Assignment2-1_Small

```
library(caret)
library(ggplot2)
library(e1071)


#Reading and Cleaning the Data
df <- read.csv("/Users/Nick/Desktop/delay.csv", stringsAsFactors=FALSE)
Day.of.Week <- as.factor(df$Day.of.Week)
Status <- factor(df$Status, levels = c("0", "1"), labels = c("on-time", "Dela
yed"))
df <- cbind(Day.of.Week, df[,-1])
df <- cbind(Status, df[,-5])

#Training Dataset
set.seed(1234)
intrain <- createDataPartition(y=df$Status, p=0.7, list=FALSE)
trainset <- df[intrain,]
testset <- df[-intrain,]

#Naive Model
nb.model <- naiveBayes(Status~., data=trainset)
nb.model
```

The following below shows the Conditional probabilities based on Days of the week, Carrier, Origin, and Destination.

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##    on-time    Delayed
## 0.92358804 0.07641196
##
## Conditional probabilities:
##          Day.of.Week
## Y                 1         2         3         4         5         6
7
##   on-time 0.1654676 0.1474820 0.1510791 0.1223022 0.1798561 0.0647482 0.16
90647
##   Delayed 0.5652174 0.4347826 0.0000000 0.0000000 0.0000000 0.0000000 0.00
```

```
00000
##
##          Carrier
## Y                  CO          DH          DL          MQ          OH
##   on-time 0.057553957 0.287769784 0.129496403 0.176258993 0.010791367
##   Delayed 0.086956522 0.304347826 0.086956522 0.304347826 0.000000000
##          Carrier
## Y                  RU          UA          US
##   on-time 0.248201439 0.007194245 0.082733813
##   Delayed 0.086956522 0.086956522 0.043478261
##
##          Origin
## Y                BWI       DCA       IAD
##   on-time 0.1043165 0.5287770 0.3669065
##   Delayed 0.0000000 0.5652174 0.4347826
##
##          Destination
## Y                EWR       JFK       LGA
##   on-time 0.4064748 0.1834532 0.4100719
##   Delayed 0.2608696 0.2173913 0.5217391

#Prediction Function
nb.model.pred <- predict(nb.model, testset, type='class')
nb.model.pred

##   [1] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [10] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [19] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [28] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [37] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [46] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [55] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [64] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [73] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [82] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
##  [91] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
## [100] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
## [109] on-time on-time on-time on-time on-time on-time on-time on-time on-t
```

```
ime
## [118] on-time on-time on-time on-time on-time on-time on-time on-time on-t
ime
## [127] on-time
## Levels: on-time Delayed

#Confusion Table
actual <- testset$Status
confusionMatrix(actual, nb.model.pred)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction on-time Delayed
##    on-time     118       0
##    Delayed       9       0
##
##                Accuracy : 0.9291
##                  95% CI : (0.8697, 0.9671)
##     No Information Rate : 1
##     P-Value [Acc > NIR] : 1.000000
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 0.007661
##
##             Sensitivity : 0.9291
##             Specificity :     NA
##          Pos Pred Value :     NA
##          Neg Pred Value :     NA
##              Prevalence : 1.0000
##          Detection Rate : 0.9291
##    Detection Prevalence : 0.9291
##       Balanced Accuracy :     NA
##
##        'Positive' Class : on-time
```

Analysis of Confusion Matrix:

There are 0 false positives, 9 false negatives, 118 True Positives, and 0 True Negatives. The Naïve Bayes Model is based on the testset is 92% accurate. This is a very accurate model.

```
#Equations
# {Status=Delayed} given {Carrier = DL, Day of Week=Saturday (7), Destination
= LGA, and Origin = DCA.}
# P{Carrier = DL} -> 0.086956522
# P{Day of the week = Saturday (7)} -> 0.0000000
# P{Destination = LGA} -> 0.5217391
# P{Origin = DCA} -> 0.5652174
```

```
# P{Status = Delayed} -> 0.07641196

Probability.Delayed = 0.086956522 * 0.0000000 * 0.5217391 * 0.5652174 * 0.076
41196
```

The Probability of Delayed is:

```
## [1] 0
```

```
# P{Carrier = DL} -> 0.129496403
# P{Day of the week = Saturday (7)} -> 0.1690647
# P{Destination = LGA} -> 0.4100719
# P{Origin = DCA} -> 0.5287770
# P{Status = on-time} -> 0.92358804

Probability.ontime = 0.129496403 * 0.1690647 * 0.4100719 * 0.5287770 *0.92358
804
```

The Probability of being ontime is:

```
## [1] 0.004384515
```

```
# P{Status=Delayed} given {Carrier = DL, Day of Week=Saturday (7), Destinatio
n = LGA, and Origin = DCA.}
Probability = (0)/(0.004384515+0)
Probability
```

```
## [1] 0
```