Regressão Logística

# Carregar Dados

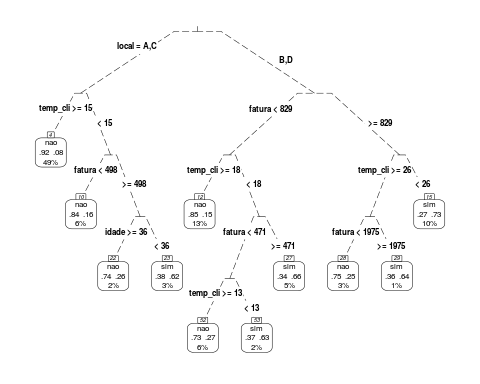
TEBA <- read\_excel("TEBA.xlsx")

# Separar em treino e teste

set.seed(1234)  
index=sample(1:2000,1200)  
  
teba.learn=TEBA[index,]  
teba.test=TEBA[-index,]

# Criar arvore

library(rpart)  
library(rpart.plot)  
  
set.seed(93)  
ad1=rpart(data=teba.learn, cancel~idade+linhas+temp\_cli+renda+fatura+temp\_rsd+local+tvcabo+debaut, method = "class")  
prp(ad1, type=3, extra=104,nn=T, fallen.leaves = F, branch.lty = 5)

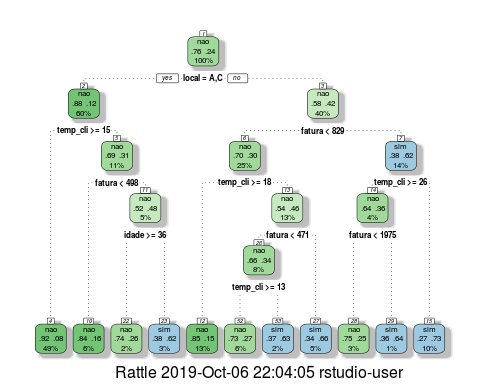


# Visualizacao alternativa

library(RColorBrewer)  
library(rattle)

## Rattle: A free graphical interface for data science with R.  
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

fancyRpartPlot(ad1)



# Saida crua da arvare

ad1

## n= 1200   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1200 288 nao (0.76000000 0.24000000)   
## 2) local=A,C 726 89 nao (0.87741047 0.12258953)   
## 4) temp\_cli>=14.5 590 47 nao (0.92033898 0.07966102) \*  
## 5) temp\_cli< 14.5 136 42 nao (0.69117647 0.30882353)   
## 10) fatura< 497.5 74 12 nao (0.83783784 0.16216216) \*  
## 11) fatura>=497.5 62 30 nao (0.51612903 0.48387097)   
## 22) idade>=35.5 23 6 nao (0.73913043 0.26086957) \*  
## 23) idade< 35.5 39 15 sim (0.38461538 0.61538462) \*  
## 3) local=B,D 474 199 nao (0.58016878 0.41983122)   
## 6) fatura< 829 302 92 nao (0.69536424 0.30463576)   
## 12) temp\_cli>=17.5 151 22 nao (0.85430464 0.14569536) \*  
## 13) temp\_cli< 17.5 151 70 nao (0.53642384 0.46357616)   
## 26) fatura< 471 93 32 nao (0.65591398 0.34408602)   
## 52) temp\_cli>=12.5 74 20 nao (0.72972973 0.27027027) \*  
## 53) temp\_cli< 12.5 19 7 sim (0.36842105 0.63157895) \*  
## 27) fatura>=471 58 20 sim (0.34482759 0.65517241) \*  
## 7) fatura>=829 172 65 sim (0.37790698 0.62209302)   
## 14) temp\_cli>=25.5 50 18 nao (0.64000000 0.36000000)   
## 28) fatura< 1975 36 9 nao (0.75000000 0.25000000) \*  
## 29) fatura>=1975 14 5 sim (0.35714286 0.64285714) \*  
## 15) temp\_cli< 25.5 122 33 sim (0.27049180 0.72950820) \*

# Classificar individuos da porcao de treino

teba.learn$pred.lrn.class=predict(ad1, newdata = teba.learn, type = "class")  
table(teba.learn$cancel,teba.learn$pred.lrn.class)

##   
## nao sim  
## nao 832 80  
## sim 116 172

# Analisar a necessidade de podar

É necessario podar quando há decrescimo no valor de xerror

printcp(ad1)

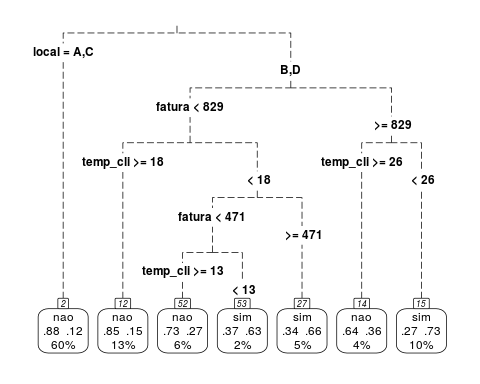
##   
## Classification tree:  
## rpart(formula = cancel ~ idade + linhas + temp\_cli + renda +   
## fatura + temp\_rsd + local + tvcabo + debaut, data = teba.learn,   
## method = "class")  
##   
## Variables actually used in tree construction:  
## [1] fatura idade local temp\_cli  
##   
## Root node error: 288/1200 = 0.24  
##   
## n= 1200   
##   
## CP nsplit rel error xerror xstd  
## 1 0.072917 0 1.00000 1.00000 0.051370  
## 2 0.048611 2 0.85417 0.89931 0.049484  
## 3 0.031250 3 0.80556 0.89236 0.049345  
## 4 0.017361 5 0.74306 0.89583 0.049414  
## 5 0.013889 6 0.72569 0.88542 0.049204  
## 6 0.010417 7 0.71181 0.89236 0.049345  
## 7 0.010000 10 0.68056 0.91319 0.049758

# Podar a arvore

ad2=prune(ad1,cp=0.015)  
ad2

## n= 1200   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1200 288 nao (0.7600000 0.2400000)   
## 2) local=A,C 726 89 nao (0.8774105 0.1225895) \*  
## 3) local=B,D 474 199 nao (0.5801688 0.4198312)   
## 6) fatura< 829 302 92 nao (0.6953642 0.3046358)   
## 12) temp\_cli>=17.5 151 22 nao (0.8543046 0.1456954) \*  
## 13) temp\_cli< 17.5 151 70 nao (0.5364238 0.4635762)   
## 26) fatura< 471 93 32 nao (0.6559140 0.3440860)   
## 52) temp\_cli>=12.5 74 20 nao (0.7297297 0.2702703) \*  
## 53) temp\_cli< 12.5 19 7 sim (0.3684211 0.6315789) \*  
## 27) fatura>=471 58 20 sim (0.3448276 0.6551724) \*  
## 7) fatura>=829 172 65 sim (0.3779070 0.6220930)   
## 14) temp\_cli>=25.5 50 18 nao (0.6400000 0.3600000) \*  
## 15) temp\_cli< 25.5 122 33 sim (0.2704918 0.7295082) \*

prp(ad2, type=3, extra=104,nn=T, fallen.leaves = T, branch.lty = 5)



# Analisar resultado na amostra teste

ad2=prune(ad1,cp=0.015)  
ad2

## n= 1200   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1200 288 nao (0.7600000 0.2400000)   
## 2) local=A,C 726 89 nao (0.8774105 0.1225895) \*  
## 3) local=B,D 474 199 nao (0.5801688 0.4198312)   
## 6) fatura< 829 302 92 nao (0.6953642 0.3046358)   
## 12) temp\_cli>=17.5 151 22 nao (0.8543046 0.1456954) \*  
## 13) temp\_cli< 17.5 151 70 nao (0.5364238 0.4635762)   
## 26) fatura< 471 93 32 nao (0.6559140 0.3440860)   
## 52) temp\_cli>=12.5 74 20 nao (0.7297297 0.2702703) \*  
## 53) temp\_cli< 12.5 19 7 sim (0.3684211 0.6315789) \*  
## 27) fatura>=471 58 20 sim (0.3448276 0.6551724) \*  
## 7) fatura>=829 172 65 sim (0.3779070 0.6220930)   
## 14) temp\_cli>=25.5 50 18 nao (0.6400000 0.3600000) \*  
## 15) temp\_cli< 25.5 122 33 sim (0.2704918 0.7295082) \*

prp(ad2, type=3, extra=104,nn=T, fallen.leaves = T, branch.lty = 5)

