MIS 545 Lab 09

Code:

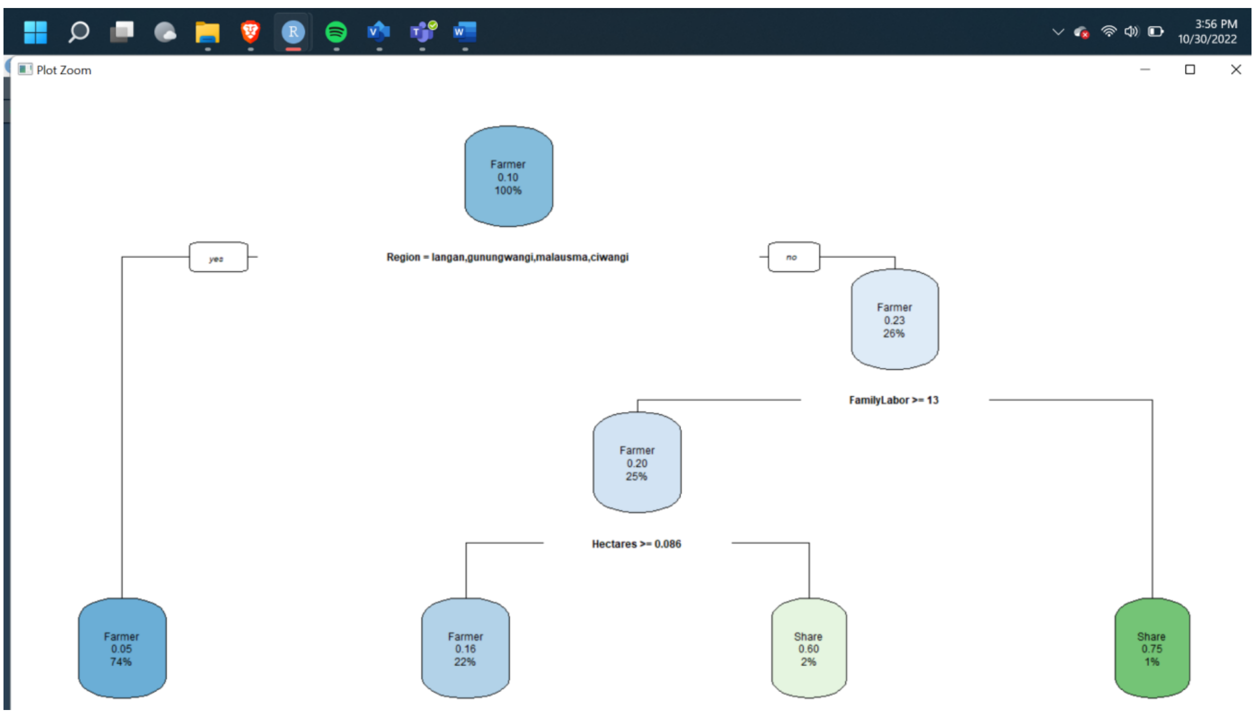
```
# Medha Singh, Varun Kapuria
# MIS 545 Section 01
# Lab09Group12SinghKapuria.R
# The following code predicts a farms ownership status (farmer owned or
# sharecropped) based on independent variables of hectares, seed price,
# hired labor, family labor, avg wage, and region.

# Installing the tidyverse and rpart.plot packages
# install.packages("tidyverse")
# install.packages("rpart.plot")

# Loading the tidyverse, rpart, and rpart.plot libraries
library(tidyverse)
library(rpart)
library(rpart.plot)

# Setting the working directory to your Lab09 folder
setwd("C:/Users/ual-laptop/Desktop/MIS545/Lab09")

# Reading IndonesianRiceFarms.csv into a tibble called riceFarms
riceFarms <- read_csv("IndonesianRiceFarms.csv",
                      col_names = TRUE,
                      col_types = "fniiinf")

# Displaying riceFarms in the console
print(riceFarms)

# Displaying the structure of riceFarms in the console
str(riceFarms)

# Displaying the summary of riceFarms in the console
summary(riceFarms)

# Randomly splitting the dataset into riceFarmsTraining (75% of records)
# and riceFarmsTesting (25% of records) using 370 as the random seed
set.seed(370)
sampleSet <- sample(nrow(riceFarms),
                    round(nrow(riceFarms) * 0.75),
                    replace = FALSE)
riceFarmsTraining <- riceFarms[sampleSet, ]
riceFarmsTesting <- riceFarms[-sampleSet, ]

# Generating the decision tree model to predict FarmOwnership based on the
```

```r
# other variables in the dataset. Use 0.01 as the complexity parameter.
riceFarmsDecisionTreeModel <- rpart(formula = FarmOwnership ~ .,
                                    method = "class",
                                    cp = 0.01,
                                    data = riceFarmsTraining)

# Displaying the decision tree visualization in R
rpart.plot(riceFarmsDecisionTreeModel)

# Predicting classes for each record in the testing dataset and
# storing them in riceFarmsPrediction
riceFarmPredictions <- predict(riceFarmsDecisionTreeModel,
                               riceFarmsTesting,
                               type = "class")

# Displaying riceFarmsPrediction on the console
print(riceFarmPredictions)

# Evaluating the model by forming a confusion matrix
riceFarmsConfusionMatrix <-  table(riceFarmsTesting$FarmOwnership,
                                   riceFarmPredictions)

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

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

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

# Creating a new decision tree model using 0.007 as the complexity parameter
riceFarmsDecisionTreeModel2 <- rpart(formula = FarmOwnership ~ .,
                                    method = "class",
                                    cp = 0.007,
                                    data = riceFarmsTraining)

# displaying the new decision tree visualization
rpart.plot(riceFarmsDecisionTreeModel2)

# predicting classes for new decision tree
riceFarmPredictions2 <- predict(riceFarmsDecisionTreeModel2,
                               riceFarmsTesting,
```

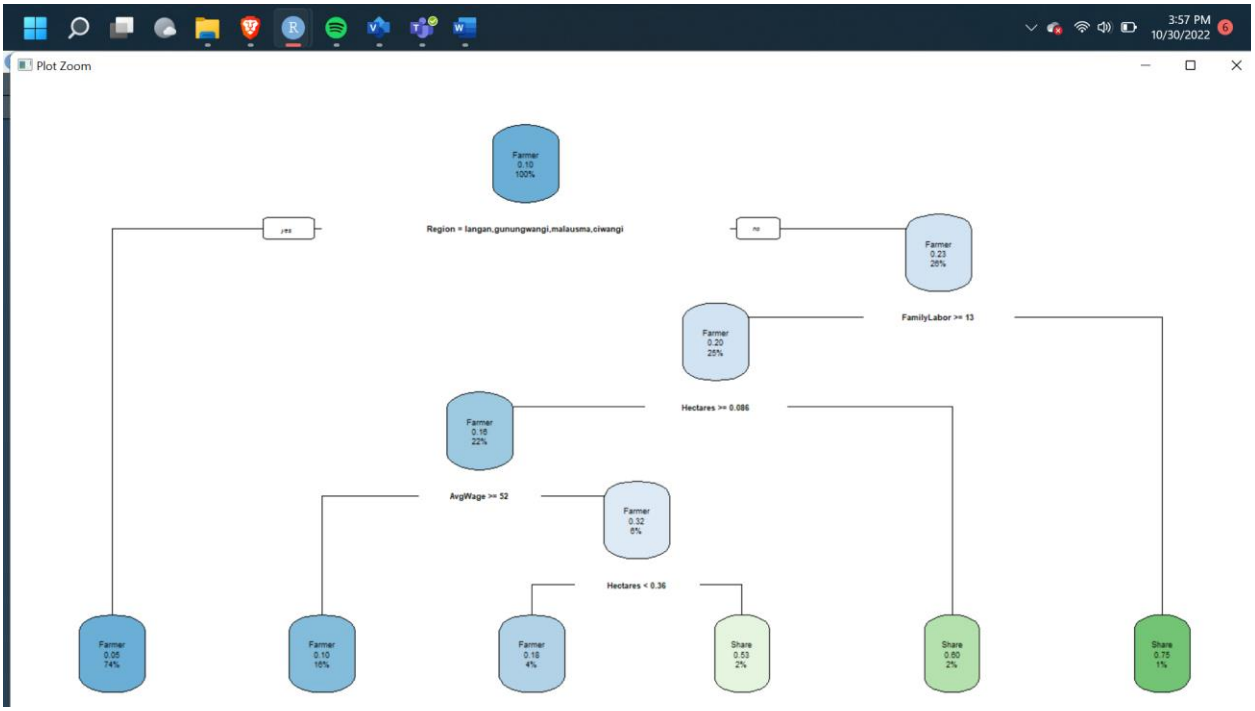
MIS 545 Lab 09

```
                                type = "class")

# Creating new predictions' confusion matrix for riceFarmsPredictions2
riceFarmsConfusionMatrix2 <-  table(riceFarmsTesting$FarmOwnership,
                                   riceFarmPredictions2)

# Calculating new predictive accuracy for riceFarmsPredictions2
predictiveAccuracy2 <- sum(diag(riceFarmsConfusionMatrix2))/
  nrow(riceFarmsTesting)

# Displaying the new predictive accuracy
print(predictiveAccuracy2)
```

Screenshots:



MIS 545 Lab 09



Question and answers:

Did increasing the complexity of the decision tree improve the model's predictive accuracy? Why do you think this is the case?

Increasing the complexity resulted in a minor decrease in predictive accuracy, this might be because decision trees are prone to overfitting.