HW\_2

2023-02-18

# Read the data from a CSV file  
UniversalBank <- read.csv("/Users/binsalim/Downloads/UniversalBank.csv")  
UniBank <- UniversalBank  
  
# Summary of Education variable  
summary(UniBank$Education)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 2.000 1.881 3.000 3.000

# Create dummy variables for Education variable  
UniBank <- mutate(UniBank, Education\_1 = ifelse(Education == 1, 1, 0))  
UniBank <- mutate(UniBank, Education\_2 = ifelse(Education == 2, 1, 0))  
UniBank <- mutate(UniBank, Education\_3 = ifelse(Education == 3, 1, 0))

# Convert Personal Loan variable to character type  
UniBank$Personal.Loan <- as.character(UniBank$Personal.Loan)  
  
# Select the relevant columns and remove missing values  
df <- UniBank[, -c(1, 4, 6)]  
df <- na.omit(df)

# Normalize the data using centering and scaling  
norm\_model <- preProcess(df, method = c("center", "scale"))  
Default\_normalized <- predict(norm\_model, df)  
  
# Split the data into training and testing sets  
Index\_Train <- createDataPartition(Default\_normalized$Personal.Loan, p = 0.6, list = FALSE)  
Train <- Default\_normalized[Index\_Train, ]  
Test <- Default\_normalized[-Index\_Train, ]  
  
# Separate predictors and labels in the training and testing sets  
Train\_Predictors <- Train[, -7]  
Test\_Predictors <- Test[, -7]  
Train\_labels <- Train[, 7]  
Test\_labels <- Test[, 7]

# Classify the data using k-NN with k=1  
Predicted\_Test\_labels <- knn(Train\_Predictors,  
 Test\_Predictors,  
 cl = Train\_labels,  
 k = 1)  
  
# Print the first few predicted labels  
head(Predicted\_Test\_labels)

## [1] 0 0 0 0 1 1  
## Levels: 0 1

# Create a confusion matrix to evaluate the model's performance  
CrossTable(x = Test\_labels, y = Predicted\_Test\_labels, prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2000   
##   
##   
## | Predicted\_Test\_labels   
## Test\_labels | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 1730 | 78 | 1808 |   
## | 0.957 | 0.043 | 0.904 |   
## | 0.946 | 0.456 | |   
## | 0.865 | 0.039 | |   
## -------------|-----------|-----------|-----------|  
## 1 | 99 | 93 | 192 |   
## | 0.516 | 0.484 | 0.096 |   
## | 0.054 | 0.544 | |   
## | 0.050 | 0.046 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 1829 | 171 | 2000 |   
## | 0.914 | 0.086 | |   
## -------------|-----------|-----------|-----------|  
##   
##

# Define the new customer data frame  
record <- data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education\_1 = 0, Education\_2 = 1, Education\_3 = 0, Mortgage = 0, Securities.Account = 0, CD.Account = 0, Online = 1, Credit.Card = 1)  
  
# Remove the "Personal Loan" column from the record  
#record <- record[, -7]  
record

## Age Experience Income Family CCAvg Education\_1 Education\_2 Education\_3  
## 1 40 10 84 2 2 0 1 0  
## Mortgage Securities.Account CD.Account Online Credit.Card  
## 1 0 0 0 1 1

# Classify the new customer using the trained model  
Predicted\_Test\_labels <- knn(Train\_Predictors,  
 record,  
 cl = Train\_labels,  
 k = 1)  
# The customer will not be given a loan   
Predicted\_Test\_labels

## [1] 0  
## Levels: 0 1