----
##      18588      (a): class 0
##      18412      (b): class 1
##
##
## Attribute usage:
##
## 100.00% total_day_minutes
## 86.57% state
## 79.76% number_customer_service_calls
## 65.74% international_plan
## 52.45% total_eve_charge
## 30.86% voice_mail_plan
## 25.18% total_day_calls
## 25.06% account_length
## 22.78% total_night_minutes
## 14.18% total_night_charge
## 12.19% total_intl_minutes
## 10.85% total_intl_calls
## 10.35% total_night_calls
## 3.67% total_eve_calls
## 2.00% total_eve_minutes
## 1.67% area_code
## 0.40% total_intl_charge

```

```
##
##
## Time: 0.4 secs
```

#Predict

```
pred1 <- predict(modelv1, test)
CrossTable(test$churn, pred1)
```

```
##
##
##      Cell Contents
## |-----|
## |                N |
## | Chi-square contribution |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  20000
##
##
##      test$churn | pred1
##      test$churn |      0 |      1 | Row Total |
## -----|-----|-----|-----|
##           0 |    9390 |    532 |    9922 |
##           |  844.517 | 1988.957 |          |
##           |    0.946 |    0.054 |    0.496 |
##           |    0.669 |    0.089 |          |
##           |    0.469 |    0.027 |          |
## -----|-----|-----|-----|
##           1 |    4649 |    5429 |    10078 |
##           |  831.444 | 1958.170 |          |
##           |    0.461 |    0.539 |    0.504 |
##           |    0.331 |    0.911 |          |
##           |    0.232 |    0.271 |          |
## -----|-----|-----|-----|
## Column Total |    14039 |    5961 |    20000 |
##           |    0.702 |    0.298 |          |
## -----|-----|-----|-----|
##
##
```

#2° model

```
normfun <- function(dataset, fatures){
  for(feature in features){
    dataset[[feature]] <- scale(dataset[[feature]])
  }
  return(dataset)
}
```

```
colnames(train)
```

```
## [1] "state"
## [2] "account_length"
## [3] "area_code"
## [4] "international_plan"
## [5] "voice_mail_plan"
## [6] "number_vmail_messages"
## [7] "total_day_minutes"
## [8] "total_day_calls"
## [9] "total_day_charge"
## [10] "total_eve_minutes"
## [11] "total_eve_calls"
## [12] "total_eve_charge"
## [13] "total_night_minutes"
## [14] "total_night_calls"
## [15] "total_night_charge"
## [16] "total_intl_minutes"
## [17] "total_intl_calls"
## [18] "total_intl_charge"
## [19] "number_customer_service_calls"
## [20] "churn"
```

```
norm <- c('number_vmail_messages', "total_day_minutes", "total_day_minutes", "total_day_calls",
         "total_day_charge", "total_eve_minutes", "total_eve_calls", "total_eve_charge", "total_night_minu",
         "total_night_calls", "total_night_charge", "total_intl_minutes", "total_intl_calls", "total_intl_
```

```
train <- scalfun(train, norm)
test <- scalfun(test, norm)
```

```
#model creation
```

```
modelv2 <- C5.0(train[, -20], train$churn, trials = 5)
summary(modelv2)
```

```
##
## Call:
## C5.0.default(x = train[, -20], y = train$churn, trials = 5)
##
##
## C5.0 [Release 2.07 GPL Edition]          Sun Jan 10 15:31:24 2021
## -----
##
## Class specified by attribute 'outcome'
##
## Read 37000 cases (20 attributes) from undefined.data
##
## ----- Trial 0: -----
##
## Decision tree:
##
## total_day_minutes > 0.9500225:
## ...voice_mail_plan = yes:
```

```

## :      :...state in {AK,AL,AR,AZ,CA,CO,CT,DC,DE,FL,GA,HI,IA,ID,IL,IN,KS,KY,LA,MA,
## :      :      :      MD,ME,MI,MN,MO,MS,MT,NC,ND,NE,NH,NJ,NM,NY,OH,OK,OR,PA,RI,SC,
## :      :      :      SD,TN,TX,UT,VA,VT,WA,WI,WV,WY}: 0 (380)
## :      :      state = NV: 1 (92)
## :      voice_mail_plan = no:
## :      :...total_eve_charge > -0.1871857:
## :      :      :...total_night_minutes > -1.123915: 1 (4262)
## :      :      :      total_night_minutes <= -1.123915:
## :      :      :      :...total_day_minutes <= 1.351725: 0 (37)
## :      :      :      :      total_day_minutes > 1.351725: 1 (350)
## :      :      total_eve_charge <= -0.1871857:
## :      :      :...state in {AL,CA,CO,CT,DC,DE,GA,IA,ID,IL,LA,ME,MI,MN,MO,MS,MT,NC,ND,
## :      :      :      :      :      NE,NM,OH,OR,PA,RI,SC,TN,VT,WI,WV,WY}: 0 (250)
## :      :      :      state in {AK,AR,AZ,FL,HI,IN,KS,KY,MA,MD,NH,NJ,NV,NY,OK,SD,TX,UT,VA,
## :      :      :      :      :      WA}:
## :      :      :      :...total_day_minutes > 1.560808: 1 (1292)
## :      :      :      :      total_day_minutes <= 1.560808:
## :      :      :      :      :...state in {AK,AZ,FL,HI,IN,KS,KY,MA,NH,NJ,NV,OK,SD,VA,
## :      :      :      :      :      :      :      WA}: 0 (177)
## :      :      :      :      :      state in {AR,MD,NY,TX,UT}:
## :      :      :      :      :      :...total_day_minutes <= 1.061972: 0 (56)
## :      :      :      :      :      :      total_day_minutes > 1.061972:
## :      :      :      :      :      :      :...total_day_calls <= 0.3335073: 1 (486)
## :      :      :      :      :      :      :      total_day_calls > 0.3335073:
## :      :      :      :      :      :      :      :...total_night_charge <= 1.227654: 0 (29)
## :      :      :      :      :      :      :      :      total_night_charge > 1.227654: 1 (78)
## :      total_day_minutes <= 0.9500225:
## :      :...number_customer_service_calls > 0.7277807:
## :      :      :...total_day_minutes <= -0.5069725:
## :      :      :      :...state in {AR,CA,CT,DE,FL,ID,KY,LA,MA,ME,MI,MN,MO,MS,MT,NE,NH,NM,NY,
## :      :      :      :      :      :      OH,OR,SD,TN,TX,UT,WI,WV}: 1 (3057)
## :      :      :      :      :      :      state in {AK,AL,AZ,CO,DC,GA,HI,IA,IL,IN,KS,NC,ND,NV,OK,PA,RI,SC,VA,
## :      :      :      :      :      :      :      :      VT,WA,WY}: 0 (31)
## :      :      :      :      :      :      :      state in {MD,NJ}:
## :      :      :      :      :      :      :      :...total_day_minutes <= -1.035442: 1 (256)
## :      :      :      :      :      :      :      :      total_day_minutes > -1.035442: 0 (27)
## :      :      :      :      :      :      :      total_day_minutes > -0.5069725:
## :      :      :      :      :      :      :      :...total_intl_minutes <= -1.606722: 0 (95)
## :      :      :      :      :      :      :      :      total_intl_minutes > -1.606722:
## :      :      :      :      :      :      :      :...account_length <= 62:
## :      :      :      :      :      :      :      :...international_plan = no: 0 (146)
## :      :      :      :      :      :      :      :      :      international_plan = yes: 1 (79)
## :      :      :      :      :      :      :      :      :      :...account_length > 62:
## :      :      :      :      :      :      :      :...state in {DC,GA,IA,KS,LA,MA,MT,TX}: 1 (721)
## :      :      :      :      :      :      :      :      :      state in {AK,AL,AR,AZ,CA,CO,CT,DE,FL,HI,IL,IN,KY,MD,MI,MN,
## :      :      :      :      :      :      :      :      :      :      :      MO,MS,NC,ND,NE,NH,NM,NV,NY,OH,OK,OR,PA,RI,SC,SD,
## :      :      :      :      :      :      :      :      :      :      :      TN,UT,VA,WI,WY}: 0 (376)
## :      :      :      :      :      :      :      :      :      :      state in {ID,ME,NJ,VT,WA,WV}:
## :      :      :      :      :      :      :      :      :      :...total_day_minutes > 0.1268615: 0 (63)
## :      :      :      :      :      :      :      :      :      :      total_day_minutes <= 0.1268615:
## :      :      :      :      :      :      :      :      :      :...total_day_minutes > 0.02643584: 1 (355)
## :      :      :      :      :      :      :      :      :      :      :      total_day_minutes <= 0.02643584:
## :      :      :      :      :      :      :      :      :      :...total_eve_calls <= -0.3883498: 1 (153)
## :      :      :      :      :      :      :      :      :      :      :      :      total_eve_calls > -0.3883498: 0 (54)

```

```

## number_customer_service_calls <= 0.7277807:
## :...international_plan = yes:
## :...state in {FL,MA}: 1 (0)
## : state in {CO,ID,LA,MI,MT,NE,NM,NV,NY,PA,RI,SC,SD,TX,UT,VA,VT,WI,
## : : WY}: 0 (542)
## : state in {AK,AL,AR,AZ,CA,CT,DC,DE,GA,HI,IA,IL,IN,KS,KY,MD,ME,MN,MO,
## : : MS,NC,ND,NH,NJ,OH,OK,OR,TN,WA,WV}:
## : :...total_intl_calls <= -0.9095257: 1 (1767)
## : total_intl_calls > -0.9095257:
## : :...total_intl_minutes > 0.8883637: 1 (1578)
## : total_intl_minutes <= 0.8883637:
## : :...state in {AK,AL,AR,AZ,CA,CT,DC,DE,GA,HI,IA,IL,IN,KS,KY,
## : : MD,MN,MS,NC,ND,NH,NJ,OH,OK,OR,TN,WA,
## : : WV}: 0 (478)
## : state in {ME,MO}:
## : :...area_code = area_code_510: 1 (0)
## : area_code = area_code_408: 0 (18)
## : area_code = area_code_415:
## : :...total_day_minutes <= -0.908675: 0 (10)
## : total_day_minutes > -0.908675: 1 (163)
## international_plan = no:
## :...state in {AK,CO,CT,FL,GA,HI,IA,KS,MA,MD,ME,MS,NC,NE,NH,NJ,NM,OK,PA,
## : SC,SD,VA,VT,WI}: 0 (7154)
## : state in {AL,AR,AZ,CA,DC,DE,ID,IL,IN,KY,LA,MI,MN,MO,MT,ND,NV,NY,OH,
## : OR,RI,TN,TX,UT,WA,WV,WY}:
## :...total_eve_charge > 0.9780852:
## :...voice_mail_plan = yes: 0 (307)
## : voice_mail_plan = no:
## : :...total_day_calls > 1.170056: 0 (145)
## : total_day_calls <= 1.170056:
## : :...total_night_calls <= -1.314227: 0 (79)
## : total_night_calls > -1.314227:
## : :...total_day_minutes <= -0.3061212:
## : :...account_length > 73: 0 (264)
## : : account_length <= 73:
## : : :...total_day_minutes <= -1.740068: 0 (30)
## : : total_day_minutes > -1.740068: 1 (162)
## : total_day_minutes > -0.3061212:
## : :...state in {AR,CA,DE,MI,MN,MO,ND,NV,NY,RI,TN,
## : : WA,WV}: 1 (1502)
## : state in {AL,AZ,DC,ID,IL,IN,KY,LA,OR,TX,UT,
## : : WY}: 0 (104)
## : state in {MT,OH}:
## : :...total_intl_minutes <= -0.6665444: 0 (35)
## : total_intl_minutes > -0.6665444: 1 (186)
## total_eve_charge <= 0.9780852:
## :...state in {AR,DE,MI,MO,MT,ND,NV,NY,RI,WY}: 0 (2704)
## : state in {AL,AZ,CA,DC,ID,IL,IN,KY,LA,MN,OH,OR,TN,TX,UT,WA,
## : : WV}:
## : :...account_length <= 48: 0 (653)
## : account_length > 48:
## : :...total_day_calls > 1.170056: 0 (524)
## : total_day_calls <= 1.170056:
## : :...account_length > 165:

```

```

##                                     :...state in {DC,ID,LA,TX}: 1 (0)
##                                     :   state in {AL,AZ,CA,IL,KY,OH,TN,UT,
##                                     :   :           WV}: 0 (139)
##                                     :   state in {IN,MN,OR,WA}:
##                                     :   :...account_length > 186: 0 (27)
##                                     :       account_length <= 186:
##                                     :       :...total_eve_calls <= -0.6947184: 0 (26)
##                                     :       :       total_eve_calls > -0.6947184: 1 (361)
##                                     account_length <= 165:
##                                     :...total_night_charge <= -1.856685:
##                                     :   :...state in {AL,AZ,CA,DC,ID,IL,KY,MN,OR,
##                                     :   :   :           TN,TX,WA}: 1 (0)
##                                     :   :   state in {IN,LA,OH}: 0 (37)
##                                     :   :   state in {UT,WV}:
##                                     :   :   :...voice_mail_plan = no: 1 (168)
##                                     :   :       voice_mail_plan = yes: 0 (10)
##                                     total_night_charge > -1.856685:
##                                     :...state in {MN,OR,WV}: 0 (773)
##                                     :   state in {AL,AZ,CA,DC,ID,IL,IN,KY,LA,
##                                     :   :           OH,TN,TX,UT,WA}:
##                                     :   :...account_length > 149: 0 (180)
##                                     :       account_length <= 149: [S1]
##
## SubTree [S1]
##
## total_night_charge <= -1.396776: 0 (191)
## total_night_charge > -1.396776:
## :...total_night_minutes <= -1.355551: 1 (76)
##     total_night_minutes > -1.355551:
##     :...account_length > 148:
##     :   :...state = CA: 0 (10)
##     :   :   state in {AL,AZ,DC,ID,IL,IN,KY,LA,OH,TN,TX,UT,WA}: 1 (85)
##     :   account_length <= 148:
##     :   :...total_day_calls > 0.5949287:
##     :   :   :...state = UT: 1 (0)
##     :   :   :   state in {AL,DC,IL,IN,KY,OH,TX,WA}: 0 (215)
##     :   :   :   state in {AZ,CA,ID,LA,TN}:
##     :   :   :   :...voice_mail_plan = yes: 0 (22)
##     :   :   :   :   voice_mail_plan = no:
##     :   :   :   :   :...total_night_calls <= 0.6061517: 1 (415)
##     :   :   :   :   :       total_night_calls > 0.6061517:
##     :   :   :   :   :       :...state in {AZ,CA,ID,TN}: 1 (73)
##     :   :   :   :   :       :   state = LA: 0 (14)
##     :   :   total_day_calls <= 0.5949287:
##     :   :   :...number_customer_service_calls > 0.07807057:
##     :   :   :   :...state in {DC,UT}: 1 (0)
##     :   :   :   :   state in {AZ,CA,ID,IN,KY,LA,TN,WA}: 0 (214)
##     :   :   :   :   state in {AL,IL,OH,TX}:
##     :   :   :   :   :...total_night_calls <= -0.8710629: 0 (31)
##     :   :   :   :   :       total_night_calls > -0.8710629:
##     :   :   :   :   :       :...total_eve_calls <= -0.2862269: 0 (27)
##     :   :   :   :   :       :       total_eve_calls > -0.2862269:
##     :   :   :   :   :       :       :...voice_mail_plan = no: 1 (409)
##     :   :   :   :   :       :       :       voice_mail_plan = yes: 0 (4)

```

```

##          number_customer_service_calls <= 0.07807057:
##          :...state in {AL,AZ,CA,DC,ID,IL,LA,OH,TN,TX}: 0 (1447)
##          state in {IN,KY,UT,WA}:
##          :...total_eve_minutes > -0.3669789: 0 (243)
##          total_eve_minutes <= -0.3669789:
##          :...total_night_calls > 1.000076: 0 (70)
##          total_night_calls <= 1.000076:
##          :...area_code = area_code_510: 0 (65)
##          area_code in {area_code_408,area_code_415}:
##          :...total_night_charge <= -0.4853985: 0 (37)
##          total_night_charge > -0.4853985:
##          :...total_eve_calls <= -0.5415341: 1 (147)
##          total_eve_calls > -0.5415341:
##          :...total_night_calls > -0.08321509: 0 (28)
##          total_night_calls <= -0.08321509:
##          :...total_intl_charge <= -0.2681369: 0 (10)
##          total_intl_charge > -0.2681369: 1 (139)
##
## *** boosting reduced to 1 trial since last classifier is very accurate
##
## *** boosting abandoned (too few classifiers)
##
## Evaluation on training data (37000 cases):
##
##      Decision Tree
##      -----
##      Size      Errors
##
##      78      0( 0.0%)  <<
##
##      (a)   (b)   <-classified as
##      ----  ----
##      18588      (a): class 0
##      18412      (b): class 1
##
##
## Attribute usage:
##
## 100.00% total_day_minutes
## 86.57% state
## 79.76% number_customer_service_calls
## 65.74% international_plan
## 52.45% total_eve_charge
## 30.86% voice_mail_plan
## 25.18% total_day_calls
## 25.06% account_length
## 22.78% total_night_minutes
## 14.18% total_night_charge
## 12.19% total_intl_minutes
## 10.85% total_intl_calls
## 10.35% total_night_calls
## 3.67% total_eve_calls

```

```
##      2.00% total_eve_minutes
##      1.67% area_code
##      0.40% total_intl_charge
##
##
## Time: 0.6 secs
```

```
#predict
predi2 <- predict(modelv2, test)
CrossTable(test$churn, predi2)
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |          N / Row Total |
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table:  20000
##
##
##      | predi2
## test$churn |      0 |      1 | Row Total |
## -----|-----|-----|-----|
##          0 |    8981 |    941 |    9922 |
##          |  695.070 | 1518.176 |          |
##          |    0.905 |    0.095 |    0.496 |
##          |    0.655 |    0.150 |          |
##          |    0.449 |    0.047 |          |
## -----|-----|-----|-----|
##          1 |    4738 |    5340 |    10078 |
##          |  684.311 | 1494.676 |          |
##          |    0.470 |    0.530 |    0.504 |
##          |    0.345 |    0.850 |          |
##          |    0.237 |    0.267 |          |
## -----|-----|-----|-----|
## Column Total |    13719 |    6281 |    20000 |
##          |    0.686 |    0.314 |          |
## -----|-----|-----|-----|
##
##
```

```
#Using randomForest

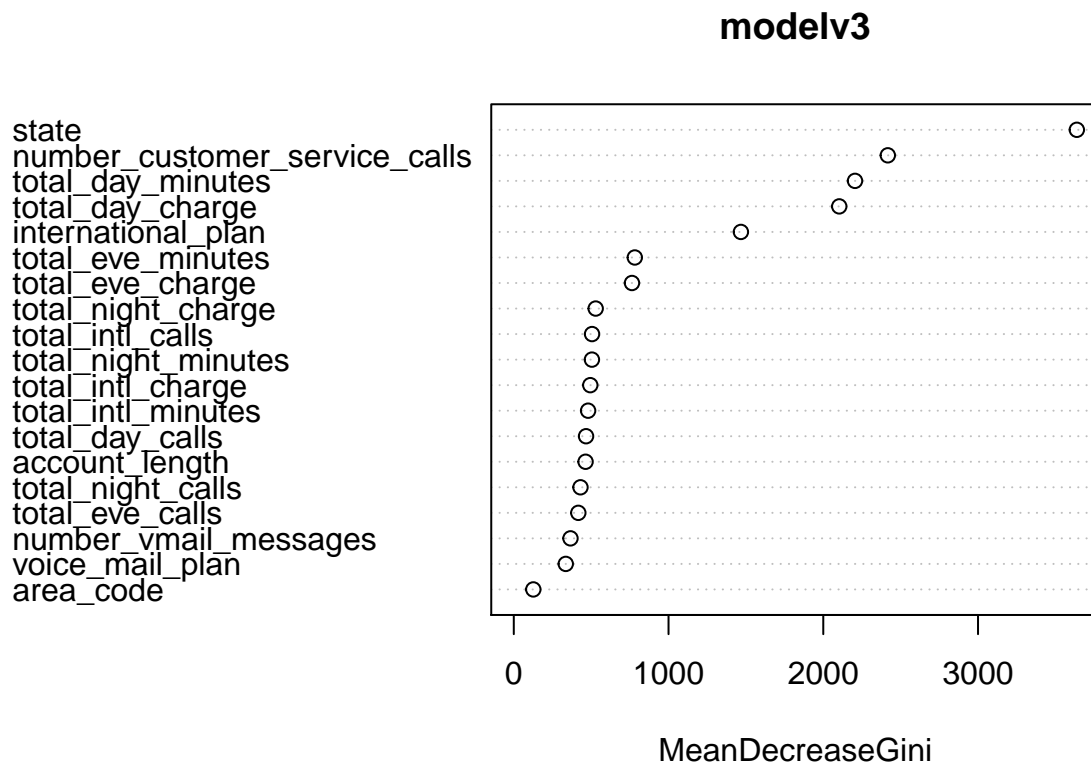
modelv3 <- randomForest(churn ~., data = train)
modelv3
```

```
##
```



```
## Call:
## randomForest(formula = churn ~ ., data = train)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 4
##
##           OOB estimate of  error rate: 0%
## Confusion matrix:
##           0      1 class.error
## 0 18588      0      0
## 1      0 18412      0
```

```
varImpPlot(modelv3)
```



```
#predict
target <- test$churn
test1 <- test[, -20]

predict3 <- predict(modelv3, test)

confusionMatrix(table(predict3, target))
```

```
## Confusion Matrix and Statistics
```

```
##
##      target
## predict3    0    1
##           0 9791 4413
##           1  131 5665
##
##           Accuracy : 0.7728
##           95% CI : (0.7669, 0.7786)
##      No Information Rate : 0.5039
##      P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.5471
##
## Mcnemar's Test P-Value : < 2.2e-16
##
##           Sensitivity : 0.9868
##           Specificity : 0.5621
##           Pos Pred Value : 0.6893
##           Neg Pred Value : 0.9774
##           Prevalence : 0.4961
##           Detection Rate : 0.4895
##      Detection Prevalence : 0.7102
##           Balanced Accuracy : 0.7745
##
##           'Positive' Class : 0
##
```

```
#Model 2 removing features
train$number_vmail_messages <- NULL
train$area_code <- NULL
train$voice_mail_plan <- NULL
train$total_night_calls <- NULL

test1$number_vmail_messages <- NULL
test1$area_code <- NULL
test1$voice_mail_plan <- NULL
test1$total_night_calls <- NULL

View(train)
#Model creation wito C5.0

modelc1 <- C5.0(train[,16], train$churn)

#predict
predcc <- predict(modelc1, test)
CrossTable(test$churn, predcc)
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
```

```
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
##
##
## Total Observations in Table: 20000
##
##
##           | predcc
## test$churn |      0 |      1 | Row Total |
## -----|-----|-----|-----|
##           0 |   8797 |   1125 |   9922 |
##           | 552.984 | 1233.429 |         |
##           |   0.887 |   0.113 |   0.496 |
##           |   0.637 |   0.182 |         |
##           |   0.440 |   0.056 |         |
## -----|-----|-----|-----|
##           1 |   5012 |   5066 |  10078 |
##           | 544.424 | 1214.336 |         |
##           |   0.497 |   0.503 |   0.504 |
##           |   0.363 |   0.818 |         |
##           |   0.251 |   0.253 |         |
## -----|-----|-----|-----|
## Column Total |   13809 |    6191 |   20000 |
##           |   0.690 |   0.310 |         |
## -----|-----|-----|-----|
##
##
```

```
#Model with Random Forest
modelv1_rf <- randomForest(churn ~., data = train)
```

```
#predict
predrf <- predict(modelv1_rf, test1)

confusionMatrix(table(predrf, target))
```

```
## Confusion Matrix and Statistics
##
##           target
## predrf      0      1
##      0 9710 4847
##      1  212 5231
##
##              Accuracy : 0.747
##              95% CI : (0.741, 0.7531)
##      No Information Rate : 0.5039
##      P-Value [Acc > NIR] : < 2.2e-16
##
##              Kappa : 0.4959
##
##      McNemar's Test P-Value : < 2.2e-16
```

```
##
##      Sensitivity : 0.9786
##      Specificity : 0.5191
##      Pos Pred Value : 0.6670
##      Neg Pred Value : 0.9611
##      Prevalence : 0.4961
##      Detection Rate : 0.4855
##      Detection Prevalence : 0.7278
##      Balanced Accuracy : 0.7488
##
##      'Positive' Class : 0
##
```

#Model modelv3 has presented the best performance with 78% of accuracy