Page 1

**Capstone Project**

**Preliminary Stage Assignment 2**

**Course code:**CSA1643

**Course :**Data warehousing and Data Mining for Data Science

**S.No**: 34

**Name**:S.Bhadradri Satti Reddy

**Reg. No**: 192224240

**Slot** :c

**Title :** : Fraud Detection in online Auction Platforms using data warehousing.

**Assignment Release Date** :14/02/2024

**Assignment** **Preliminary Stage ( Assignment 2) submission Date** : 16/02/2024  
**Mentor Name :** DR.B.Rengnathan

**Mentor Phone number and Department** : 9043936537 and Nano Electronic Materialsa and Sensors

**R PROGRAMMING :  
  
# Simulated data generation**

**set.seed(123)**

**n <- 1000 # Number of observations**

**# Generate random features**

**features <- data.frame(**

**feature1 = runif(n),**

**feature2 = rnorm(n),**

**feature3 = sample(letters[1:3], n, replace = TRUE),**

**feature4 = rbinom(n, 1, 0.5)**

**)**

**# Generate random labels (0 for non-fraud, 1 for fraud)**

**labels <- sample(0:1, n, replace = TRUE)**

**# Combine features and labels into a single dataframe**

**data <- cbind(features, label = labels)**

**# Split the data into training and testing sets**

**set.seed(456)**

**train\_index <- sample(1:n, 0.8\*n)**

**train\_data <- data[train\_index, ]**

**test\_data <- data[-train\_index, ]**

**# Train logistic regression model**

**model <- glm(label ~ ., data = train\_data, family = "binomial")**

**# Make predictions on test data**

**predictions <- predict(model, newdata = test\_data, type = "response")**

**# Convert predicted probabilities to binary predictions (0 or 1)**

**binary\_predictions <- ifelse(predictions > 0.5, 1, 0)**

**# Evaluate model performance**

**confusion\_matrix <- table(binary\_predictions, test\_data$label)**

**accuracy <- sum(diag(confusion\_matrix)) / sum(confusion\_matrix)**

**precision <- confusion\_matrix[2, 2] / sum(confusion\_matrix[, 2])**

**recall <- confusion\_matrix[2, 2] / sum(confusion\_matrix[2, ])**

**f1\_score <- 2 \* precision \* recall / (precision + recall)**

**# Output**

**print("Confusion Matrix:")**

**print(confusion\_matrix)**

**cat("\n")**

**print("Metrics:")**

**print(paste("Accuracy:", accuracy))**

**print(paste("Precision:", precision))**

**print(paste("Recall:", recall))**

**print(paste("F1-score:", f1\_score))**