```
tinytex::install_tinytex()
```

```
## Warning in file(con, "r"): cannot open file '/texmf-dist/web2c/fmtutil.cnf': No
## such file or directory
```

```
## tlmgr conf auxtrees add "C:/PROGRA~1/R/R-40~1.2/share/texmf"
```

```
FlightDelays <- read.csv("FlightDelays.csv")
library(caret)
```

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

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

```
library(ISLR)
library(naivebayes)
```

```
## Warning: package 'naivebayes' was built under R version 4.0.3
```

```
## naivebayes 0.9.7 loaded
```

```
# Data preparation
FlightDelays$DAY_WEEK <- factor(FlightDelays$DAY_WEEK)
FlightDelays$DAY_OF_MONTH <- factor(FlightDelays$DAY_OF_MONTH)
FlightDelays$CRS_DEP_TIME<-factor(round(FlightDelays$CRS_DEP_TIME/100))
```

```
DelayVariables <- c(1, 2, 4, 8, 10, 13)
set.seed(15)
delay.part = createDataPartition(FlightDelays$Flight.Status, p = 0.6, list = FALSE)
delay.train = FlightDelays[delay.part, DelayVariables]
delay.validate <- FlightDelays[-delay.part, DelayVariables]
delay.model <- naive_bayes(Flight.Status~., data = delay.train)
delay.model
```

```
##
## ================================== Naive Bayes ==================================
##
##  Call:
## naive_bayes.formula(formula = Flight.Status ~ ., data = delay.train)
##
## ---------------------------------------------------------------------------------
##
## Laplace smoothing: 0
##
## ---------------------------------------------------------------------------------
##
##  A priori probabilities:
##
##   delayed    ontime
## 0.1945496 0.8054504
```

```
##
## ------------------------------------------------------------------------------
##
##  Tables:
##
## ------------------------------------------------------------------------------
##  ::: CRS_DEP_TIME (Categorical)
## ------------------------------------------------------------------------------
##
## CRS_DEP_TIME    delayed     ontime
##           6  0.02723735 0.05263158
##           7  0.05836576 0.06484962
##           8  0.06225681 0.07988722
##           9  0.01945525 0.05169173
##           10 0.02334630 0.05357143
##           11 0.01556420 0.03571429
##           12 0.05447471 0.05921053
##           13 0.03501946 0.06860902
##           14 0.05836576 0.06109023
##           15 0.21011673 0.11466165
##           16 0.07003891 0.08082707
##           17 0.14007782 0.10808271
##           18 0.02723735 0.04229323
##           19 0.09338521 0.04981203
##           20 0.02334630 0.02631579
##           21 0.08171206 0.05075188
##
## ------------------------------------------------------------------------------
##  ::: CARRIER (Categorical)
## ------------------------------------------------------------------------------
##
## CARRIER     delayed      ontime
##      CO 0.054474708 0.040413534
##      DH 0.295719844 0.217105263
##      DL 0.136186770 0.190789474
##      MQ 0.194552529 0.125939850
##      OH 0.007782101 0.017857143
##      RU 0.233463035 0.184210526
##      UA 0.011673152 0.016917293
##      US 0.066147860 0.206766917
##
## ------------------------------------------------------------------------------
##  ::: DEST (Categorical)
## ------------------------------------------------------------------------------
##
## DEST    delayed    ontime
##   EWR 0.3891051 0.2838346
##   JFK 0.1712062 0.1616541
##   LGA 0.4396887 0.5545113
##
## ------------------------------------------------------------------------------
##  ::: ORIGIN (Categorical)
## ------------------------------------------------------------------------------
##
```

```
## ORIGIN    delayed     ontime
##    BWI 0.08171206 0.06578947
##    DCA 0.52140078 0.64943609
##    IAD 0.39688716 0.28477444
##
## ----------------------------------------------------------------------------
##  ::: DAY_WEEK (Categorical)
## ----------------------------------------------------------------------------
##
## DAY_WEEK    delayed     ontime
##        1 0.17509728 0.12593985
##        2 0.15564202 0.12969925
##        3 0.14007782 0.14379699
##        4 0.12451362 0.19078947
##        5 0.18287938 0.17669173
##        6 0.06225681 0.13251880
##        7 0.15953307 0.10056391
##
## ----------------------------------------------------------------------------
```

```
#Counts table and Proportion
table(delay.train$Flight.Status, delay.train$DEST)
```

```
##
##           EWR JFK LGA
##   delayed 100  44 113
##   ontime  302 172 590
```

```
prop.table(table(delay.train$Flight.Status, delay.train$DEST), margin = 1)
```

```
##
##                 EWR       JFK       LGA
##   delayed 0.3891051 0.1712062 0.4396887
##   ontime  0.2838346 0.1616541 0.5545113
```

```
library(gmodels)
new.train <- predict(delay.model, newdata = delay.train, type = "class")
```

```
## Warning: predict.naive_bayes(): more features in the newdata are provided as
## there are probability tables in the object. Calculation is performed based on
## features to be found in the tables.
```

```
CrossTable(x=delay.train$Flight.Status, y=new.train, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
```

```
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1321
##
##
##                              | new.train
## delay.train$Flight.Status |   delayed |    ontime | Row Total |
## --------------------------|-----------|-----------|-----------|
##                    delayed |        31 |       226 |       257 |
##                            |     0.121 |     0.879 |     0.195 |
##                            |     0.437 |     0.181 |           |
##                            |     0.023 |     0.171 |           |
## --------------------------|-----------|-----------|-----------|
##                     ontime |        40 |      1024 |      1064 |
##                            |     0.038 |     0.962 |     0.805 |
##                            |     0.563 |     0.819 |           |
##                            |     0.030 |     0.775 |           |
## --------------------------|-----------|-----------|-----------|
##              Column Total |        71 |      1250 |      1321 |
##                            |     0.054 |     0.946 |           |
## --------------------------|-----------|-----------|-----------|
##
##
```

```
new.val <- predict(delay.model, newdata=delay.validate, type="prob")
```

```
## Warning: predict.naive_bayes(): more features in the newdata are provided as
## there are probability tables in the object. Calculation is performed based on
## features to be found in the tables.
```

```
classpredict <- predict(delay.model, newdata = delay.validate)
```

```
## Warning: predict.naive_bayes(): more features in the newdata are provided as
## there are probability tables in the object. Calculation is performed based on
## features to be found in the tables.
```

```
CrossTable(x= delay.validate$Flight.Status, y=classpredict, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  880
```

```
##
##
##                              | classpredict
## delay.validate$Flight.Status |   delayed |    ontime | Row Total |
## -----------------------------|-----------|-----------|-----------|
##                      delayed |        16 |       155 |       171 |
##                              |     0.094 |     0.906 |     0.194 |
##                              |     0.444 |     0.184 |           |
##                              |     0.018 |     0.176 |           |
## -----------------------------|-----------|-----------|-----------|
##                       ontime |        20 |       689 |       709 |
##                              |     0.028 |     0.972 |     0.806 |
##                              |     0.556 |     0.816 |           |
##                              |     0.023 |     0.783 |           |
## -----------------------------|-----------|-----------|-----------|
##                 Column Total |        36 |       844 |       880 |
##                              |     0.041 |     0.959 |           |
## -----------------------------|-----------|-----------|-----------|
##
##
```

```
require(pROC)
```

```
## Loading required package: pROC

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

##
## Attaching package: 'pROC'

## The following object is masked from 'package:gmodels':
##
##     ci

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

```
roc(delay.validate$Flight.Status,new.val[,1])
```

```
## Setting levels: control = delayed, case = ontime

## Setting direction: controls > cases

##
## Call:
## roc.default(response = delay.validate$Flight.Status, predictor = new.val[,     1])
##
## Data: new.val[, 1] in 171 controls (delay.validate$Flight.Status delayed) > 709 cases (delay.validat
## Area under the curve: 0.6716
```

```
plot.roc(delay.validate$Flight.Status,new.val[,1])
```

```
## Setting levels: control = delayed, case = ontime
## Setting direction: controls > cases
```