Mini Project 02 : Machine learning Classification

```
library(tidyverse)
library(caret)
library(titanic)

glimpse(titanic_train)
titanic <- na.omit(titanic_train)
titanic$Survived <- factor(titanic$Survived)
```

```
# split data 80% to train model, 20% to test model

set.seed(50)
n <- nrow(titanic)
id <- sample(1:n , size = 0.8*n)
train_data <- titanic[id,]
test_data <- titanic[-id,]
```

```
# train data in logistic regression , random forest and decision trees to find su
ctrl <- trainControl( method = "cv",
                      number = 5,
                      verboseIter = TRUE)

glm_model <- train(Survived ~ Pclass + Sex + Age,
                   data = train_data,
                   method = "glm", #logostic regression model
                   metric = "Accuracy",
                   trControl = ctrl)

rf_model <- train(Survived ~ Pclass + Sex + Age,
                   data = train_data,
                   method = "ranger", #random forest model
                   metric = "Accuracy",
                   trControl = ctrl)

rpart_model <- train(Survived ~ Pclass + Sex + Age,
                   data = train_data,
                   method = "rpart", #decision tree model
                   metric = "Accuracy",
                   trControl = ctrl)
```

```
# prediction survived
glm_pred <- predict(glm_model,newdata = test_data)
rf_pred <- predict(rf_model,newdata = test_data)
rpart_pred <- predict(rpart_model,newdata = test_data)
```

```
#evaluate model

list_model <- list(glm_model = glm_model,
                   rf_model = rf_model,
                   rpart_model = rpart_model)

result <- resamples(list_model)
summary(result) #show model's accuracy

mean(glm_pred == test_data$Survived) #accuracy = 0.804
mean(rf_pred == test_data$Survived) #accuracy = 0.769
mean(rpart_pred == test_data$Survived) #accuracy = 0.762

confusionMatrix(glm_pred,test_data$Survived,mode = "prec_recall") #show Confusion
```