FML_3

Bharadwaj

2023-11-05

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.2.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.3

## Loading required package: lattice
```

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.2.3
```

```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.2.3
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.2.3
```

```
#loading data set required
data_asign3 <- read.csv("C:/Users/CherRyY/Desktop/dataset/UniversalBank1.csv")
```

conversion of the variables into factor

```
data_asign3$Personal.Loan<-factor(data_asign3$Personal.Loan)
data_asign3$Online<-factor(data_asign3$Online)
data_asign3$CreditCard<-factor(data_asign3$CreditCard)
```

Splitting data into two sets for training and validation.

```
set.seed(1237)
training<-createDataPartition(data_asign3$Personal.Loan,p=0.6,list = FALSE)
training_setPart<-data_asign3[training,]
validation_setPart<-data_asign3[-training,]
nrow(training_setPart)
```

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

```r
nrow(validation_setPart)
```

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

Question-A

```r
table1<-xtabs(~CreditCard+Personal.Loan+Online,data=training_setPart)
ftable(table1)
```

```
##                          Online    0    1
## CreditCard Personal.Loan
## 0          0                     783 1133
##            1                      82  115
## 1          0                     319  477
##            1                      37   54
```

Question-B

```r
46/(46+460)
```

```
## [1] 0.09090909
```

Question-C

```r
table(Personal.Loan=training_setPart$Personal.Loan,
      Online=training_setPart$Online)
```

```
##              Online
## Personal.Loan    0    1
##             0 1102 1610
##             1  119  169
```

```r
table(Personal.Loan=training_setPart$Personal.Loan,
      CreditCard=training_setPart$CreditCard)
```

```
##              CreditCard
## Personal.Loan    0    1
##             0 1916  796
##             1  197   91
```

```r
table(Personal.Loan=training_setPart$Personal.Loan)
```

```
## Personal.Loan
##    0    1
## 2712  288
```

Question-D

```r
#i.P(CC = 1 | Loan = 1)
P1=80/(80+208)
P1
```

```
## [1] 0.2777778
```

```r
#ii. P(Online = 1 | Loan = 1)
P2=179/(179+109)
P2
```

```
## [1] 0.6215278
```

```r
#iii. P(Loan = 1)
P3=288/(288+2712)
P3
```

```
## [1] 0.096
```

```r
#iv. P(CC = 1 | Loan = 0)
P4=779/(779+1933)
P4
```

```
## [1] 0.2872419
```

```r
#v. P(Online = 1 | Loan = 0)
P5=1599/(1599+1113)
P5
```

```
## [1] 0.5896018
```

```r
#vi. P(Loan = 0)
P6=2712/(288+2712)
P6
```

```
## [1] 0.904
```

Question-E
calculating the naive Bayes probability for P(Loan = 1 | CC = 1, Online = 1).

```r
(P1*P2*P3)/((P1*P2*P3)+(P4*P5*P6))
```

```
## [1] 0.09768187
```

Question-F

The probability from pivot table is 0.1005587 and the calculated naive Bayes probability is 0.1120411. Naive Bayes works on the assumption that attributes are independent of each other. this suggests that probability from the pivot table is more accurate.

Question-G

```r
Naivebayes_model<-naiveBayes(Personal.Loan~CreditCard+Online,data = training_setPart)
testing<-data.frame(CreditCard=1,Online=1)
testing$CreditCard<-factor(testing$CreditCard)
testing$Online<-factor(testing$Online)
predict(Naivebayes_model,testing,type = 'raw')
```

```
##              0         1
## [1,] 0.8984709 0.1015291
```

The probability of the test data coincides with the probability calculated in question E, which is 0.09768187. This suggests that the Naive Bayes algorithm has made an identical prediction to the calculated probability.