FDS homework 3

Sadiya Amreen- 2079690

2022-10-20

Problem 4

**Problem 4 (25 points)**

**In this problem you will identify the most important independent variables used in a classification model. Use the Bank\_Modified.csv data. As a preprocessing step, remove the ID column and make sure to convert the target variable, approval, from a string to a factor.**

#Setting directory

setwd("C:/Sadiya Studies/Data Science/DS441-Fundamts DS/homework")

#Reading Data from the .CSV file

Data04 <- read.csv("Bank\_Modified.csv")

head(Data04)

## X cont1 cont2 cont3 bool1 bool2 cont4 bool3 cont5 cont6 approval credit.score  
## 1 1 30.83 0.000 1.25 t t 1 f 202 0 + 664.60  
## 2 2 58.67 4.460 3.04 t t 6 f 43 560 + 693.88  
## 3 3 24.50 0.500 1.50 t f 0 f 280 824 + 621.82  
## 4 4 27.83 1.540 3.75 t t 5 t 100 3 + 653.97  
## 5 5 20.17 5.625 1.71 t f 0 f 120 0 + 670.26  
## 6 6 32.08 4.000 2.50 t f 0 t 360 0 + 672.16  
## ages  
## 1 58  
## 2 54  
## 3 62  
## 4 51  
## 5 58  
## 6 37

#### removing blank columns  
Data04 <- na.omit(Data04)

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Data4 <- Data4 %>% select(-c("X"))  
head(Data4)

## cont1 cont2 cont3 bool1 bool2 cont4 bool3 cont5 cont6 approval credit.score  
## 1 30.83 0.000 1.25 t t 1 f 202 0 + 664.60  
## 2 58.67 4.460 3.04 t t 6 f 43 560 + 693.88  
## 3 24.50 0.500 1.50 t f 0 f 280 824 + 621.82  
## 4 27.83 1.540 3.75 t t 5 t 100 3 + 653.97  
## 5 20.17 5.625 1.71 t f 0 f 120 0 + 670.26  
## 6 32.08 4.000 2.50 t f 0 t 360 0 + 672.16  
## ages  
## 1 58  
## 2 54  
## 3 62  
## 4 51  
## 5 58  
## 6 37

### converting the variable from a string to a factor  
Data04$approval <- as.factor(Data04$approval)

##### 

**a. Build your initial decision tree model with minsplit=10 and maxdepth=20**

library(rattle)

## Loading required package: tibble

## Loading required package: bitops

## Rattle: A free graphical interface for data science with R.  
## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

library(tidyverse)

## ── Attaching packages  
## ───────────────────────────────────────  
## tidyverse 1.3.2 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.5  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1  
## ✔ readr 2.1.3 ✔ forcats 0.5.2  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

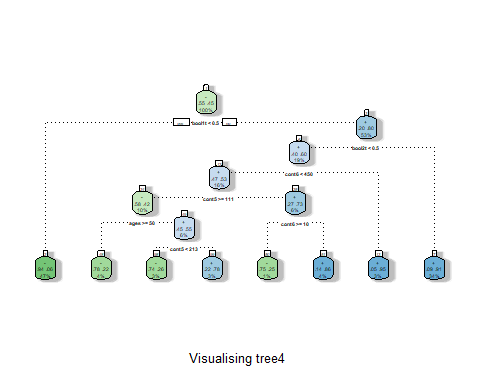
library(ggplot2)  
library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(lattice)  
library(rpart)  
train\_control = trainControl(method = "cv", number = 10)  
hypers4 = rpart.control(minsplit = 10, maxdepth = 20)  
tree4 <- train(approval ~., data = Data04, control = hypers4, trControl = train\_control, method = "rpart1SE")  
tree4

## CART   
##   
## 666 samples  
## 11 predictor  
## 2 classes: '-', '+'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 600, 599, 600, 600, 599, 599, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.8738354 0.7468813

fancyRpartPlot(tree4$finalModel, caption= "Visualising tree4")



**b. Run variable importance analysis on the model and print the result.**

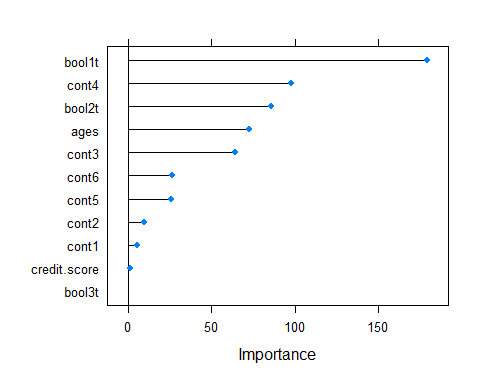
varimpanalysis <- varImp(tree4, scale= FALSE)  
print(varimpanalysis)

## rpart1SE variable importance  
##   
## Overall  
## bool1t 179.282  
## cont4 97.700  
## bool2t 85.622  
## ages 72.800  
## cont3 64.343  
## cont6 26.828  
## cont5 25.878  
## cont2 9.620  
## cont1 5.646  
## credit.score 1.504  
## bool3t 0.000

**c. Generate a plot to visualize the variables by importance.**

#### 

plot(varimpanalysis)



**d. Rebuild your model with the top six variables only, based on the variable relevance analysis. Did this change have an effect on the accuracy?**

#After reducing the number of variables in the model, the accuracy of the model did not change full data accuracy : 0.8738354 , sicx variables data Accuracy: 0.8679783

rebuild<- Data04 %>% select(c("approval", "bool1", "cont4","bool2", "ages", "cont3", "cont6"))  
rebuild

## approval bool1 cont4 bool2 ages cont3 cont6  
## 1 + t 1 t 58 1.250 0  
## 2 + t 6 t 54 3.040 560  
## 3 + t 0 f 62 1.500 824  
## 4 + t 5 t 51 3.750 3  
## 5 + t 0 f 58 1.710 0  
## 6 + t 0 f 37 2.500 0  
## 7 + t 0 f 47 6.500 31285  
## 8 + t 0 f 67 0.040 1349  
## 9 + t 0 f 61 3.960 314  
## 10 + t 0 f 62 3.165 1442  
## 11 + f 0 f 41 2.165 0  
## 12 + t 0 f 51 4.335 200  
## 13 + t 0 f 51 1.000 0  
## 14 + f 0 f 32 0.040 2690  
## 15 + t 7 t 51 5.000 0  
## 16 + t 10 t 59 0.250 0  
## 17 + t 3 t 49 0.960 0  
## 18 + t 10 t 49 3.170 245  
## 19 + t 0 f 46 0.665 0  
## 20 + t 7 t 55 0.750 0  
## 21 + t 17 t 58 2.500 1208  
## 22 + t 0 f 43 0.835 0  
## 23 + t 6 t 43 7.875 1260  
## 24 + t 1 t 47 3.085 11  
## 25 + t 3 t 43 0.500 0  
## 26 + t 2 t 63 1.500 0  
## 27 + t 9 t 51 5.165 0  
## 28 + t 17 t 45 15.000 0  
## 29 + t 3 t 40 7.000 0  
## 30 + t 6 t 49 5.000 10000  
## 31 + t 5 t 38 5.040 0  
## 32 + t 8 t 52 7.960 0  
## 33 + t 15 t 54 7.585 5000  
## 34 + t 0 f 42 5.000 4000  
## 35 + t 5 t 45 0.415 560  
## 36 + t 11 t 62 2.000 35  
## 37 + t 12 t 50 1.835 713  
## 38 + t 2 t 57 0.500 551  
## 39 + t 2 t 43 0.250 500  
## 40 + t 11 t 59 14.415 300  
## 41 + t 12 t 39 4.500 221  
## 42 + t 11 t 36 5.335 2283  
## 43 + t 1 t 61 0.750 100  
## 44 + t 6 t 48 8.625 0  
## 45 + t 40 t 58 28.500 15  
## 46 + t 11 t 61 2.625 284  
## 47 + t 23 t 57 0.125 1236  
## 48 + t 3 t 50 6.040 300  
## 49 + f 0 f 48 3.500 0  
## 50 + f 0 f 55 0.165 0  
## 51 + t 0 f 35 0.875 0  
## 52 + t 0 f 45 1.750 0  
## 53 + t 0 f 49 0.040 5800  
## 54 + t 0 f 41 0.000 200  
## 55 + t 0 f 51 7.415 0  
## 56 + t 0 f 43 0.835 300  
## 57 + t 0 f 44 0.085 0  
## 58 + t 0 f 35 5.000 0  
## 59 + f 0 f 54 5.750 0  
## 60 + t 11 t 57 6.000 0  
## 61 + t 4 t 52 1.250 0  
## 62 + t 9 t 54 3.000 730  
## 63 + t 2 t 35 1.500 400  
## 64 + t 1 t 35 1.585 0  
## 65 + t 1 t 60 4.290 0  
## 66 + t 1 t 45 1.540 50000  
## 67 + t 11 t 54 2.000 456  
## 68 + t 3 t 42 0.250 15108  
## 69 + t 7 t 32 1.460 2954  
## 70 + t 1 t 46 1.625 500  
## 71 - t 0 f 52 1.585 0  
## 73 - t 0 f 58 13.500 0  
## 74 - t 0 f 56 10.750 0  
## 75 - f 0 f 53 1.625 2  
## 76 - t 1 t 52 0.375 0  
## 77 - t 0 f 55 0.125 0  
## 78 - t 0 f 44 0.585 0  
## 79 - t 1 t 51 2.500 20  
## 80 - t 0 f 50 0.250 0  
## 81 - t 1 t 52 0.000 0  
## 82 - t 0 f 52 2.000 0  
## 83 - t 0 f 54 0.250 0  
## 85 - t 0 f 54 0.455 0  
## 86 - t 0 f 54 5.000 0  
## 88 - t 0 f 36 4.000 0  
## 89 - t 0 f 51 1.000 0  
## 90 - t 0 f 48 0.000 27  
## 91 - t 0 f 40 5.000 0  
## 92 - t 0 f 39 0.500 0  
## 94 - t 0 f 55 9.460 100  
## 95 - t 0 f 36 1.500 225  
## 96 - t 0 f 50 0.500 0  
## 97 - t 0 f 84 0.125 1  
## 99 - t 0 f 61 3.000 0  
## 100 - t 2 t 71 1.000 500  
## 101 - t 0 f 44 0.250 400  
## 102 - t 0 f 31 4.000 0  
## 103 - t 2 t 40 0.375 38  
## 104 - t 2 t 56 2.250 5  
## 105 - t 2 t 80 5.750 0  
## 106 - t 20 t 49 0.000 130  
## 107 - t 0 f 58 0.500 0  
## 108 - t 0 f 55 4.500 0  
## 109 - t 0 f 45 10.000 0  
## 110 - t 5 t 58 0.795 5  
## 111 - t 3 t 53 3.500 0  
## 112 - t 3 t 56 0.500 147  
## 113 - t 0 f 52 0.875 0  
## 114 - t 3 t 64 1.000 0  
## 115 - t 3 t 34 1.375 210  
## 116 - t 2 t 55 1.290 0  
## 117 - t 7 t 48 11.500 5  
## 118 + t 15 t 69 6.290 11202  
## 119 + t 6 t 56 14.000 1332  
## 120 + t 1 t 51 0.335 50  
## 121 + t 1 t 29 0.040 300  
## 122 + t 67 t 61 1.210 258  
## 123 + t 12 t 51 1.500 567  
## 124 + t 3 t 43 7.375 0  
## 125 + t 5 t 51 8.500 0  
## 126 + t 6 t 59 7.500 1000  
## 127 + t 12 t 52 2.500 2510  
## 128 + t 7 t 41 2.500 809  
## 129 + t 2 t 36 3.250 610  
## 130 + t 0 f 45 0.835 0  
## 131 + t 1 t 25 13.000 0  
## 132 + t 1 t 52 2.250 150  
## 133 + t 6 t 47 6.500 51100  
## 134 + t 6 t 44 2.500 367  
## 135 + t 12 t 42 5.500 1000  
## 136 + t 0 f 43 6.000 0  
## 137 + t 3 t 44 0.500 1000  
## 138 + t 6 t 58 4.250 0  
## 139 + t 6 t 35 1.625 600  
## 140 + t 2 t 30 5.000 5000  
## 141 + t 9 t 55 0.625 0  
## 142 + t 15 t 34 0.000 247  
## 143 + t 8 t 63 1.750 375  
## 144 + t 1 t 44 2.000 278  
## 145 + t 9 t 41 5.085 827  
## 146 + t 6 t 51 2.750 2072  
## 147 + t 3 t 57 2.375 582  
## 148 + t 14 t 50 8.000 2300  
## 149 + t 7 t 33 4.000 3065  
## 150 + t 14 t 29 5.500 2200  
## 151 + t 11 t 53 0.415 6  
## 152 + t 14 t 74 4.000 1602  
## 153 + t 12 t 47 4.250 0  
## 154 + t 11 t 51 1.085 2184  
## 155 + t 3 t 74 5.500 0  
## 156 + t 11 t 42 0.000 0  
## 157 + t 1 t 45 2.540 0  
## 158 + t 14 t 41 0.000 3376  
## 159 + t 2 t 52 4.165 0  
## 160 + t 1 t 57 0.040 2000  
## 161 + t 4 t 36 1.000 7544  
## 162 + t 2 t 58 1.750 15  
## 163 + t 1 t 43 1.665 20  
## 164 + t 0 f 58 0.040 0  
## 165 + t 0 f 50 11.000 10561  
## 166 + t 0 f 50 1.750 837  
## 167 + t 0 f 46 1.000 400  
## 168 + t 0 f 46 0.040 11177  
## 169 + t 0 f 50 9.000 639  
## 170 + f 0 f 49 1.500 0  
## 171 + t 0 f 58 0.250 0  
## 172 + t 0 f 40 15.000 0  
## 173 + t 0 f 39 8.000 2028  
## 174 + t 0 f 43 8.500 0  
## 175 + t 0 f 43 1.335 0  
## 176 + t 0 f 49 0.375 1065  
## 177 + t 0 f 41 3.000 0  
## 178 + t 0 f 46 1.415 150  
## 179 + t 4 t 55 1.210 540  
## 180 + t 14 t 56 1.960 158  
## 181 + t 20 t 59 15.000 15000  
## 182 + t 4 t 62 0.500 0  
## 183 + t 3 t 57 0.165 6  
## 184 + t 7 t 37 5.500 3000  
## 185 + t 7 t 52 2.585 3257  
## 186 + t 9 t 48 12.500 1655  
## 187 + t 1 t 37 3.500 500  
## 188 + t 7 t 66 5.000 3065  
## 189 + t 5 t 37 2.500 1430  
## 190 + t 2 t 48 1.625 0  
## 191 + t 0 f 47 3.000 0  
## 192 + t 0 f 47 5.125 0  
## 193 + t 0 f 64 2.500 600  
## 194 + t 6 t 45 3.085 0  
## 195 + t 7 t 47 8.500 0  
## 196 + t 8 t 46 1.500 7  
## 197 + t 3 t 55 3.165 0  
## 198 + t 12 t 44 15.500 790  
## 199 + t 3 t 54 2.000 560  
## 200 + t 9 t 53 0.040 396  
## 201 + t 1 t 43 1.250 678  
## 202 + t 0 f 46 2.250 300  
## 204 + t 2 t 55 0.710 0  
## 205 + t 1 t 43 0.085 1187  
## 206 + t 8 t 39 14.000 6590  
## 208 + t 6 t 62 5.665 168  
## 209 + t 7 t 30 4.500 1270  
## 210 + t 16 t 27 6.500 1210  
## 211 + t 14 t 47 10.000 0  
## 212 + t 0 f 29 5.500 0  
## 213 + t 15 t 43 18.000 1000  
## 214 + t 9 t 55 3.500 742  
## 215 + t 1 t 47 5.250 0  
## 216 + t 0 f 49 3.500 0  
## 217 + t 8 t 50 7.000 0  
## 218 + t 5 t 38 5.000 8851  
## 219 + t 5 t 47 8.665 0  
## 220 + t 4 t 57 1.000 500  
## 221 + t 3 t 46 2.290 0  
## 222 + t 7 t 61 20.000 0  
## 223 + t 0 f 41 1.375 0  
## 224 + t 0 f 54 0.085 0  
## 225 + t 0 f 50 0.125 0  
## 226 + t 0 f 43 0.250 0  
## 227 + t 0 f 50 0.500 0  
## 228 + f 0 f 41 2.460 0  
## 229 + t 2 t 50 2.000 0  
## 230 + t 0 f 58 0.665 0  
## 231 + t 7 t 34 0.625 7059  
## 232 + t 2 t 29 13.875 1704  
## 233 + t 4 t 37 4.500 857  
## 234 + t 0 f 25 5.750 500  
## 235 + t 13 t 31 10.000 6700  
## 236 + t 5 t 33 2.085 2503  
## 237 + t 0 f 22 0.000 0  
## 238 + t 1 t 26 1.415 9800  
## 239 + t 0 f 33 4.580 0  
## 240 + t 6 t 36 2.500 196  
## 241 + t 0 f 40 2.500 0  
## 242 + t 3 t 32 1.750 14  
## 243 + t 0 f 34 11.000 0  
## 245 + t 2 t 27 1.500 300  
## 246 + t 1 t 33 2.040 18027  
## 247 + t 1 t 29 14.000 2000  
## 248 + t 11 t 35 0.290 99  
## 249 + t 2 t 24 4.750 444  
## 250 + t 6 t 20 0.290 0  
## 251 + t 11 t 26 20.000 1200  
## 252 + t 6 t 31 5.000 0  
## 253 + t 11 t 29 1.000 3000  
## 254 + t 1 t 25 0.460 0  
## 256 - f 0 f 33 1.085 13  
## 257 - f 0 f 25 2.000 0  
## 258 - f 0 f 28 0.500 0  
## 259 - f 0 f 33 0.040 1000  
## 260 - f 0 f 34 0.165 0  
## 261 - f 0 f 27 5.750 0  
## 262 - f 0 f 31 0.000 0  
## 263 - f 0 f 35 0.335 120  
## 264 - f 0 f 24 0.000 32  
## 265 - f 0 f 38 0.000 0  
## 266 - f 0 f 26 0.040 722  
## 267 - f 0 f 38 0.000 0  
## 268 - f 0 f 30 1.250 0  
## 269 + t 0 f 24 0.125 0  
## 270 + f 0 f 24 0.210 40  
## 272 - f 0 f 38 1.250 0  
## 273 - f 0 f 32 0.040 0  
## 274 - f 0 f 29 0.125 0  
## 275 - f 0 f 27 2.250 0  
## 276 - f 0 f 30 0.540 0  
## 277 - f 0 f 26 0.290 484  
## 278 - f 0 f 30 0.165 0  
## 280 - t 0 f 19 0.085 0  
## 281 - f 0 f 32 0.250 204  
## 282 - f 0 f 34 0.125 1  
## 283 - f 0 f 38 0.250 0  
## 284 - f 1 t 29 0.250 98  
## 285 - f 2 t 28 0.125 5552  
## 286 - f 1 t 21 0.165 1  
## 288 - f 1 t 27 0.290 2803  
## 289 - f 1 t 24 0.165 1  
## 290 - f 0 f 32 0.000 0  
## 291 - f 0 f 25 0.500 0  
## 292 - f 1 t 33 1.000 1  
## 293 - f 1 t 33 0.165 444  
## 294 - f 2 t 27 0.125 1  
## 295 - f 1 t 19 0.000 126  
## 296 - f 2 t 32 3.040 4  
## 297 - f 1 t 36 4.000 6  
## 298 - f 2 t 30 1.750 0  
## 299 - f 1 t 25 0.665 21  
## 300 - f 2 t 27 3.335 173  
## 301 - f 1 t 36 6.500 10  
## 302 - f 0 f 29 0.250 0  
## 303 - f 0 f 30 0.500 0  
## 304 - f 0 f 32 0.085 0  
## 305 - f 0 f 29 1.500 1  
## 306 - t 0 f 31 0.000 0  
## 307 - f 0 f 31 4.500 25  
## 308 - f 0 f 34 0.415 0  
## 309 - f 0 f 45 0.250 0  
## 310 - f 0 f 30 0.000 20  
## 311 - f 0 f 30 1.000 6  
## 312 - f 0 f 35 2.335 6  
## 313 - f 0 f 21 0.125 1  
## 314 - f 0 f 32 0.415 42  
## 315 - f 0 f 20 0.250 0  
## 316 - t 0 f 24 0.500 0  
## 317 - f 0 f 21 0.250 204  
## 318 + f 0 f 28 0.000 100000  
## 319 + f 0 f 39 0.000 1  
## 320 + f 0 f 27 1.500 113  
## 321 + f 0 f 29 1.500 8  
## 322 + f 0 f 34 10.000 0  
## 323 + f 0 f 34 0.375 44  
## 324 + t 11 t 34 0.250 2732  
## 325 - f 0 f 28 1.165 0  
## 326 - f 0 f 36 1.000 13  
## 327 - f 0 f 31 0.040 179  
## 328 - f 0 f 30 0.500 0  
## 329 - f 0 f 34 3.000 0  
## 332 - f 0 f 28 2.500 2  
## 333 - f 0 f 23 1.000 16  
## 334 - f 0 f 27 1.000 1062  
## 335 - f 0 f 21 0.500 0  
## 336 - f 0 f 27 0.165 251  
## 337 - f 0 f 35 1.000 228  
## 338 - f 0 f 29 0.500 0  
## 339 - f 0 f 29 2.000 0  
## 340 - f 0 f 31 0.750 67  
## 341 - f 0 f 32 3.000 0  
## 342 - f 0 f 18 0.040 100  
## 343 - f 0 f 32 0.500 4000  
## 344 - f 0 f 30 0.000 0  
## 345 - f 0 f 23 0.500 2  
## 346 - f 0 f 30 0.000 12  
## 347 - f 0 f 39 0.250 122  
## 348 - f 0 f 32 2.500 1210  
## 349 - t 3 t 27 0.585 0  
## 350 - f 1 t 27 2.250 3  
## 351 - f 0 f 33 0.000 1  
## 352 - f 0 f 42 0.000 0  
## 353 - f 0 f 25 1.500 4000  
## 354 - f 0 f 32 0.585 0  
## 355 - f 0 f 27 0.250 0  
## 356 - f 2 t 26 1.000 1  
## 357 - f 0 f 26 0.165 0  
## 358 - f 0 f 35 0.040 0  
## 359 - f 0 f 29 0.165 0  
## 360 - f 0 f 24 0.000 0  
## 361 - f 0 f 25 5.500 0  
## 362 - f 0 f 27 0.085 4208  
## 363 - f 0 f 28 0.000 0  
## 364 - f 0 f 31 0.290 0  
## 365 - f 2 t 25 0.165 1300  
## 366 - f 1 t 31 13.875 112  
## 367 - f 0 f 30 0.165 1000  
## 368 - f 0 f 27 0.165 0  
## 369 - f 0 f 24 3.000 16  
## 370 - f 0 f 32 0.750 2  
## 371 - f 0 f 23 7.000 0  
## 372 - f 0 f 31 0.000 1110  
## 373 - f 0 f 39 1.000 0  
## 374 - f 0 f 32 0.125 0  
## 375 - f 0 f 28 0.040 1004  
## 376 - f 0 f 29 1.000 0  
## 377 - f 0 f 28 0.125 286  
## 378 - f 0 f 29 2.000 0  
## 379 - f 0 f 35 0.125 4500  
## 380 - f 0 f 34 4.000 0  
## 381 - f 0 f 17 2.250 0  
## 382 - f 0 f 32 0.165 0  
## 383 - f 0 f 27 4.500 456  
## 384 - f 0 f 25 5.000 4  
## 385 - f 0 f 30 1.585 1212  
## 386 - f 0 f 26 1.500 0  
## 387 - f 0 f 27 0.540 67  
## 388 - f 0 f 33 0.500 0  
## 389 - f 0 f 21 0.000 0  
## 390 - f 0 f 36 0.085 0  
## 391 - f 0 f 28 1.000 0  
## 392 - f 0 f 36 0.210 0  
## 393 - f 0 f 30 0.250 0  
## 394 - f 0 f 28 1.000 1  
## 395 - f 0 f 31 0.250 195  
## 396 - f 0 f 26 0.540 0  
## 397 - f 0 f 33 0.040 1  
## 398 - f 0 f 33 0.125 87  
## 399 - f 0 f 25 1.250 17  
## 400 - f 0 f 34 0.085 0  
## 401 - f 0 f 34 0.290 184  
## 402 - f 0 f 34 0.290 140  
## 403 - f 0 f 33 3.085 0  
## 404 - f 0 f 23 0.750 0  
## 405 - f 0 f 27 1.085 0  
## 406 - f 0 f 30 0.000 0  
## 408 - f 1 t 45 1.000 2  
## 409 - f 1 t 33 0.085 6  
## 410 - f 4 t 23 0.335 8  
## 411 - f 5 t 34 0.000 146  
## 412 - f 1 t 31 1.250 22  
## 413 - f 0 f 29 0.085 0  
## 414 - f 0 f 30 0.000 0  
## 415 - f 0 f 28 0.125 55  
## 416 - f 0 f 29 3.250 0  
## 417 - f 0 f 26 0.125 70  
## 418 - f 0 f 33 0.540 1  
## 419 - f 2 t 35 0.375 500  
## 420 - f 0 f 36 0.000 60  
## 421 - f 0 f 28 1.000 0  
## 422 - f 0 f 39 1.500 7  
## 423 - f 0 f 31 1.750 0  
## 424 - f 0 f 24 1.165 0  
## 425 - f 0 f 26 1.500 0  
## 426 - f 0 f 30 0.250 0  
## 427 - f 0 f 34 0.250 50  
## 428 - f 0 f 39 0.040 5  
## 429 - f 0 f 26 0.290 3  
## 430 - f 0 f 30 0.085 0  
## 431 - f 0 f 34 1.500 4  
## 432 - f 0 f 34 0.165 1058  
## 433 - f 0 f 31 0.085 0  
## 434 - f 0 f 31 0.500 0  
## 435 - f 0 f 33 2.415 0  
## 436 - f 4 t 33 0.000 1  
## 437 - f 3 t 36 0.000 769  
## 438 - f 0 f 26 0.000 27  
## 439 - f 1 t 24 0.000 300  
## 440 - f 2 t 33 0.375 3  
## 441 - f 11 t 32 1.000 0  
## 442 - f 2 t 31 0.000 1  
## 443 - f 0 f 32 0.125 0  
## 444 - f 0 f 26 0.040 40  
## 445 - f 0 f 29 0.000 0  
## 447 - f 0 f 33 0.165 0  
## 448 - f 0 f 26 0.000 1  
## 449 - f 1 t 27 0.000 19  
## 450 - f 0 f 32 0.500 0  
## 452 - f 0 f 29 1.500 316  
## 453 - f 0 f 27 3.500 50  
## 454 - f 0 f 32 0.250 0  
## 455 - f 0 f 29 3.750 350  
## 456 - f 0 f 27 0.085 3552  
## 458 - f 0 f 27 0.040 0  
## 459 - f 0 f 36 5.000 687  
## 460 - f 0 f 28 1.500 0  
## 461 - f 0 f 27 0.000 0  
## 462 - f 4 t 48 0.085 1950  
## 463 - f 0 f 44 0.125 0  
## 464 - f 10 t 54 0.000 18  
## 465 - f 1 t 30 0.000 53  
## 466 - f 1 t 44 2.790 10  
## 467 - f 2 t 38 2.500 41  
## 468 - f 3 t 41 0.040 33  
## 469 - f 0 f 38 0.750 0  
## 470 - f 0 f 43 0.415 0  
## 471 - f 0 f 42 0.085 5  
## 472 - f 0 f 46 0.040 100  
## 473 - f 0 f 34 0.125 100  
## 474 - f 0 f 51 1.000 1000  
## 475 - f 0 f 37 0.125 44  
## 476 - f 0 f 48 0.750 0  
## 477 - f 0 f 43 0.085 5  
## 478 - f 0 f 40 10.000 0  
## 479 - f 0 f 44 0.415 0  
## 480 - f 0 f 40 0.085 0  
## 481 - f 6 t 36 0.165 35  
## 482 - f 1 t 37 0.415 80  
## 483 - f 10 t 34 1.750 10  
## 484 - f 2 t 40 0.040 6  
## 485 - f 0 f 33 1.165 0  
## 486 - f 2 t 41 0.040 351  
## 487 - f 0 f 49 0.085 2100  
## 488 - f 0 f 35 0.040 475  
## 489 - f 0 f 38 0.000 1  
## 490 - f 0 f 32 0.125 0  
## 491 + t 1 t 35 2.750 892  
## 492 + t 2 t 38 4.625 2000  
## 493 + t 14 t 41 6.500 4607  
## 494 + t 0 f 42 6.000 0  
## 495 + t 0 f 43 3.000 0  
## 496 + t 0 f 27 1.500 2206  
## 497 + t 0 f 47 1.040 5860  
## 498 + t 3 t 37 1.665 28  
## 499 + t 5 t 46 1.460 0  
## 500 + t 3 t 44 1.625 1391  
## 502 + t 10 t 36 3.500 0  
## 503 + t 1 t 53 0.000 100  
## 504 + t 2 t 43 4.750 7  
## 505 + t 1 t 32 1.085 0  
## 506 + t 16 t 34 7.000 5000  
## 507 + t 2 t 45 0.750 591  
## 508 + t 19 t 32 1.835 500  
## 509 + t 1 t 29 2.000 19  
## 510 + t 1 t 40 2.250 300  
## 511 + t 2 t 35 1.750 1000  
## 512 + t 0 f 42 0.000 960  
## 513 + t 0 f 31 2.500 0  
## 514 + t 0 f 34 0.000 0  
## 515 + t 0 f 40 2.585 0  
## 517 + t 4 t 49 4.000 99  
## 518 + t 5 t 53 1.750 690  
## 519 + t 4 t 36 0.585 0  
## 520 + t 5 t 32 0.125 0  
## 521 + t 1 t 37 1.500 234  
## 522 + t 5 t 47 2.250 500  
## 523 + t 1 t 39 1.290 800  
## 524 - t 10 t 54 1.750 990  
## 525 - t 0 f 32 2.415 0  
## 526 - t 0 f 34 2.500 0  
## 527 - t 0 f 36 0.210 0  
## 528 - t 0 f 36 1.000 2197  
## 529 - t 3 t 53 6.750 50  
## 530 - t 1 t 50 0.210 90  
## 531 - t 0 f 36 2.750 1  
## 532 - t 0 f 57 1.750 0  
## 533 - t 0 f 52 0.750 0  
## 534 - t 0 f 32 0.000 1  
## 535 - t 0 f 53 7.500 0  
## 536 - t 0 f 47 0.000 0  
## 537 - t 0 f 22 0.000 0  
## 538 - t 0 f 49 3.750 0  
## 539 - t 0 f 45 0.250 0  
## 540 - t 0 f 36 0.540 340  
## 541 - t 0 f 40 2.000 20  
## 542 - t 0 f 31 1.000 200  
## 543 - t 0 f 51 0.835 0  
## 544 - t 0 f 40 1.165 0  
## 545 - t 10 t 45 0.500 28  
## 546 - t 0 f 39 1.500 0  
## 547 - t 6 t 43 2.625 347  
## 548 + t 6 t 33 1.875 327  
## 549 + t 7 t 33 0.750 4071  
## 550 + t 0 f 49 16.000 0  
## 551 + t 1 t 43 12.750 109  
## 552 + t 11 t 56 0.000 1249  
## 553 + t 9 t 47 5.375 134  
## 554 + t 14 t 42 4.000 1344  
## 555 + t 4 t 39 0.750 321  
## 556 + t 6 t 43 0.210 948  
## 557 + t 2 t 45 7.500 0  
## 558 + t 16 t 40 1.085 2079  
## 559 + t 0 f 59 0.040 3000  
## 560 + t 7 t 29 2.290 2384  
## 561 + t 6 t 30 3.500 458  
## 562 + t 0 f 41 1.250 5298  
## 563 + t 0 f 42 1.415 200  
## 564 + t 0 f 44 1.585 0  
## 565 + t 0 f 47 12.750 0  
## 566 + t 0 f 22 0.040 0  
## 567 + t 11 t 24 2.125 284  
## 568 + t 0 f 47 0.875 0  
## 569 + t 8 t 62 0.375 162  
## 570 + t 4 t 49 0.750 1583  
## 571 + t 5 t 42 1.750 58  
## 572 + t 8 t 39 1.085 1  
## 573 + t 1 t 34 0.040 59  
## 574 + t 11 t 44 0.000 1400  
## 575 + t 4 t 47 1.000 1465  
## 576 + t 1 t 34 3.250 8000  
## 577 + t 11 t 45 1.750 540  
## 578 + t 3 t 44 1.000 0  
## 579 + t 10 t 45 1.500 4700  
## 580 + t 5 t 43 1.290 1097  
## 581 + t 8 t 44 1.335 3290  
## 582 + f 1 t 37 0.040 0  
## 583 + t 0 f 29 0.125 0  
## 584 + t 0 f 42 5.250 0  
## 585 + t 0 f 44 0.000 13212  
## 586 + t 0 f 26 0.000 0  
## 587 + t 9 t 44 17.500 1000  
## 588 + t 9 t 53 8.500 0  
## 589 + t 5 t 46 1.000 5777  
## 590 + t 7 t 34 0.290 5124  
## 591 + t 8 t 45 3.125 1200  
## 592 + t 3 t 25 4.250 150  
## 594 + f 0 f 38 0.085 6  
## 595 + t 0 f 37 0.085 0  
## 596 + t 0 f 45 0.250 23  
## 597 + t 8 t 46 2.375 4159  
## 598 + t 3 t 51 2.500 918  
## 599 + f 1 t 36 1.000 768  
## 600 + t 11 t 40 2.000 3000  
## 601 + t 4 t 42 0.540 500  
## 602 - f 0 f 42 0.000 1  
## 603 - f 0 f 37 0.250 0  
## 604 - f 0 f 38 0.125 0  
## 605 - t 0 f 34 0.085 0  
## 606 - f 0 f 32 1.250 0  
## 607 + f 0 f 36 0.040 0  
## 608 - f 0 f 39 0.500 0  
## 610 - f 0 f 44 0.085 0  
## 611 - f 0 f 32 0.125 4  
## 612 - f 2 t 37 5.085 1  
## 613 - f 1 t 45 0.290 283  
## 614 - f 2 t 42 0.585 7  
## 615 - f 0 f 57 0.125 0  
## 616 - f 2 t 33 0.250 108  
## 617 - f 1 t 42 1.585 9  
## 618 - f 2 t 32 0.000 1  
## 619 - f 1 t 52 2.000 68  
## 620 - t 0 f 34 0.125 375  
## 621 - f 0 f 47 0.125 10  
## 622 + f 0 f 42 2.250 0  
## 624 - f 0 f 39 0.665 1  
## 625 - f 0 f 43 0.665 0  
## 626 - f 0 f 49 2.085 0  
## 628 - f 0 f 39 0.000 1000  
## 629 - f 0 f 40 0.500 0  
## 630 - f 0 f 45 1.665 5  
## 631 - f 0 f 42 0.250 809  
## 632 - f 1 t 39 0.125 108  
## 633 - f 0 f 37 0.000 0  
## 634 - f 0 f 26 0.000 0  
## 635 - f 1 t 26 0.250 4  
## 636 - f 2 t 40 0.960 587  
## 637 - f 0 f 49 0.500 0  
## 638 - f 0 f 30 0.790 350  
## 639 - f 0 f 34 0.250 0  
## 640 - f 0 f 43 1.500 0  
## 641 - f 0 f 40 2.500 200  
## 642 - f 0 f 33 3.500 141  
## 643 - f 0 f 41 3.500 0  
## 644 - f 0 f 44 3.000 0  
## 645 - f 0 f 28 0.290 2  
## 646 - f 0 f 37 0.165 501  
## 647 - f 0 f 38 0.165 351  
## 648 - f 0 f 30 0.250 0  
## 649 - f 0 f 37 3.500 0  
## 650 - f 6 t 33 0.000 200  
## 651 - f 0 f 48 1.000 2  
## 652 - f 1 t 43 0.125 160  
## 653 - f 0 f 36 0.335 0  
## 654 - t 0 f 31 0.500 68  
## 655 - f 1 t 33 0.415 11  
## 656 - f 0 f 28 0.000 0  
## 657 - f 1 t 49 2.290 21  
## 658 - f 0 f 45 0.250 390  
## 659 - f 0 f 41 1.000 18  
## 660 - f 1 t 35 0.250 154  
## 661 - f 0 f 49 0.085 0  
## 662 - f 0 f 39 0.165 0  
## 663 - f 0 f 36 0.875 0  
## 664 - f 0 f 36 1.500 0  
## 665 - f 0 f 38 0.040 0  
## 666 - f 0 f 38 0.040 0  
## 667 - f 0 f 29 0.250 0  
## 668 - f 1 t 40 1.750 5  
## 669 - f 0 f 33 0.085 0  
## 670 - f 0 f 43 1.500 1  
## 671 - f 0 f 31 5.500 150  
## 672 - f 0 f 47 0.500 2  
## 673 - f 0 f 27 0.500 117  
## 674 - f 0 f 46 2.000 17  
## 675 - f 0 f 34 0.210 246  
## 676 - f 0 f 40 0.665 237  
## 677 - f 12 t 43 0.085 3  
## 678 - f 1 t 36 0.040 1  
## 679 - f 0 f 35 0.000 50  
## 680 - f 0 f 35 0.000 0  
## 681 - f 0 f 39 0.290 364  
## 682 - f 0 f 48 3.000 537  
## 683 - f 0 f 38 0.335 2  
## 684 - f 0 f 41 0.585 3  
## 685 - f 0 f 25 3.500 0  
## 686 - f 0 f 36 1.250 0  
## 687 - f 2 t 32 2.000 394  
## 688 - f 1 t 38 2.000 1  
## 689 - f 0 f 35 0.040 750  
## 690 - f 0 f 45 8.290 0

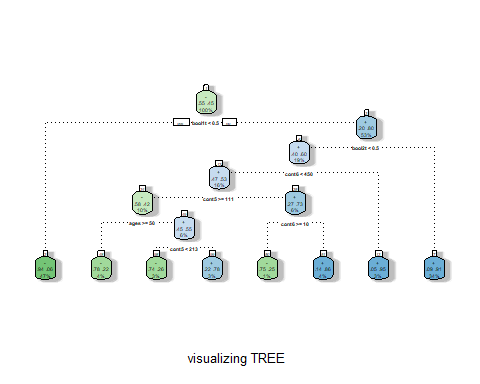
#partition into training and test sets  
indexdata = createDataPartition(y=rebuild$approval, p=0.7, list=FALSE)  
treeupdated1 <- train(approval ~., data = rebuild, method = "rpart1SE", trControl =train\_control)  
treeupdated1

## CART   
##   
## 666 samples  
## 6 predictor  
## 2 classes: '-', '+'   
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 599, 599, 599, 600, 599, 600, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.8679783 0.7346151

# Evaluate the fit with a confusion matrix  
predict\_tree <- predict(treeupdated1, Data04)  
# Confusion Matrix  
confusionMatrix(rebuild$approval, predict\_tree)

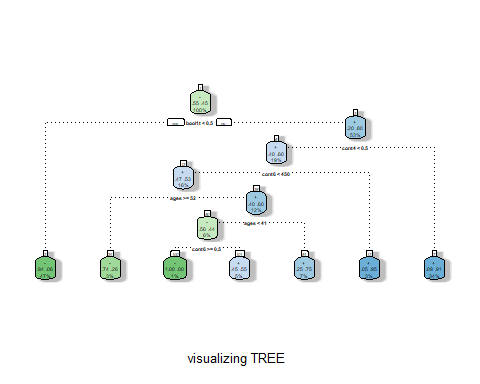
## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction - +  
## - 320 47  
## + 25 274  
##   
## Accuracy : 0.8919   
## 95% CI : (0.8658, 0.9144)  
## No Information Rate : 0.518   
## P-Value [Acc > NIR] : < 2e-16   
##   
## Kappa : 0.783   
##   
## Mcnemar's Test P-Value : 0.01333   
##   
## Sensitivity : 0.9275   
## Specificity : 0.8536   
## Pos Pred Value : 0.8719   
## Neg Pred Value : 0.9164   
## Prevalence : 0.5180   
## Detection Rate : 0.4805   
## Detection Prevalence : 0.5511   
## Balanced Accuracy : 0.8906   
##   
## 'Positive' Class : -   
##

fancyRpartPlot(tree4$finalModel, caption = "visualizing TREE")



**e. Visualize the trees from (a) and (d) and report if reducing the number of variables had an effect on the size of the tree?**

fancyRpartPlot(treeupdated1$finalModel, caption = "visualizing TREE")



**Inference after visualization of both the decision trees**:

By reducing the number of variables in the model, the decision tree has less number of branches and leaf nodes in it when compared to the model with the full data