MIS 545 Lab 08 Naïve Bayes

Code:

```
# Amandeep Kaur, Varun Kapuria
# MIS 545 Section 01
# Lab08Group15KaurKapuria.R
# The following code loads the DwellingType csv into a tibble called
# dwellingType and generates Naive Baiyes Model to predict the dwelling type.

# Installing the tidyverse and e1071 packages
# install.packages("tidyverse")
# install.packages("e1071")

# Loading the tidyverse and e1071 libraries
library(tidyverse)
library(e1071)

# Settting the working directory to Lab08 folder
setwd("C:/Users/Apple Kaur/Desktop/FALL1 2022/MIS545 Data
Mining/Rlab08Naives")
getwd()

# Reading DwellingType.csv into a tibble called dwellingType
dwellingType <- read_csv("DwellingType.csv",
                         col_types = "filll",
                         col_names = TRUE)

# Displaying dwellingType in the console
print(dwellingType)

# Displaying the structure of dwellingType in the console
print(str(dwellingType))

# Displaying the summary of dwellingType in the console
print(summary(dwellingType))

# Randomly splitting the dataset into dwellingTypeTraining (75% of records)
# and dwellingTypeTesting (25% of records) using 154 as the random seed
set.seed(154)
sampleSet <- sample(nrow(dwellingType),
                    round(nrow(dwellingType) * 0.75),
                    replace = FALSE)
dwellingTypeTraining <- dwellingType[sampleSet, ]
dwellingTypeTesting <- dwellingType[-sampleSet, ]

# Generating the Naive Bayes model to predict DwellingType based on the
# other variables in the dataset
dwellingTypeModel <- naiveBayes(formula = DwellingType ~ . ,
                                data = dwellingTypeTraining,
```

MIS 545 Lab 08 Naïve Bayes

```
                                laplace = 1)

# Building probabilities for each record in the testing dataset and
# storing them in dwellingTypeProbability
dwellingTypeProbability <- predict(dwellingTypeModel,
                                   dwellingTypeTesting,
                                   type = "raw")

# Displaying dwellingTypeProbability on the console
print(dwellingTypeProbability)

# Predicting classes for each record in the testing dataset and storing
# them in dwellingTypePrediction
dwellingTypePrediction <- predict(dwellingTypeModel,
                                  dwellingTypeTesting,
                                  type = "class")

# Displaying dwellingTypePrediction on the console
print(dwellingTypePrediction)

# Evaluating the model by forming a confusion matrix
dwellingTypeconfusionMatrix <- table(dwellingTypeTesting$DwellingType,
                                     dwellingTypePrediction)

# Displaying the confusion matrix on the console
print(dwellingTypeconfusionMatrix)

# Calculating the model predictive accuracy and store it into a
# variable called predictiveAccuracy
predictiveAccuracy <- sum(diag(dwellingTypeconfusionMatrix))/
  nrow(dwellingTypeTesting)

# Displaying the predictive accuracy on the console
print(predictiveAccuracy)
```

MIS 545 Lab 08 Naïve Bayes

Questions and Answers:

1. Provide a possible explanation as to why the model was better able to predict home and apartment dwellers than condo dwellers?
    - Home dwellers had ~50% more tuples than Condo dwellers, which is why it was able to accurately predict Home dwellers. We believe that the tuples with dwellingType “Condo” too closely resembled the dwellingType “Home” or “Apartment”, and since both of the latter had larger number of tuples, Naïve Bayes classified them as either Home or Apartment. We would need a sizable amount of data for the model to increase accuracy.

2. How could this model be used by a direct mailing marketing company to optimally target their mailings?
    - The direct mailing companies can use this data to send personalized mailings to specific sets of people. For example, a 41 male and not living alone is more likely to buy an Apartment or a Home rather than a Condo.