

# Machine learning\_Assignment\_3

Manasa Chelukala

3/5/2022

#Importing the Dataset

```
UniversalBank <- read.csv('C:/Users/HP/Documents/csv file/UniversalBank.csv')  
summary(UniversalBank)
```

```
##           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  
## Min.      :1.000      Min.     : 0.000      Min.     :1.000      Min.     : 0.0  
## 1st Qu.:1.000      1st Qu.: 0.700      1st Qu.:1.000      1st Qu.: 0.0  
## 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  
## Personal.Loan      Securities.Account      CD.Account      Online  
## Min.      :0.000      Min.     :0.0000      Min.     :0.0000      Min.     :0.0000  
## 1st Qu.:0.000      1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.0000  
## Median :0.000      Median :0.0000      Median :0.0000      Median :1.0000  
## Mean    :0.096      Mean    :0.1044      Mean    :0.0604      Mean    :0.5968  
## 3rd Qu.:0.000      3rd Qu.:0.0000      3rd Qu.:0.0000      3rd Qu.:1.0000  
## Max.     :1.000      Max.     :1.0000      Max.     :1.0000      Max.     :1.0000  
##           CreditCard  
## Min.      :0.000  
## 1st Qu.:0.000  
## Median :0.000  
## Mean    :0.294  
## 3rd Qu.:1.000  
## Max.     :1.000
```

#calling Libraries

```
library(caret)
```

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

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

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

```
library(class)
library(ISLR)
```

```
#Converting Personal.Loan Variable
```

```
UniversalBank$Personal.Loan = as.factor(UniversalBank$Personal.Loan)
```

```
summary(UniversalBank)
```

```
##      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    Min.   :0.0000    Min.   :0.000
## 1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.000
## Median :0.0000    Median :0.0000    Median :1.0000    Median :0.000
## Mean   :0.1044    Mean   :0.0604    Mean   :0.5968    Mean   :0.294
## 3rd Qu.:0.0000    3rd Qu.:0.0000    3rd Qu.:1.0000    3rd Qu.:1.000
## Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.000
```

```
#Converting Online Variable
```

```
UniversalBank$Online = as.factor(UniversalBank$Online)
```

```
summary(UniversalBank)
```

```
##      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 Min. :0.000
## 1st Qu.:0.0000 1st Qu.:0.0000 1:2984 1st Qu.:0.000
## Median :0.0000 Median :0.0000 Median :0.000
## Mean :0.1044 Mean :0.0604 Mean :0.294
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.000
## Max. :1.0000 Max. :1.0000 Max. :1.000
```

#Converting Creditcard Variable

```
UniversalBank$CreditCard = as.factor(UniversalBank$CreditCard)
summary(UniversalBank)
```

```
## 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
```

#Task\_A #Data Partition

```
set.seed(64060)
Train_Index = createDataPartition(UniversalBank$Personal.Loan,p=0.6, list=FALSE) # 60% reserved for Tra
Train.df=UniversalBank[Train_Index,]
Validation.df=UniversalBank[-Train_Index,]
```

```
mytable <- xtabs(~ CreditCard+Online+Personal.Loan, data=Train.df)
ftable(mytable)
```

```
## Personal.Loan 0 1
## CreditCard Online
## 0 0 772 75
## 1 1152 120
## 1 0 309 34
## 1 1 479 59
```

#Task\_B: what is the probability that this customer will accept the loan offer? [This is the probability of loan acceptance (Loan = 1) conditional on having a bank credit card (CC = 1) and being an active user of online banking services (Online = 1)]

```
Probability = 59/(479+59)
Probability
```

```
## [1] 0.1096654
```

#Task\_C:

#pivot table with Personal loan as row and credit card as column using training data.

```
table(CreditCard=Train.df$CreditCard, Personal.Loan=Train.df$Personal.Loan)
```

```
##           Personal.Loan
## CreditCard    0      1
##           0 1924  195
##           1   788   93
```

#pivot table with Personal loan as row and Online as column using training data.

```
table(Online=Train.df$Online, Personal.Loan=Train.df$Personal.Loan)
```

```
##           Personal.Loan
## Online    0      1
##           0 1081  109
##           1 1631  179
```

#pivot table for Personal loan

```
table(Personal.Loan=Train.df$Personal.Loan)
```

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

#Task\_D: #i.P(CC = 1 | Loan = 1)(the proportion of credit card holders among the loan acceptors)

```
Probability_1 = 93/(195+93)
Probability_1
```

```
## [1] 0.3229167
```

#ii.P(Online = 1 | Loan = 1)

```
Probability_2 = 179/(109+179)
Probability_2
```

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

#iii.P(Loan = 1) (the proportion of loan acceptors)

```
Probability_3 = 288/(2712+288)
Probability_3
```

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

```
#iv.P(CC = 1 | Loan = 0)
```

```
Probability_4 = 788/(1924+788)
Probability_4
```

```
## [1] 0.2905605
```

```
#v.P(Online = 1 | Loan = 0)
```

```
Probability_5 = 1631/(1631+1081)
Probability_5
```

```
## [1] 0.6014012
```

```
#vi.P(Loan = 0)
```

```
Probability_6 = 2712/(2712+288)
Probability_6
```

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

```
#Task_E: #P(Loan = 1 | CC = 1, Online = 1).
```

```
naive_Bayes_probability <- (Probability_1*Probability_2*Probability_3) /
  ((Probability_1*Probability_2*Probability_3) +
   (Probability_4*Probability_5*Probability_6))
naive_Bayes_probability
```

```
## [1] 0.1087106
```

```
#Task_F: Compare this value with the one obtained from the pivot table in (B). #Which is a more accurate estimate?
```

```
#0.1087106 in task-E is very similar to the 0.1096654 in task-B.
#The difference between the exact and naive bayes methods is that
#the exact approach requires the same independent variable classifications to predict,
#whereas the naive bayes method does not.
```

```
#Task_G: # P(Loan = 1 | CC = 1, Online = 1) #Run naive Bayes on the data. Examine the model output on training data, and find the entry #that corresponds to P(Loan = 1 | CC = 1, Online = 1). Compare this to the number you #obtained in (E).
```

```
library(e1071)
```

```
nb.model<-naiveBayes (Personal.Loan~Online+CreditCard, data=Train.df)
To_Predict=data.frame(Online= '1', CreditCard= '1')
predict(nb.model,To_Predict,type='raw')
```

```
##           0           1
## [1,] 0.8912894 0.1087106
```

#The task-G value of 0.1087106 and the task-E value of 0.1087106 are identical. #As a result, the naive bayes produces the same results as the prior approaches.