Progressive Education Society's

**Modern College of Engineering**

**MCA Department**

**A.Y. 2024-25**

**Subject Code: 410908: Data Science Laboratory**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Class: SY MCA Div: A Batch: S3 Roll Number: 52062

Name: Laxman Shinde Assignment No: 03 Date of Implementation: 15/09/24

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**1. We have four things grape, green bean, nuts and orange with two characteristics sweetness (8, 3, 3, 7) and Crunchiness (5, 7, 6, 3). Among them two are fruits, one is protein and one is vegetable. Suppose we wanted to classify tomato into one of the classes. Is tomato a fruit, vegetable or protein? Tomato has the following characteristics: sweetness = 6, crunchiness = 4. Let’s add Carrot with characteristics sweetness = 4 and crunchiness = 9 keep k=1. Try for k=4 also.**

**Program-**

# Load necessary library library(class)

# Define the data items <- data.frame( name = c("grape", "green bean", "nuts", "orange"), sweetness = c(8, 3, 3, 7), crunchiness = c(5, 7, 6, 3), class = c("fruit", "vegetable", "protein", "fruit") )

# Define the new items to classify new\_items <- data.frame( name = c("tomato", "carrot"), sweetness = c(6, 4),

crunchiness = c(4, 9)

)

# Prepare the training data

train\_data <- items[, c("sweetness", "crunchiness")] train\_labels <- items$class # k-NN classification with k=1

k1\_predictions <- knn(train = train\_data, test = new\_items[, c("sweetness", "crunchiness")], cl = train\_labels, k = 1)

# k-NN classification with k=4

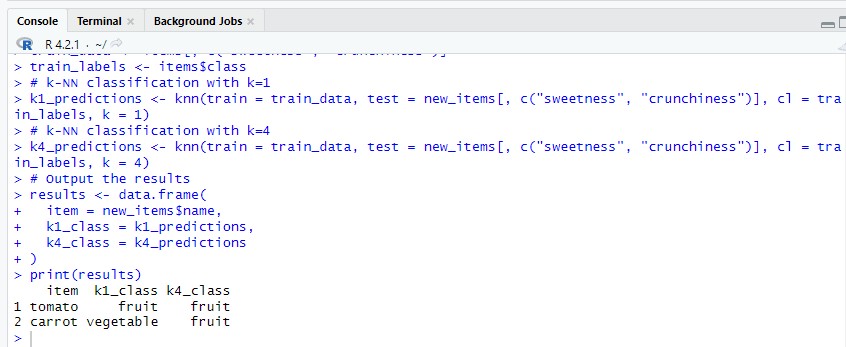
k4\_predictions <- knn(train = train\_data, test = new\_items[, c("sweetness", "crunchiness")], cl = train\_labels, k = 4) # Output the results

results <- data.frame( item = new\_items$name, k1\_class = k1\_predictions,

k4\_class = k4\_predictions

) print(results)

**Output-**



**2. Using Titanic.CSV file predict which people are more likely to survive after the collision with the iceberg using Decision Trees.**

**Program-**

install.packages("rpart.plot")

set.seed(678) titanic <-read.csv("D:/titanic.csv") head(titanic)

tail(titanic) shuffle\_index <- sample(1:nrow(titanic)) head(shuffle\_index) titanic <- titanic[shuffle\_index, ] head(titanic) library(dplyr)

# Drop variables clean\_titanic <- titanic %>%

select(-c(Cabin,Name, Fare,Ticket)) %>% #Convert to factor level mutate(Pclass = factor(Pclass, levels = c(1, 2, 3), labels = c('Upper', 'Middle', 'Lower')), Survived =

factor(Survived, levels = c(0, 1), labels = c('No', 'Yes'))) %>%

na.omit()

glimpse(clean\_titanic) create\_train\_test <- function(data, size = 0.8, train = TRUE) { n\_row = nrow(data) total\_row = size \* n\_row

train\_sample <- 1: total\_row if (train == TRUE) {

return (data[train\_sample, ])

} else {

return (data[-train\_sample, ])

} } data\_train <- create\_train\_test(clean\_titanic, 0.8, train = TRUE) data\_test <- create\_train\_test(clean\_titanic, 0.8, train = FALSE) dim(data\_train) dim(data\_test) prop.table(table(data\_train$Survived)) prop.table(table(data\_test$Survived))

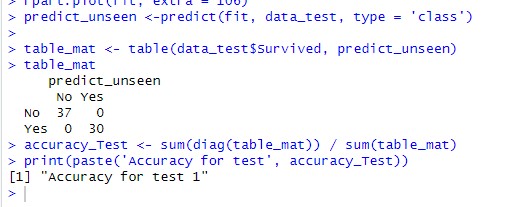
library(rpart)

library(rpart.plot) fit <- rpart(Survived~., data = data\_train, method = 'class') rpart.plot(fit, extra = 106)

predict\_unseen <-predict(fit, data\_test, type = 'class')

table\_mat <- table(data\_test$Survived, predict\_unseen) table\_mat accuracy\_Test <- sum(diag(table\_mat)) / sum(table\_mat) print(paste('Accuracy for test', accuracy\_Test))

**Output-**



**Plot-**

