**Agricultural Data Analysis**

# AIM:

he purpose of the subsequent experiment is to work with the agricultural dataset, attempt to evaluate a performance analysis of it, and attempt to understand the outcomes of the experiment.

# DESCRIPTION:

The following dataset based on yield include country, item, year starting from 1961 to 2016 and yield value and the top ten most grown crops around the world which includes,

* Cassava
* Maize
* Plantains and others
* Potatoes
* Rice, paddy
* Sorghum
* Soybeans
* Sweet potatoes
* Wheat
* Yams

A confusion matrix is been created to evaluate the performance of our model and the observations are recorded.

# PROGRAM:

data <- read.csv(file.choose()) #read the yield data print(data)

str(data) library("Metrics") library("gmodels") library("caret") library("forcats")

data$Area.Code <- as.factor(data$Area.Code) data$Item.Code <- as.factor(data$Item.Code) data$Year.Code <- as.factor(data$Year.Code) str(data)

trainIndex <- createDataPartition(data$Value, p = 0.8, list = FALSE) trainData <- data[trainIndex, ]

testData <- data[-trainIndex, ]

glmModel <- glm(Value ~ Area.Code + Item.Code + Year.Code, data = trainData, family = gaussian)

predictions <- predict(glmModel, newdata = testData) predicted\_labels <- ifelse(predictions < 0.5, 0, 1) View(predicted\_labels)

length(predicted\_labels)

dim(predicted\_labels) nrow(actual\_labels)

testData$norm\_value <- scale(testData$Value, center = FALSE, scale = max(testData$Value))

mae(testData$norm\_value, predicted\_labels) mse(testData$norm\_value, predicted\_labels) rmse(testData$norm\_value, predicted\_labels)

actual\_labels <- testData$norm\_value actual\_labels <- ifelse(actual\_labels < 0.5, 0, 1)

confusion\_matrix <- table(actual\_labels, predicted\_labels) print(confusion\_matrix)

acc <- 1287/11341

acc

errorrate <- 10054/11341 errorrate

precesion <- 2/10054 precesion

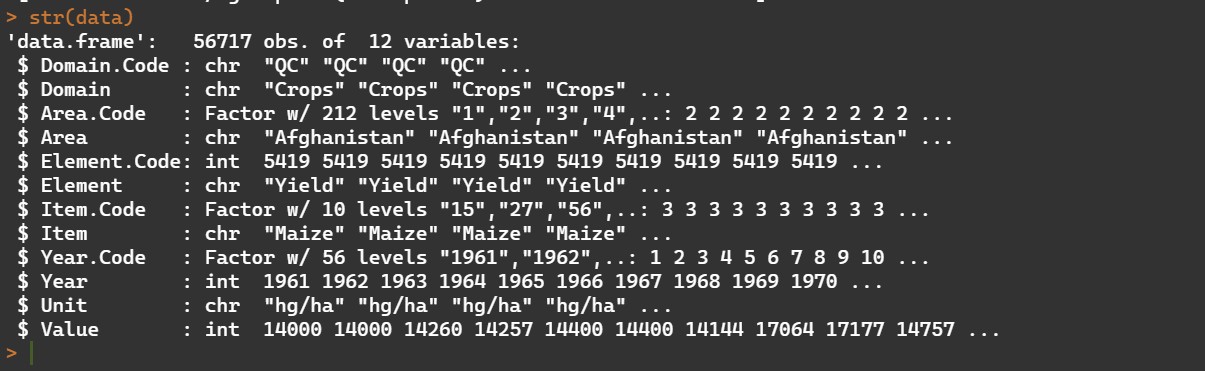
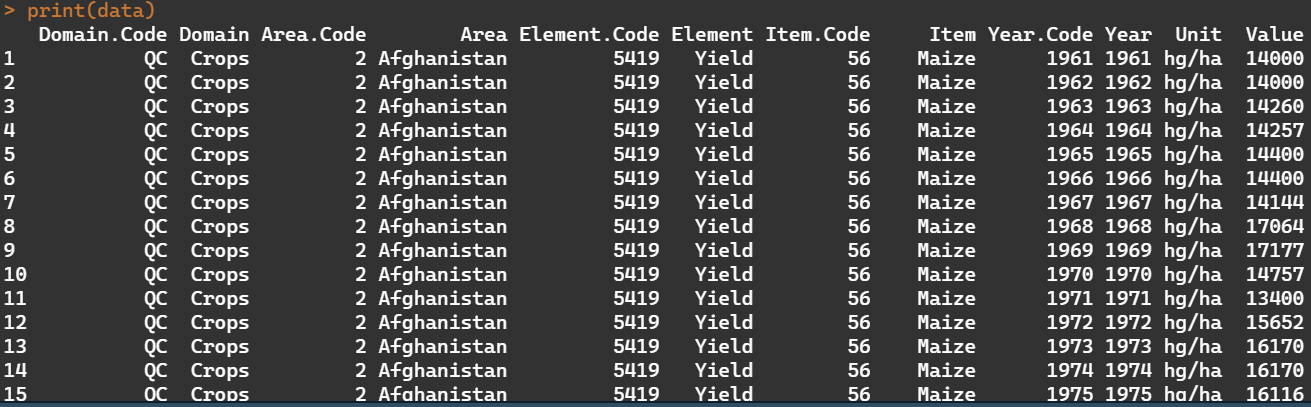
recall <- 2/2 recall

specificity <- 1285/11339 specificity

f1score <- 2 \*(precesion \* recall)/ (precesion + recall) f1score

# OUTPUT:

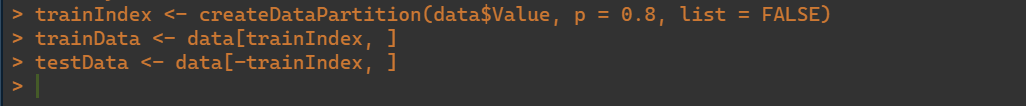
After loading the data set we try to print its values and analyse its structure.



Importing the necessary libraries



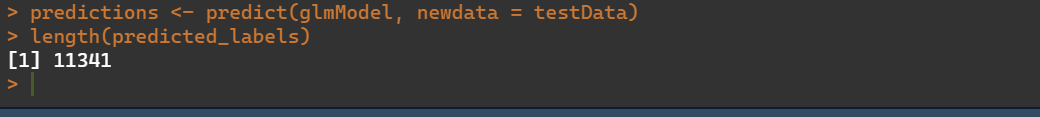
Splitting the data set into train and split data



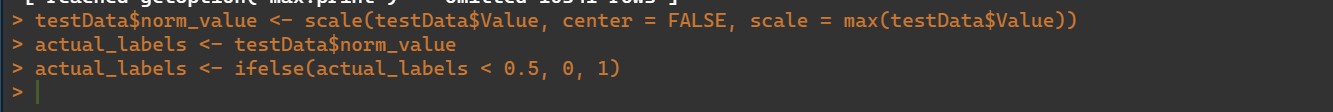
Since, dealing with categorical variables as our dependent variable and our predicted variable is (Value) is numeric in nature we use a linear regression model.



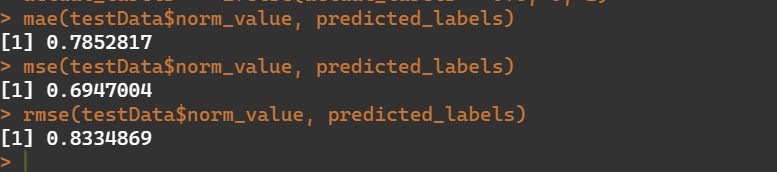
Confusion matrix works best when the target variable is categorical in nature. So we try to convert the output variable to categorical in nature.



Now to calculate the actual labels, we are making it with equal dimensions



Mae, Mse, Rmse



Accuracy, error rate, precision, recall, specificity and f1\_scre calculation:

