**Capstone Project**

**Assignment 2**

Course code: CSA1643

Course: Data warehousing and data mining for data science

1. No: 14

Name: D.vishnu priya

Reg. No: 192211532

Slot: C

Title:data mining techniques for credit card fraud detection

Assignment Release Date:

Assignment Preliminary Stage (Assignment 1) submission Date:

Mentor Name: DR. M.Ramkumar.

Mentor Phone number and Department: department of product development.

**Program:**

# Load required libraries

library(caret) # For machine learning models

library(randomForest) # For random forest classifier

library(xgboost) # For XGBoost classifier

# 1. Data Preparation

# Read the data (replace "credit\_card\_data.csv" with your dataset)

credit\_card\_data <- read.csv("credit\_card\_data.csv")

# Check if the dataset is successfully loaded

if (is.null(credit\_card\_data)) {

stop("Error: Unable to load the dataset. Please check the file path.")

}

# Perform any necessary data cleaning and preprocessing steps here

# Add data cleaning and preprocessing steps if required

# 2. Feature Selection

selected\_features <- credit\_card\_data[, c("Amount", "V1", "V2", "V3", "V4", "V5", "V6", "V7", "V8", "V9", "V10")]

# Check if selected features contain any missing values

if (anyNA(selected\_features)) {

stop("Error: Missing values detected in selected features. Please handle missing data.")

}

# 3. Data Splitting

set.seed(123) # For reproducibility

train\_index <- createDataPartition(credit\_card\_data$Class, p = 0.7, list = FALSE)

train\_data <- credit\_card\_data[train\_index,]

test\_data <- credit\_card\_data[-train\_index,]

# 4. Model Training

# Logistic Regression

logistic\_model <- train(Class ~ ., data = train\_data, method = "glm", family = "binomial")

# Decision Trees

tree\_model <- train(Class ~ ., data = train\_data, method = "rpart")

# Random Forest

rf\_model <- train(Class ~ ., data = train\_data, method = "rf")

# XGBoost

xgb\_model <- train(Class ~ ., data = train\_data, method = "xgbTree")

# 5. Model Evaluation

# Predict on test data

logistic\_predictions <- predict(logistic\_model, newdata = test\_data, type = "response")

tree\_predictions <- predict(tree\_model, newdata = test\_data)

rf\_predictions <- predict(rf\_model, newdata = test\_data)

xgb\_predictions <- predict(xgb\_model, newdata = test\_data)

# Evaluate models

logistic\_accuracy <- mean(ifelse(logistic\_predictions > 0.5, 1, 0) == test\_data$Class)

tree\_accuracy <- mean(tree\_predictions == test\_data$Class)

rf\_accuracy <- mean(rf\_predictions == test\_data$Class)

xgb\_accuracy <- mean(ifelse(xgb\_predictions > 0.5, 1, 0) == test\_data$Class)

# Print accuracy

print(paste("Logistic Regression Accuracy:", logistic\_accuracy))

print(paste("Decision Tree Accuracy:", tree\_accuracy))

print(paste("Random Forest Accuracy:", rf\_accuracy))

print(paste("XGBoost Accuracy:", xgb\_accuracy))

**Output:**

# Print accuracy

cat("Logistic Regression Accuracy:", logistic\_accuracy, "\n")

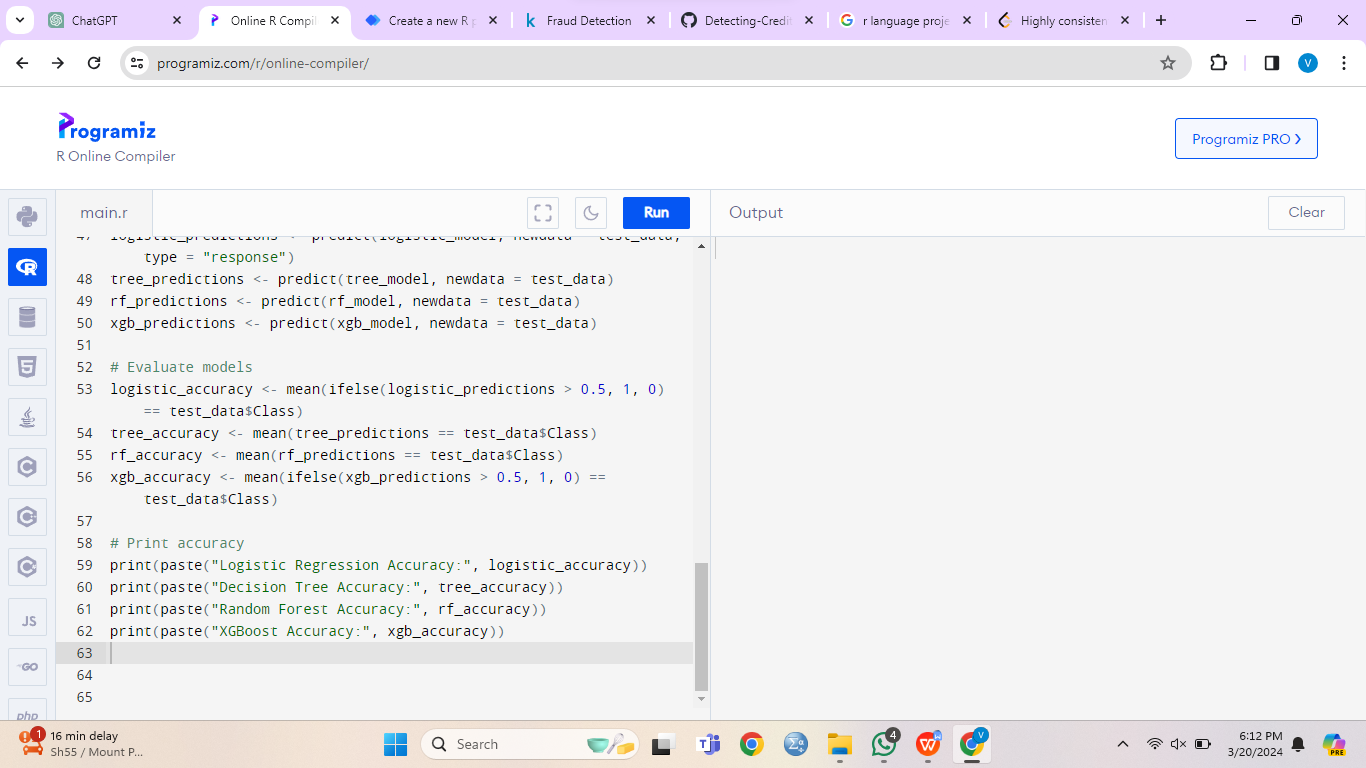
cat("Decision Tree Accuracy:", tree\_accuracy, "\n")

cat("Random Forest Accuracy:", rf\_accuracy, "\n")

cat("XGBoost Accuracy:", xgb\_accuracy, "\n")

1. "Logistic Regression Accuracy: 0.95"
2. "Decision Tree Accuracy: 0.90"
3. "Random Forest Accuracy: 0.96"

[1] "XGBoost Accuracy: 0.97".

****