Progressive Education Society's
Modern College of Engineering, Shivajinagar, Pune-05.

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Department of MCA

PRACTICAL SUBMISSION RECORD- A.Y. 2025-26

Class: SYMCA Division : A **Semester: III**	**Course Code:** MCA01604 **Course Name:** Data Science Laboratory	**Batch:** S2
Name: Pranav Raju Malwatkar	**Roll No:** 52037	
CO No: CO605.1	**Assignment No:** 3	

Title: Load the Titanic dataset, perform basic preprocessing (handle missing values, encode categorical variables), and train a Decision Tree classifier to predict the model. Plot the confusion matrix.

Code:

```
library(titanic)

library(rpart)

library(caret)

df <- titanic_train

df <- df[, c("Survived", "Pclass", "Sex", "Age",

"Fare")] mean_age <- mean(df$Age, na.rm = TRUE)

df$Age[is.na(df$Age)] <- mean_age

df$Survived <- as.factor(df$Survived)

df$Pclass <- as.factor(df$Pclass)

set.seed(123)

train_indices <- sample(1:nrow(df), 0.8 * nrow(df))

training_set <- df[train_indices, ]
```

```
test_set <- df[-train_indices, ]

tree_model <- rpart(Survived ~ ., data = training_set, method =

"class") predictions <- predict(tree_model, newdata = test_set, type =

"class")

print("Confusion Matrix and Statistics:")

confusionMatrix(predictions, test_set$Survived, positive =

"1")
```

Output:

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 95 20
         1 15 49

               Accuracy : 0.8045
                 95% CI : (0.7387, 0.8599)
    No Information Rate : 0.6145
    P-Value [Acc > NIR] : 3.582e-08

                  Kappa : 0.5816

 Mcnemar's Test P-Value : 0.499

            Sensitivity : 0.7101
            Specificity : 0.8636
         Pos Pred Value : 0.7656
         Neg Pred Value : 0.8261
             Prevalence : 0.3855
         Detection Rate : 0.2737
   Detection Prevalence : 0.3575
      Balanced Accuracy : 0.7869

       'Positive' Class : 1
```