Assignment 3

SAI ROHITH GUDA

10-14-2023

###Task a

```
# Load necessary libraries
library(reshape2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load necessary library
library(dplyr)

# Read the CSV file
data <- read.csv("UniversalBank.csv")

# Select the relevant columns
selected_data <- data %>% select(Online, CreditCard, Personal.Loan)

# Set seed for reproducibility
set.seed(42)

# Split data into training and validation sets
sample_size <- floor(0.6 * nrow(selected_data))
train_index <- sample(seq_len(nrow(selected_data)), size = sample_size)

train_data <- selected_data[train_index, ]
validation_data <- selected_data[-train_index, ]

pivot_table <- table(train_data$CreditCard, train_data$Online, train_data$Personal.Loan)
print(paste("Pivot table for CreditCard, Online, and Personal.Loan:\n", pivot_table))
```

```
## [1] "Pivot table for CreditCard, Online, and Personal.Loan:\n 803"
```

```
## [2] "Pivot table for CreditCard, Online, and Personal.Loan:\n 312"
## [3] "Pivot table for CreditCard, Online, and Personal.Loan:\n 1125"
## [4] "Pivot table for CreditCard, Online, and Personal.Loan:\n 472"
## [5] "Pivot table for CreditCard, Online, and Personal.Loan:\n 77"
## [6] "Pivot table for CreditCard, Online, and Personal.Loan:\n 32"
## [7] "Pivot table for CreditCard, Online, and Personal.Loan:\n 131"
## [8] "Pivot table for CreditCard, Online, and Personal.Loan:\n 48"
```

Task b

```
prob_acceptance <- pivot_table[2, 2, 2] / (pivot_table[2, 2, 1] + pivot_table[2, 2, 2])
print(paste("Probability of loan acceptance given CC=1 and Online=1:", round(prob_acceptance, 4)))
```

```
## [1] "Probability of loan acceptance given CC=1 and Online=1: 0.0923"
```

Task c

```
pivot_loan_online <- table(train_data$Personal.Loan, train_data$Online)

pivot_loan_cc <- table(train_data$Personal.Loan, train_data$CreditCard)

print("Pivot table for Personal.Loan and Online:")
```

```
## [1] "Pivot table for Personal.Loan and Online:"
```

```
print(pivot_loan_online)
```

```
##
##        0    1
##   0 1115 1597
##   1  109  179
```

```
print("\nPivot table for Personal.Loan and CreditCard:")
```

```
## [1] "\nPivot table for Personal.Loan and CreditCard:"
```

```
print(pivot_loan_cc)
```

```
##
##        0    1
##   0 1928  784
##   1  208   80
```

Task d

```r
# P(CC = 1 | Loan = 1)
p_cc_given_loan1 <- pivot_loan_cc[2, 2] / (pivot_loan_cc[2, 1] + pivot_loan_cc[2, 2])

# P(Online = 1 | Loan = 1)
p_online_given_loan1 <- pivot_loan_online[2, 2] / (pivot_loan_online[2, 1] + pivot_loan_online[2, 2])

# P(Loan = 1)
p_loan1 <- sum(train_data$Personal.Loan == 1) / nrow(train_data)

# P(CC = 1 | Loan = 0)
p_cc_given_loan0 <- pivot_loan_cc[1, 2] / (pivot_loan_cc[1, 1] + pivot_loan_cc[1, 2])

# P(Online = 1 | Loan = 0)
p_online_given_loan0 <- pivot_loan_online[1, 2] / (pivot_loan_online[1, 1] + pivot_loan_online[1, 2])

# P(Loan = 0)
p_loan0 <- sum(train_data$Personal.Loan == 0) / nrow(train_data)

print(paste("P(CC = 1 | Loan = 1):", round(p_cc_given_loan1, 4)))
```

```
## [1] "P(CC = 1 | Loan = 1): 0.2778"
```

```r
print(paste("P(Online = 1 | Loan = 1):", round(p_online_given_loan1, 4)))
```

```
## [1] "P(Online = 1 | Loan = 1): 0.6215"
```

```r
print(paste("P(Loan = 1):", round(p_loan1, 4)))
```

```
## [1] "P(Loan = 1): 0.096"
```

```r
print(paste("P(CC = 1 | Loan = 0):", round(p_cc_given_loan0, 4)))
```

```
## [1] "P(CC = 1 | Loan = 0): 0.2891"
```

```r
print(paste("P(Online = 1 | Loan = 0):", round(p_online_given_loan0, 4)))
```

```
## [1] "P(Online = 1 | Loan = 0): 0.5889"
```

```r
print(paste("P(Loan = 0):", round(p_loan0, 4)))
```

```
## [1] "P(Loan = 0): 0.904"
```

Task e

```r
p_cc <- sum(train_data$CreditCard == 1) / nrow(train_data)
p_online <- sum(train_data$Online == 1) / nrow(train_data)

# Naive Bayes Probability
p_naive_bayes <- (p_cc_given_loan1 * p_online_given_loan1 * p_loan1) / (p_cc * p_online)
print(paste("Naive Bayes Probability P(Loan = 1 | CC = 1, Online = 1):", round(p_naive_bayes, 4)))
```

```
## [1] "Naive Bayes Probability P(Loan = 1 | CC = 1, Online = 1): 0.0972"
```

Task f

```
comparison <- ifelse(abs(prob_acceptance - p_naive_bayes) < 0.01, "similar", "different")
print(paste("The value from the pivot table in (b) is", round(prob_acceptance, 4),
            "and the value from the naive Bayes estimate in (e) is", round(p_naive_bayes, 4),
            ". The two values are", comparison, "."))
```

```
## [1] "The value from the pivot table in (b) is 0.0923 and the value from the naive Bayes estimate in
```

Task g

```
#install.packages("e1071")
library(e1071)

model <- naiveBayes(Personal.Loan ~ CreditCard + Online, data=train_data)
predicted_probs <- predict(model, newdata=data.frame(CreditCard=1, Online=1), type="raw")
p_naive_bayes_R <- predicted_probs[2]
print(paste("Predicted Probability using naiveBayes for P(Loan = 1 | CC = 1, Online = 1):", round(p_nai
```

```
## [1] "Predicted Probability using naiveBayes for P(Loan = 1 | CC = 1, Online = 1): 0.0962"
```