FML_3

Randheer Gonuguntla

2023-10-15

```r
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```r
library(e1071)
library(ISLR)
library(reshape2)
#loading data set required
data_ub<-read.csv("C:/Users/sidda/Downloads/UniversalBank.csv")
```

converting the variables into factor

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

partitioning of data into two training and validation datasets

```r
set.seed(6354)
train<-createDataPartition(data_ub$Personal.Loan,p=0.6,list = FALSE)
train_set<-data_ub[train,]
validation_set<-data_ub[-train,]
nrow(train_set)
```

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

```r
nrow(validation_set)
```

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

Question-A

```r
table_1<-xtabs(~CreditCard+Personal.Loan+Online,data=train_set)
ftable(table_1)
```

```
##                            Online    0    1
## CreditCard Personal.Loan                  
## 0          0                      778 1135
##            1                       71  124
## 1          0                      314  485
##            1                       45   48
```

Question-B

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

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

Question-C

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

```
##              Online
## Personal.Loan    0    1
##             0 1092 1620
##             1  116  172
```

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

```
##              CreditCard
## Personal.Loan    0    1
##             0 1913  799
##             1  195   93
```

```r
table(Personal.Loan=train_set$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 = train_set)
test<-data.frame(CreditCard=1,Online=1)
test$CreditCard<-factor(test$CreditCard)
test$Online<-factor(test$Online)
predict(Naivebayes_model,test,type = 'raw')
```

```
##              0         1
## [1,] 0.8957591 0.1042409
```

Probability of test data is same as the probability obtained in the question E which is equal to 0.09768187. This implies that Naive bayes algorithm has predicted same as that of calculated probability