R script code

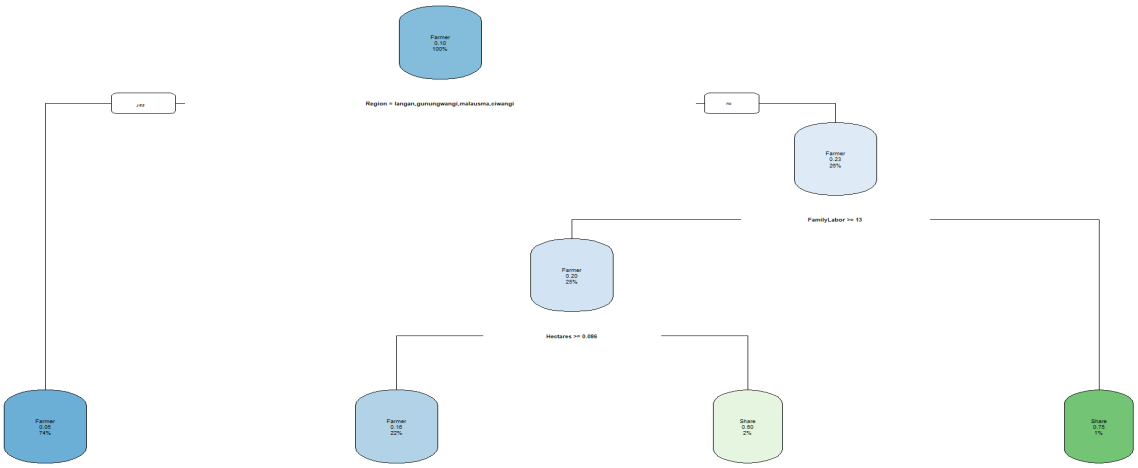
```
# Vihari Reddy Tummuru and LNU Abhinav
# MIS 545 Section 01
# Lab09Group24AbhinavTummuru.R
# This code demonstrates the summary of the IndonesianRiceFarms
# available in the form of a csv file. Performs data analysis for various
# features and displays DecisionTree visualizations.

# Environment Setup ------------------------------------------------------
# Installing rpart.plot packages
# install.packages("tidyverse")
# install.packages("rpart.plot")

# Loading tidyverse and rpart.plot packages
library(rpart.plot)
library(tidyverse)

# Set the working directory to your Lab09 Folder
setwd("C:/Users/ual-laptop/Documents/MIS545/Lab09")

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

# Displaying structure of riceFarms tibble
print(str(riceFarms))
```

```
# Displaying summary of riceFarms tibble
print(summary(riceFarms))

# Set seed using 154 as random seed
set.seed(370)

# Split the dataset into riceFarmsTraining (75% of records) and
# riceFarmsTesting (25% of records)
sampleSet <- sample(nrow(riceFarms),
                    round(nrow(riceFarms)*.75),
                    replace = FALSE)

# Set riceFarmsTraining (75% of records)
riceFarmsTraining <- riceFarms[sampleSet,]

# Set riceFarmsTesting (25% of records)
riceFarmsTesting <- riceFarms[-sampleSet,]

# Generating Decision Tree model for riceFarmsTraining tibble
riceFarmDecisionTreeModel <- rpart(formula = FarmOwnership ~ .,
                                   method = "class",
                                   cp = 0.01,
                                   data = riceFarmsTraining)

# Generating Rpart.plot for the decision tree model using default cp
rpart.plot(riceFarmDecisionTreeModel)

# Generating prediction based on the riceFarmDecisionTreeModel model
riceFarmPrediction <- predict(riceFarmDecisionTreeModel,
```

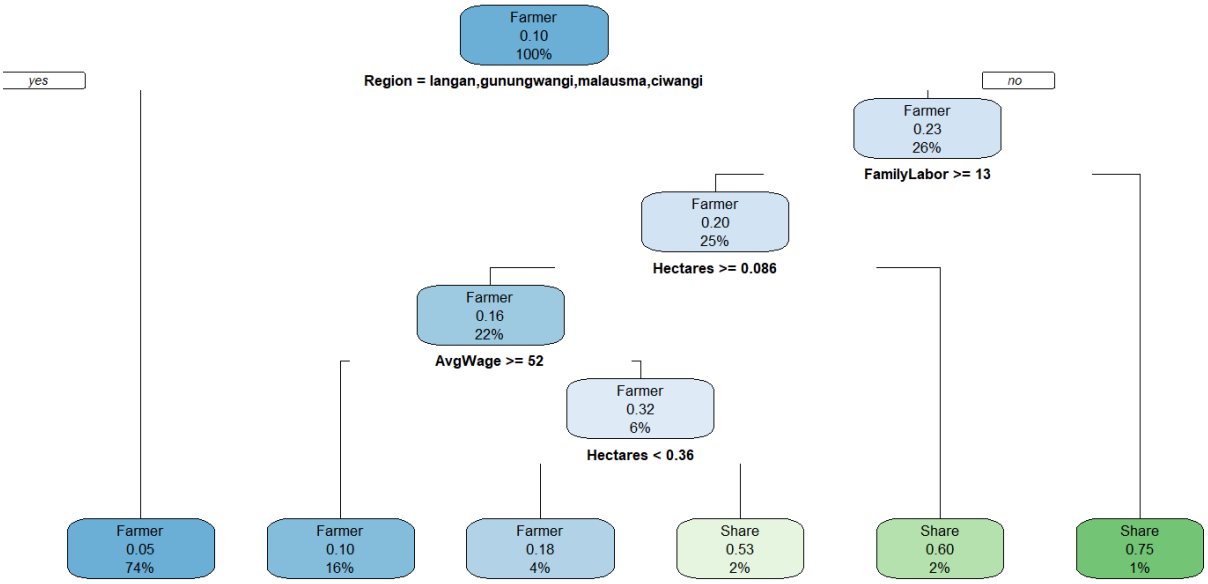
```
                                riceFarmsTesting,
                                type = 'class')

# Displaying the prediction data riceFarmPrediction on console
print(riceFarmPrediction)

# Generating confusion matrix to evauate te model
riceFramConfusionMatrix <- table(riceFarmsTesting$FarmOwnership,
                                 riceFarmPrediction)

# Displaying confusion matrix on the console
print(riceFramConfusionMatrix)

# Generating predictiveAccuracy of the model and storing it in a varaiable
# called predictiveAccuracy
predictiveAccuracy <- sum(diag(riceFramConfusionMatrix)) /
  nrow(riceFarmsTesting)

# Displaying predictiveAccuracy of the model on console
print(predictiveAccuracy)

# Generating Decision Tree model for tibble riceFarmsTraining using cp =
0.007
riceFarmDecisionTreeModelNew <- rpart(formula = FarmOwnership ~ .,
                                      method = "class",
                                      cp = 0.007,
                                      data = riceFarmsTraining)

# Generating Rpart.plot for the model
rpart.plot(riceFarmDecisionTreeModelNew)
```

```
# Generating prediction based on model
riceFarmPredictionNew <- predict(riceFarmDecisionTreeModelNew,
                                 riceFarmsTesting,
                                 type = 'class')

# Displaying the prediction data on the console
print(riceFarmPredictionNew)

# Generating confusion matrix to evaluate the model
riceFramConfusionMatrixNew <- table(riceFarmsTesting$FarmOwnership,
                                    riceFarmPredictionNew)

# Displaying predictiveAccuracy model on the console
print(riceFramConfusionMatrixNew)

# Generating predictiveAccuracy of the model and storing it in a varaiable
# called predictiveAccuracy
predictiveAccuracyNew <- sum(diag(riceFramConfusionMatrixNew)) /
  nrow(riceFarmsTesting)

# Displaying predictiveAccuracy on the console
print(predictiveAccuracyNew)
```

Decision tree visualization using cp = 0.01



Decision tree visualization using cp = 0.007



Questions:

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

   Answer: No, increasing the complexity of the decision tree improve the model's predictive accuracy, on contrast it decreased the model accuracy from 87.74% to 87.25%. This is because of overfitting i.e., performance on an independent set (validation data) improves up to a point, then starts to get worse.