Bank Marketing\_Naive Bayes

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Importing the dataset

bank <- read.csv("bank.csv",header = TRUE, sep = ",")

summary(bank)

## age job marital education   
## Min. :19.00 management :969 divorced: 528 primary : 678   
## 1st Qu.:33.00 blue-collar:946 married :2797 secondary:2306   
## Median :39.00 technician :768 single :1196 tertiary :1350   
## Mean :41.17 admin. :478 unknown : 187   
## 3rd Qu.:49.00 services :417   
## Max. :87.00 retired :230   
## (Other) :713   
## default balance housing loan contact   
## no :4445 Min. :-3313 no :1962 no :3830 cellular :2896   
## yes: 76 1st Qu.: 69 yes:2559 yes: 691 telephone: 301   
## Median : 444 unknown :1324   
## Mean : 1423   
## 3rd Qu.: 1480   
## Max. :71188   
##   
## day month duration campaign   
## Min. : 1.00 may :1398 Min. : 4 Min. : 1.000   
## 1st Qu.: 9.00 jul : 706 1st Qu.: 104 1st Qu.: 1.000   
## Median :16.00 aug : 633 Median : 185 Median : 2.000   
## Mean :15.92 jun : 531 Mean : 264 Mean : 2.794   
## 3rd Qu.:21.00 nov : 389 3rd Qu.: 329 3rd Qu.: 3.000   
## Max. :31.00 apr : 293 Max. :3025 Max. :50.000   
## (Other): 571   
## pdays previous poutcome y   
## Min. : -1.00 Min. : 0.0000 failure: 490 no :4000   
## 1st Qu.: -1.00 1st Qu.: 0.0000 other : 197 yes: 521   
## Median : -1.00 Median : 0.0000 success: 129   
## Mean : 39.77 Mean : 0.5426 unknown:3705   
## 3rd Qu.: -1.00 3rd Qu.: 0.0000   
## Max. :871.00 Max. :25.0000   
##

str(bank)

## 'data.frame': 4521 obs. of 17 variables:  
## $ age : int 30 33 35 30 59 35 36 39 41 43 ...  
## $ job : Factor w/ 12 levels "admin.","blue-collar",..: 11 8 5 5 2 5 7 10 3 8 ...  
## $ marital : Factor w/ 3 levels "divorced","married",..: 2 2 3 2 2 3 2 2 2 2 ...  
## $ education: Factor w/ 4 levels "primary","secondary",..: 1 2 3 3 2 3 3 2 3 1 ...  
## $ default : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ balance : int 1787 4789 1350 1476 0 747 307 147 221 -88 ...  
## $ housing : Factor w/ 2 levels "no","yes": 1 2 2 2 2 1 2 2 2 2 ...  
## $ loan : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 1 1 1 2 ...  
## $ contact : Factor w/ 3 levels "cellular","telephone",..: 1 1 1 3 3 1 1 1 3 1 ...  
## $ day : int 19 11 16 3 5 23 14 6 14 17 ...  
## $ month : Factor w/ 12 levels "apr","aug","dec",..: 11 9 1 7 9 4 9 9 9 1 ...  
## $ duration : int 79 220 185 199 226 141 341 151 57 313 ...  
## $ campaign : int 1 1 1 4 1 2 1 2 2 1 ...  
## $ pdays : int -1 339 330 -1 -1 176 330 -1 -1 147 ...  
## $ previous : int 0 4 1 0 0 3 2 0 0 2 ...  
## $ poutcome : Factor w/ 4 levels "failure","other",..: 4 1 1 4 4 1 2 4 4 1 ...  
## $ y : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...

The new bank dataset after removing age, job, previous, default, pdays, balance

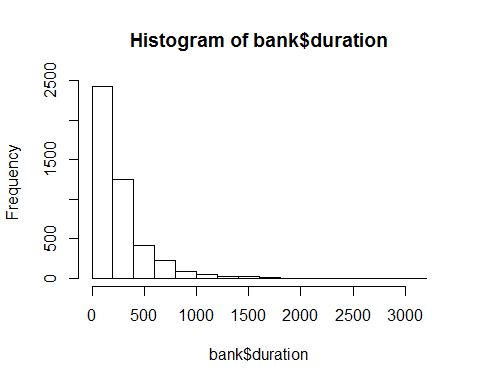
bank <- bank[c(3:4, 7:13, 16:17)]  
str(bank)

## 'data.frame': 4521 obs. of 11 variables:  
## $ marital : Factor w/ 3 levels "divorced","married",..: 2 2 3 2 2 3 2 2 2 2 ...  
## $ education: Factor w/ 4 levels "primary","secondary",..: 1 2 3 3 2 3 3 2 3 1 ...  
## $ housing : Factor w/ 2 levels "no","yes": 1 2 2 2 2 1 2 2 2 2 ...  
## $ loan : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 1 1 1 2 ...  
## $ contact : Factor w/ 3 levels "cellular","telephone",..: 1 1 1 3 3 1 1 1 3 1 ...  
## $ day : int 19 11 16 3 5 23 14 6 14 17 ...  
## $ month : Factor w/ 12 levels "apr","aug","dec",..: 11 9 1 7 9 4 9 9 9 1 ...  
## $ duration : int 79 220 185 199 226 141 341 151 57 313 ...  
## $ campaign : int 1 1 1 4 1 2 1 2 2 1 ...  
## $ poutcome : Factor w/ 4 levels "failure","other",..: 4 1 1 4 4 1 2 4 4 1 ...  
## $ y : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...

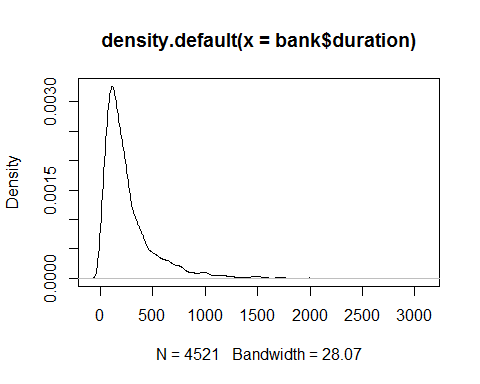
bank$campaign <- as.factor(bank$campaign)

Check for normality of duration

hist(bank$duration)



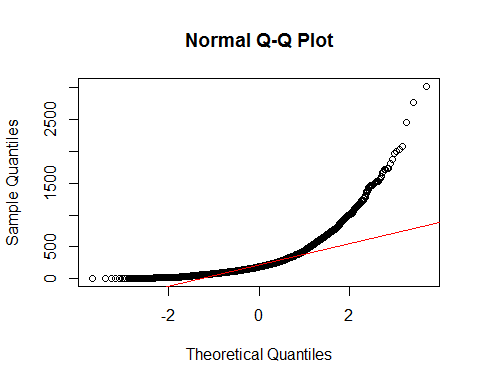
plot(density(bank$duration))



shapiro.test(bank$duration)

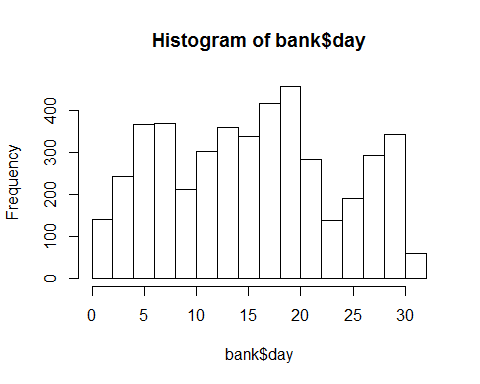
##   
## Shapiro-Wilk normality test  
##   
## data: bank$duration  
## W = 0.74754, p-value < 2.2e-16

qqnorm(bank$duration);qqline(bank$duration, col = 2)

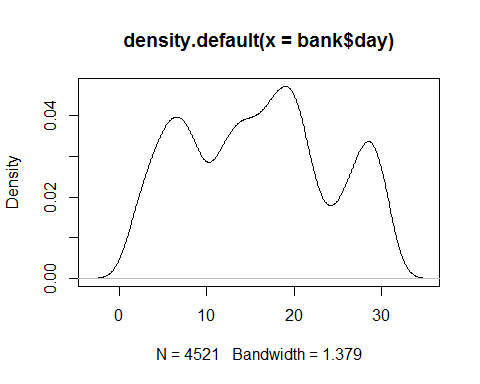


Check for normality of day

hist(bank$day)



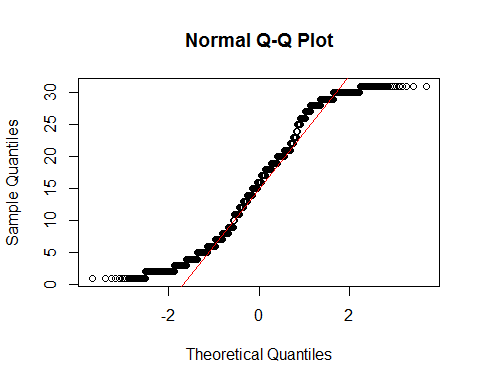
plot(density(bank$day))



shapiro.test(bank$day)

##   
## Shapiro-Wilk normality test  
##   
## data: bank$day  
## W = 0.96072, p-value < 2.2e-16

qqnorm(bank$day);qqline(bank$day, col = 2)



for(i in 1 : nrow(bank))  
{  
 if (bank$duration[i] < 240)  
 {  
 bank$duration[i] = '00 to 04 min'  
 }  
 else if (bank$duration[i] < 480 & bank$duration[i] > 240)  
 {  
 bank$duration[i] = '04 to 08 min'  
 }   
 else if (bank$duration[i] < 720 & bank$duration[i] > 480)  
 {  
 bank$duration[i] = '08 to 12 min'  
 }   
 else if (bank$duration[i] < 960 & bank$duration[i] > 720)  
 {  
 bank$duration[i] = '12 to 16 min'  
 }  
 else if (bank$duration[i] < 1200 & bank$duration[i] > 960)  
 {  
 bank$duration[i] = '16 to 20 min'  
 }  
 else if (bank$duration[i] < 1440 & bank$duration[i] > 1200)  
 {  
 bank$duration[i] = '20 to 24 min'  
 }  
 else if (bank$duration[i] < 1680 & bank$duration[i] > 1440)  
 {  
 bank$duration[i] = '24 to 28 min'  
 }  
 else if (bank$duration[i] < 1920 & bank$duration[i] > 1680)  
 {  
 bank$duration[i] = '28 to 32 min'  
 }  
 else if (bank$duration[i] < 2160 & bank$duration[i] > 1920)  
 {  
 bank$duration[i] = '32 to 36 min'  
 }  
 else if (bank$duration[i] > 2160)  
 {  
 bank$duration[i] = ' > 36 min '  
 }  
}

bank$duration <- as.factor(bank$duration)  
table(bank$duration)

##   
## > 36 min 00 to 04 min 04 to 08 min 08 to 12 min 12 to 16 min   
## 74 2002 1262 655 528

for(i in 1 : nrow(bank))  
{  
 if (bank$day[i] < 7)  
 {  
 bank$day[i] = '1st week'  
 }  
 else if (bank$day[i] < 14 & bank$day[i] > 7)  
 {  
 bank$day[i] = '2nd week'  
 }   
 else if (bank$day[i] < 21 & bank$day[i] > 14)  
 {  
 bank$day[i] = '3rd week'  
 }   
 else if (bank$day[i] < 28 & bank$day[i] > 21)  
 {  
 bank$day[i] = '4th week'  
 }  
   
 else if (bank$day[i] > 28)  
 {  
 bank$day[i] = '5th week'  
 }  
}

bank$day <- as.factor(bank$day)  
table(bank$day)

##   
## 1st week 3rd week 5th week   
## 3987 1 533

Splitting the dataset into training set and test set

# install.packages("caTools")  
library(caTools)  
set.seed(123)  
split = sample.split(bank$y, SplitRatio = 2/3)  
training\_set = subset(bank, split == TRUE)  
test\_set = subset(bank, split == FALSE)

Fitting classifier to the training test

# install.packages("e1071")  
library(e1071)  
classifier = naiveBayes(y ~ ., data = training\_set)  
print(classifier)

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## no yes   
## 0.8848706 0.1151294   
##   
## Conditional probabilities:  
## marital  
## Y divorced married single  
## no 0.1106112 0.6355456 0.2538433  
## yes 0.1296830 0.5504323 0.3198847  
##   
## education  
## Y primary secondary tertiary unknown  
## no 0.15673041 0.51143607 0.28871391 0.04311961  
## yes 0.12391931 0.46397695 0.37463977 0.03746398  
##   
## housing  
## Y no yes  
## no 0.4150731 0.5849269  
## yes 0.5878963 0.4121037  
##   
## loan  
## Y no yes  
## no 0.84214473 0.15785527  
## yes 0.90489914 0.09510086  
##   
## contact  
## Y cellular telephone unknown  
## no 0.61717285 0.06374203 0.31908511  
## yes 0.79538905 0.08357349 0.12103746  
##   
## day  
## Y 1st week 3rd week 5th week  
## no 0.8758905137 0.0003749531 0.1237345332  
## yes 0.9135446686 0.0000000000 0.0864553314  
##   
## month  
## Y apr aug dec feb jan  
## no 0.058867642 0.142482190 0.002624672 0.044244469 0.033370829  
## yes 0.118155620 0.141210375 0.014409222 0.074927954 0.037463977  
## month  
## Y jul jun mar may nov  
## no 0.162729659 0.122984627 0.007874016 0.322084739 0.083614548  
## yes 0.106628242 0.103746398 0.037463977 0.181556196 0.086455331  
## month  
## Y oct sep  
## no 0.010498688 0.008623922  
## yes 0.060518732 0.037463977  
##   
## duration  
## Y > 36 min 00 to 04 min 04 to 08 min 08 to 12 min 12 to 16 min  
## no 0.01424822 0.46831646 0.27071616 0.13348331 0.11323585  
## yes 0.02881844 0.30547550 0.32564841 0.18443804 0.15561960  
##   
## campaign  
## Y 1 2 3 4 5  
## no 0.3715785527 0.2759655043 0.1229846269 0.0738657668 0.0438695163  
## yes 0.4524495677 0.2737752161 0.1095100865 0.0806916427 0.0201729107  
## campaign  
## Y 6 7 8 9 10  
## no 0.0359955006 0.0183727034 0.0119985002 0.0067491564 0.0071241095  
## yes 0.0345821326 0.0144092219 0.0028818444 0.0028818444 0.0000000000  
## campaign  
## Y 11 12 13 14 15  
## no 0.0048743907 0.0048743907 0.0033745782 0.0029996250 0.0029996250  
## yes 0.0000000000 0.0028818444 0.0000000000 0.0000000000 0.0000000000  
## campaign  
## Y 16 17 18 19 20  
## no 0.0014998125 0.0011248594 0.0014998125 0.0003749531 0.0011248594  
## yes 0.0000000000 0.0028818444 0.0000000000 0.0000000000 0.0000000000  
## campaign  
## Y 21 22 23 24 25  
## no 0.0003749531 0.0007499063 0.0007499063 0.0007499063 0.0011248594  
## yes 0.0000000000 0.0000000000 0.0000000000 0.0028818444 0.0000000000  
## campaign  
## Y 28 29 30 31 32  
## no 0.0007499063 0.0003749531 0.0003749531 0.0003749531 0.0003749531  
## yes 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  
## campaign  
## Y 44 50  
## no 0.0003749531 0.0003749531  
## yes 0.0000000000 0.0000000000  
##   
## poutcome  
## Y failure other success unknown  
## no 0.10498688 0.04311961 0.01162355 0.84026997  
## yes 0.10662824 0.08357349 0.16138329 0.64841499

Predicting the test results

y\_pred = predict(classifier, newdata = test\_set)  
y\_pred

## [1] no no no no no no no no no no no no no no no no no   
## [18] no no no no no no no no no no no no no no no no yes  
## [35] no no no no no yes no no no no no no no no no no no   
## [52] no no no no no no no yes no no no no no no yes yes no   
## [69] no no no no no no no no no no no no no no no no no   
## [86] no no no no no no no no no no no no no no no no no   
## [103] no no no no no no no no no no no no no no no no no   
## [120] no no no no no no no no no no no no no no yes no no   
## [137] no no no no no no no no no no no no no no no no no   
## [154] no no no no no no no no no no no no no no no no no   
## [171] no no no no no no no no no no no no no no no no no   
## [188] no no no no no no no no no no no yes no no no no no   
## [205] no no yes no no no no no no no no no no no no no no   
## [222] no no no no no no no no no yes no no no no no no no   
## [239] no no no no no no no no no no no yes no no no no no   
## [256] no no no no no no no no no no no no no no no no no   
## [273] no no no no no no no no no no no yes no no no no no   
## [290] no yes no no no no no no no no no no no no no no no   
## [307] no no no no no no no no no no no no no no no no no   
## [324] no no no no no yes no no no no no no no no no no no   
## [341] no no no no no no no no no no no no no no no no no   
## [358] no yes no no no no no no no no no no no no no no no   
## [375] no no no no no no no no no no no no no no no yes yes  
## [392] no no yes no no no no no no no no no no no no no no   
## [409] yes no no no no no no no no no no no no no no no no   
## [426] no no no no no no no no yes no no no no no no no no   
## [443] no no no yes no no no no no no no no no no no no no   
## [460] no no no no no no no no no no no no no no no no no   
## [477] no no no no no no yes no no no no no no no no no no   
## [494] no no no no no no no no no no no no no no yes no no   
## [511] no no no no no no no no no no yes no no no no no no   
## [528] no no no no no no no no no no no no no no no no no   
## [545] no no no no no no no no no no no no no no no no no   
## [562] yes no no no no no no no no no no no yes no no no yes  
## [579] no no yes no no no no no no yes no no no no no no no   
## [596] no no no no no no no no no no no no no no no no no   
## [613] no no no no no no no no no no no no no no no no no   
## [630] yes no no no no no no no no no no no no no no no no   
## [647] no no no no no no no no no no no no no no no no no   
## [664] yes no no no no no no no no no no no no no no no yes  
## [681] no no no no no no no no no no no no no no no no no   
## [698] no no no no no no no no no no no no no no no no no   
## [715] no no no no no no no no no no no yes no no no no no   
## [732] no no no no no yes no no yes no no no no no no no no   
## [749] no no no no no no no no no no no no no no no no no   
## [766] no no no no no no no no no yes no no no no no no no   
## [783] no no no no no no no no no yes no yes no no no no no   
## [800] no no yes yes no no no no no no no no no no no no no   
## [817] no no no no no no no no no no no no no no no no no   
## [834] no no no no no no no no no no no no no no no no no   
## [851] no no no no no no no no no no no no no no no no no   
## [868] no no no no no no yes no no no no no no yes no no no   
## [885] no no no no no yes no no no no no no no no no no no   
## [902] no yes no no no no no no no no no no no no no no no   
## [919] no no no no no no no no no no no no no no no no no   
## [936] no no no no no no no no no no no no no no no yes yes  
## [953] no no no yes no no no no no no no no no no no no no   
## [970] no no no no no no no no no no no yes no no no no no   
## [987] no no no no no no no no no no no no no no no no no   
## [1004] no no yes no no no no no no no no no no no no no no   
## [1021] no no no no no no no no no no no no no no no no no   
## [1038] no yes no no no no no no no no no no no yes no no no   
## [1055] no no no no no no no no no no no no no no no no yes  
## [1072] no no no no no no no no no no no no no no no no yes  
## [1089] no no no no no no no no no no yes no no no no yes no   
## [1106] no no no no no no no no no no no no yes no no no no   
## [1123] no no no no no no no no no no no no yes no no no no   
## [1140] no no no no no no no no no no no no no no no no no   
## [1157] no no no no no yes no no no no no no yes no no no no   
## [1174] no no no no no no no no no no no no no no no no no   
## [1191] yes no no yes no no no no no no no no no no no no no   
## [1208] no no no no no no no no no no no no no no no no no   
## [1225] no no no no no no no no no no no no no no no no no   
## [1242] no no no yes no no no no no no no no no no no no no   
## [1259] no no no yes no no yes no no no no no no no no no no   
## [1276] no no yes no no no no no no no no no no no no yes no   
## [1293] yes no no no no no no no no no yes no no no no no no   
## [1310] no no no no no no no no no no no no no no no no no   
## [1327] no no no no no no no no no no no no no no no no no   
## [1344] no no no no no no no no no no no no yes yes no no no   
## [1361] no no no no no no no no no no no no no no yes no no   
## [1378] no no no no no no no no no no no no no no no no no   
## [1395] no no no no no yes no no no yes no no no no no no no   
## [1412] no no no no no no no no no no no no no no no yes no   
## [1429] no no no no no no no no no no yes no yes no no no no   
## [1446] no no no no no no no no no no no no no no no no no   
## [1463] no no no no no no no no no no no yes no no no no no   
## [1480] no no no no no no yes no no no no no no no no no no   
## [1497] no no no no no no no no no no no   
## Levels: no yes

Confusion matrix

cm = table(test\_set$y, y\_pred)  
cm

## y\_pred  
## no yes  
## no 1299 34  
## yes 131 43

Accuracy is the ratio of correctly predicted observation to the total observations

accuracy <- sum(diag(cm))/sum(cm)   
accuracy

## [1] 0.8905109

So we got 89% accuracy Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

precision <- cm[2,2]/(sum(cm[,2]))  
precision

## [1] 0.5584416

We got 56% precision Recall is the ratio of correctly predicted positive observations to the all observations in actual class

recall <- cm[2,2]/(sum(cm[2,]))  
recall

## [1] 0.2471264

We got 25% recall

F1 <- 2\*(recall \* precision) / (recall + precision)  
F1

## [1] 0.3426295

34% is the F1 score