Assignment-3

Harika

2023-11-05

```
library(caret)
```

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

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

```
library(ggplot2)
library(ISLR)
library(class)
library(reshape2)
#install.packages("melt")
library(melt)
```

```
## Warning: package 'melt' was built under R version 4.3.2
```

1.DATA LOADING

```
File_path <- "C:/Users/Harika/Downloads/UniversalBank-1.csv"
```

```
###specified the file path above that to be loaded.
```

```
UniversalBank_1 <- read.csv("C:/Users/Harika/Downloads/UniversalBank-1.csv")
```

```
### Loaded the dataset above
```

```
UniversalBank_1$Personal.Loan <- as.factor(UniversalBank_1$Personal.Loan)
UniversalBank_1$Online <- as.factor(UniversalBank_1$Online)
UniversalBank_1$CreditCard <- as.factor(UniversalBank_1$CreditCard)
```

```
summary(UniversalBank_1)
```

```
##        ID            Age          Experience       Income          ZIP.Code
##  Min.   :   1   Min.   :23.00   Min.   :-3.0   Min.   :  8.00   Min.   : 9307
##  1st Qu.:1251   1st Qu.:35.00   1st Qu.:10.0   1st Qu.: 39.00   1st Qu.:91911
##  Median :2500   Median :45.00   Median :20.0   Median : 64.00   Median :93437
##  Mean   :2500   Mean   :45.34   Mean   :20.1   Mean   : 73.77   Mean   :93153
##  3rd Qu.:3750   3rd Qu.:55.00   3rd Qu.:30.0   3rd Qu.: 98.00   3rd Qu.:94608
##  Max.   :5000   Max.   :67.00   Max.   :43.0   Max.   :224.00   Max.   :96651
```

```
##      Family          CCAvg          Education        Mortgage     Personal.Loan
##  Min.   :1.000   Min.   : 0.000   Min.   :1.000   Min.   :  0.0   0:4520
##  1st Qu.:1.000   1st Qu.: 0.700   1st Qu.:1.000   1st Qu.:  0.0   1: 480
##  Median :2.000   Median : 1.500   Median :2.000   Median :  0.0
##  Mean   :2.396   Mean   : 1.938   Mean   :1.881   Mean   : 56.5
##  3rd Qu.:3.000   3rd Qu.: 2.500   3rd Qu.:3.000   3rd Qu.:101.0
##  Max.   :4.000   Max.   :10.000   Max.   :3.000   Max.   :635.0
##  Securities.Account   CD.Account     Online   CreditCard
##  Min.   :0.0000     Min.   :0.0000   0:2016   0:3530
##  1st Qu.:0.0000     1st Qu.:0.0000   1:2984   1:1470
##  Median :0.0000     Median :0.0000
##  Mean   :0.1044     Mean   :0.0604
##  3rd Qu.:0.0000     3rd Qu.:0.0000
##  Max.   :1.0000     Max.   :1.0000
```

```
###the above data represents summary for the given dataset.
```

2.DATA PARTITIONING (60:40)

```
set.seed(23)
```

```
###to set a seed for reproducibility.
```

```
SplitData <- createDataPartition(UniversalBank_1$Personal.Loan, p=0.60, list = FALSE)
train_data <- UniversalBank_1[SplitData,]
test_data <- UniversalBank_1[-SplitData,]
```

```
###the data is now splitted into training (60%) and testing (40%) sets above.
```

```
dim(train_data)
```

```
## [1] 3000   14
```

```
dim(test_data)
```

```
## [1] 2000   14
```

```
###to check the dimensions of the training and testing sets above.
```

A. TO CREATE A PIVOT TABLE

```
Pivot_Table1 <- ftable(train_data$CreditCard, train_data$Personal.Loan, train_data$Online)
Pivot_Table1
```

```
##         0    1
##
## 0 0   773 1127
##   1    82  114
## 1 0   315  497
##   1    39   53
```

```
###the pivot table created above.
```

B. Based on the pivot table created, we can determine the probability that this customer would accept the loan offer, when equals 53/(53+497) ~ 0.096.

C. Two separate pivot tables were created using the training data. Where one will have the internet (columns) are a function of the personal loan (rows), and whereas the credit card is a function of the other.

```
Melt_1 <- melt(train_data,id=c("Personal.Loan"),variable="Online")
```

```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```

```
Melt_2 <- melt(train_data,id=c("Personal.Loan"),variable="CreditCard")
```

```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```

```
cast1 = dcast(Melt_1, Personal.Loan~Online)
```

```
## Aggregation function missing: defaulting to length
```

```
cast2 = dcast(Melt_2, Personal.Loan~CreditCard)
```

```
## Aggregation function missing: defaulting to length
```

D. Calculating specified amounts P(A/B) signifies the probability which A will occur given B.

```
ftable(train_data$Personal.Loan, train_data$Online)
```

```
##       0    1
##
## 0  1088 1624
## 1   121  167
```

```
ftable(train_data$Personal.Loan, train_data$CreditCard)
```

```
##       0    1
##
## 0  1900  812
## 1   196   92
```

ftable(train_data[,10]) #10 is Personal.Loan column.

1. P(CC = 1 | Loan = 1) = (92/92+196) = 0.319
2. P(Online = 1 | Loan = 1) = (167/167+121) = 0.579
3. P(Loan = 1) = (288/288+2712) = 0.096

4. P(CC = 1 | Loan = 0) = (812/812+1900) = 0.299
5. P(Online = 1 | Loan = 0) = (1624/ 1624+1088) = 0.598
6. P(Loan = 0) = (2712/ 2712+288) = 0.904

E. Using quantities computed above to compute the naive bayes probability P(Loan=1|CC=1, Online=1).

(0.319*0.579*0.096)/(0.319*0.579*0.096)+(0.299*0.598*0.904) ~ 0.098

F.In question B, we calculated a probability value of 0.096, and in the previous question, we found a probability value of 0.098. These values have a slight difference only. As of question E, we took into account more dependent information, but it seems that the value from question B is more accurate and specific comparitively.

G.Implementing Naives Bayes below

```
#install.packages("naivebayes")
library(naivebayes)
```

```
## Warning: package 'naivebayes' was built under R version 4.3.2
```

```
## naivebayes 0.9.7 loaded
```

```
Naive_Bayes_Model <- naive_bayes(Personal.Loan~Online+CreditCard,data = train_data)
Naive_Bayes_Model
```

```
##
## ================================== Naive Bayes ==================================
##
##  Call:
## naive_bayes.formula(formula = Personal.Loan ~ Online + CreditCard,
##     data = train_data)
##
## ---------------------------------------------------------------------------------
##
## Laplace smoothing: 0
##
## ---------------------------------------------------------------------------------
##
##  A priori probabilities:
##
##     0     1
## 0.904 0.096
##
## ---------------------------------------------------------------------------------
##
##  Tables:
##
## ---------------------------------------------------------------------------------
##  ::: Online (Bernoulli)
## ---------------------------------------------------------------------------------
```

```
##
## Online         0         1
##      0 0.4011799 0.4201389
##      1 0.5988201 0.5798611
##
## ------------------------------------------------------------------------------------
##  ::: CreditCard (Bernoulli)
## ------------------------------------------------------------------------------------
##
## CreditCard         0         1
##          0 0.7005900 0.6805556
##          1 0.2994100 0.3194444
##
## ------------------------------------------------------------------------------------
```

```
###
```