**R script:**

# Aadhithya Dinesh

# MIS 545 Section 02

# Lab09DineshA.R

# to import a dataset of Indonesian rice farms and generate a decision tree

# model that will predict a farm's ownership status (farmer-owned or sharecropped)

# based on other farm data.

# install.packages("tidyverse")

# install.packages("rpart.plot")

library(tidyverse)

library(rpart)

library(rpart.plot)

# set the working directory

setwd("~/MIS/Classes/MIS545/Assignments/Lab09")

riceFarms <- read\_csv(file = "IndonesianRiceFarms.csv",

col\_types = "fniiinf",

col\_names = TRUE)

# print the riceFarms tibble

print(riceFarms)

# print the structure of riceFarms

print(str(riceFarms))

# print the summary of riceFarms

print(summary(riceFarms))

# set the seed to 370

set.seed(370)

sampleSet <- sample(nrow(riceFarms),

round(nrow(riceFarms)\*0.75),

replace = FALSE)

# loading 75% of the training dataset

riceFarmsTraining <- riceFarms[sampleSet, ]

# loading the remaining 25% of the dataset for testing

riceFarmsTesting <- riceFarms[-sampleSet, ]

# create the decsion tree model for farm ownership with cp = 0.01

farmOwnershipModel <- rpart(formula = FarmOwnership ~.,

method = "class",

cp = 0.01,

data = riceFarmsTraining)

# display the decsion tree plot

rpart.plot(farmOwnershipModel)

# predict the classes for each record

riceFarmsPrediction <- predict(farmOwnershipModel,

riceFarmsTesting,

type = "class")

# display the predictions from riceFarmsPrediction

print(riceFarmsPrediction)

# create the confusion matrix

riceFarmsConfusionMatrix <- table(riceFarmsTesting$FarmOwnership,

riceFarmsPrediction)

# display the confusion matrix

print(riceFarmsConfusionMatrix)

# displaying the predictive accuracy of the decision tree model

predictiveAccuracy <- sum(diag(riceFarmsConfusionMatrix)) /

nrow(riceFarmsTesting)

print(predictiveAccuracy)

# create the decsion tree model for farm ownership with cp = 0.007

farmOwnershipModel2 <- rpart(formula = FarmOwnership ~.,

method = "class",

cp = 0.007,

data = riceFarmsTraining)

# display the decsion tree plot

rpart.plot(farmOwnershipModel2)

# predict the classes for each record

riceFarmsPrediction2 <- predict(farmOwnershipModel2,

riceFarmsTesting,

type = "class")

# display the predictions from riceFarmsPrediction

print(riceFarmsPrediction2)

# create the confusion matrix

riceFarmsConfusionMatrix2 <- table(riceFarmsTesting$FarmOwnership,

riceFarmsPrediction2)

# display the confusion matrix

print(riceFarmsConfusionMatrix2)

# displaying the predictive accuracy of the decision tree model

predictiveAccuracy2 <- sum(diag(riceFarmsConfusionMatrix2)) /

nrow(riceFarmsTesting)

print(predictiveAccuracy2)

**Decision tree visualization for cp = 0.01**

Timeline

Description automatically generated

**Decision tree visualization for cp = 0.007**

Timeline

Description automatically generated

**Answer:**

Increasing the complexity counter-intuitively reduced the accuracy in this case from 0.877451 to 0.872549. This is because as we increase the complexity, we tend to over-fit our model for the training set and the model wouldn’t be able to predict the classes for an unknown testing set.