

Machine learning Assingment 3

```

```
#Assignment 3 Machine Learning
setwd("C:/Users/nikes/Downloads/Machine Learning Assingment/Assingment 3")
library(readr)
UniversalBank <- read_csv("UniversalBank.csv")
colnames(UniversalBank) <- c('ID', 'Age', 'Experience', 'Income', 'ZIP_Code', 'Family',
'CCAvg', 'Education', 'Mortgage', 'Personal_Loan', 'Securities_Account', 'CD_Account',
'Online', 'CreditCard')
summary(UniversalBank)
library(caret)
library(dplyr)
library(class)
library(reshape2)
library(ISLR)
library(e1071)
```

### A. Pivot table for Universal Bank

```
UniversalBankPersonalLoan = as.factor(UniversalBankPersonalLoan)
UniversalBankOnline = as.factor(UniversalBankOnline)
UniversalBankCreditCard = as.factor(UniversalBankCreditCard)
set.seed(123)
train.index <- sample(row.names(UniversalBank), 0.6*dim(UniversalBank)[1])
test.index <- setdiff(row.names(UniversalBank), train.index)
train.df <- UniversalBank[train.index,]
test.df <- UniversalBank[test.index,]
train <- UniversalBank[train.index,]
```

```
test = UniversalBank[train.index,]
melted.UniversalBank = melt(train, id=c("CreditCard", "Personal_Loan"), variable= "Online")
recast.UniversalBank= dcast(melted.UniversalBank, Personal_Loan+CreditCard ~ Online)
recast.UniversalBank[,c(1:2,14)]
```

## B. Probability of the customer accepting loan offer

Probability =  $91 / 3000 = 0.30$

## C. Separate pivot table

```
table(Personal_Loan=trainPersonalLoan, Online = trainOnline)
table(Personal_Loan=trainPersonalLoan, CreditCard = trainCreditCard)
```

## D. [P(A | B) means "the probability of A given B"]

- i. Proportion of credit card holders among the loan acceptors =  $91/278 = 0.32$
- ii.  $P(\text{Online} = 1 \mid \text{Loan} = 1) = 179 / 278 = 0.64$
- iii.  $P(\text{Loan} = 1)$  (the proportion of loan acceptors) =  $278 / 2722 = 0.10$
- iv.  $P(\text{CC} = 1 \mid \text{Loan} = 0) = 792 / 2722 = 0.29$
- v.  $P(\text{Online} = 1 \mid \text{Loan} = 0) = 1620 / 2722 = 0.59$
- vi.  $P(\text{Loan} = 0) = 2722 / 3000 = 0.90$

## E. naive Bayes probability $P(\text{Loan} = 1 \mid \text{CC} = 1, \text{Online} = 1)$ .

$= (0.32 * 0.64 * 0.1) / (0.32 * 0.64 * 0.1 + 0.29 * 0.59 * 0.9)$   
 $= 0.11$

## F. Comparing value with the one obtained from the pivot table in (B).

Pivot table Probability =  $(278/3000) = 0.092$

Naive Bayes Probability =  $0.32 * 0.59 * 0.1 / (0.32 * 0.64 * 0.1 + 0.29 * 0.59 * 0.9) = 0.11$

##As using the naive bayes the main assumption we are making is all variable are independent and have equal importance, so we can see naive bayes probability is little

higher, the accuracy of naive bayes probability may be less accurate considering the features that all variables are independent and are not correlated with each other.

### **G. Running naive Bayes on the data. $P(\text{Loan} = 1 \mid \text{CC} = 1, \text{Online} = 1)$**

```
naive.train = train.df[,c(10,13:14)]
```

```
naive.test = test.df[,c(10,13:14)]
```

```
naivebayes = naiveBayes(Personal_Loan~.,data=naive.train)
```

```
naivebayes
```

**We can see and analyze the result the prior probability is exactly 0.092 as shown here which is exactly the same as we computed above.**