FML Assignment 3

Bhavika

```
library(readr)

Data<-read.csv("UniversalBank.csv")
head(Data)
```

```
##   ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage
## 1  1  25          1     49    91107      4   1.6         1        0
## 2  2  45         19     34    90089      3   1.5         1        0
## 3  3  39         15     11    94720      1   1.0         1        0
## 4  4  35          9    100    94112      1   2.7         2        0
## 5  5  35          8     45    91330      4   1.0         2        0
## 6  6  37         13     29    92121      4   0.4         2      155
##   Personal.Loan Securities.Account CD.Account Online CreditCard
## 1             0                  1          0      0          0
## 2             0                  1          0      0          0
## 3             0                  0          0      0          0
## 4             0                  0          0      0          0
## 5             0                  0          0      0          1
## 6             0                  0          0      1          0
```

```
#Converting Personal Loan, Online, Credit Card
Data$Personal.Loan<-as.factor(Data$Personal.Loan)
Data$Online<-as.factor(Data$Online)
Data$CreditCard<-as.factor(Data$CreditCard)
```

```
#Data Partitioning into Training and Validation set with 60% and 40%
library(caret)
```

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

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

```
set.seed(3862)
Split_data<- createDataPartition(Data$Personal.Loan,p=.6,list=FALSE,times=1)
Training<-Data[Split_data,]
Validation<-Data[-Split_data,]
```

```
#Normalizing the Data:

Normalization<-preProcess(Training[,-c(10,13,14)],method=c("center","scale"))
Training_Normalized_Data<-predict(Normalization,Training)
Validation_Normalized_Data<-predict(Normalization,Validation)
```

A. Create a pivot table for the training data with Online as a column variable, CC as a row variable, and Loan as a secondary row variable:

```
Table_1<-ftable(Training_Normalized_Data[,c(14,10,13)])
Table_1
```

```
##                          Online    0    1
## CreditCard Personal.Loan
## 0          0                     789 1111
##            1                      83  128
## 1          0                     321  491
##            1                      38   39
```

```
head(Table_1)
```

```
##
##                               "Online"  "0"  "1"
##  "CreditCard" "Personal.Loan"
##  "0"          "0"                      789 1111
##               "1"                       83  128
##  "1"          "0"                      321  491
##               "1"                       38   39
```

```
prop_table1<-prop.table(Table_1)
prop_table1
```

```
##                          Online          0          1
## CreditCard Personal.Loan
## 0          0                    0.26300000 0.37033333
##            1                    0.02766667 0.04266667
## 1          0                    0.10700000 0.16366667
##            1                    0.01266667 0.01300000
```

B. The probability of customer accepting loan and using credit card plus being an online banking user = P(Loan = 1|CreditCard = 1, Online = 1) = 39/(39+491) = 0.0735

C. Creating a Pivot table having Loan as a row variable and CreditCard as column Variable:

```
Table2 <- table(Training_Normalized_Data$Personal.Loan,Training_Normalized_Data$CreditCard)
Table2
```

```
##
##        0    1
##   0 1900  812
##   1  211   77
```

```
Table3 <- table(Training_Normalized_Data$Personal.Loan,Training_Normalized_Data$Online)
Table3
```

```
##
##        0    1
##   0 1110 1602
##   1  121  167
```

D. (i)P(CC=1|Loan=1) = 77/(211+77) = 0.267

D. (ii.)P(Online=1| Loan=1) = 167/(121+167) =0.5798

D. (iii) P(Loan =1):

```
table_loan <- table(Training_Normalized_Data$Personal.Loan)
table_loan
```

```
##
##    0    1
## 2712  288
```

P(Loan=1)=288/(288+2712) = 0.096

D. (iv.)P(CC=1|Loan=0)= 812/(812+1900) =0.2994

D. (v.)P(Online=1|Loan=0)= 1602/(1602+1110) = 0.590

D. (vi.) P(Loan=0)=2712/(2712+288)=0.904

E. Naive Bayes probablitity: P(Loan = 1|CC = 1, Online = 1)

(0.2670.5790.096)/(0.2670.5790.096)+(0.2990.5900.904) = 0.014/0.014+0.159=0.0809

F. Comparison of P(Loan=1|CC=1,Online=1) *Upon comparing the above computed Naive Bayes Probability with the value obtained in B,it can be observed that the probability computed through Naive Bayes is slightly higher than the probability computed through frequency table.*

G. Naive Bayes Probablity:

```
library(e1071)
nb_model<-naiveBayes(Personal.Loan~Online+CreditCard,data=Training_Normalized_Data)
nb_model
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##     0     1
## 0.904 0.096
##
## Conditional probabilities:
##    Online
## Y           0         1
##   0 0.4092920 0.5907080
##   1 0.4201389 0.5798611
##
##    CreditCard
## Y           0         1
##   0 0.7005900 0.2994100
##   1 0.7326389 0.2673611
```

This value is equivalent to the probability obtained in E.