# Homework_Stats_Batch_06

YO

### Library

```
install.packages("titanic")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

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

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

### Drop NA (missing value)

```
titanic_train <- na.omit(titanic_train)
nrow(titanic_train)
```

```
## [1] 714
```

## View Data

```
glimpse(titanic_train)
```

```
## Rows: 714
## Columns: 12
## $ PassengerId <int> 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19~
## $ Survived    <int> 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1~
## $ Pclass      <int> 3, 1, 3, 1, 3, 1, 3, 3, 2, 3, 1, 3, 3, 3, 2, 3, 3, 2, 2, 3~
## $ Name        <chr> "Braund, Mr. Owen Harris", "Cumings, Mrs. John Bradley (Fl~
## $ Sex         <chr> "male", "female", "female", "female", "male", "male", "mal~
## $ Age         <dbl> 22, 38, 26, 35, 35, 54, 2, 27, 14, 4, 58, 20, 39, 14, 55, ~
## $ SibSp       <int> 1, 1, 0, 1, 0, 0, 3, 0, 1, 1, 0, 0, 1, 0, 0, 4, 1, 0, 0, 0~
## $ Parch       <int> 0, 0, 0, 0, 0, 0, 1, 2, 0, 1, 0, 0, 5, 0, 0, 1, 0, 0, 0, 0~
## $ Ticket      <chr> "A/5 21171", "PC 17599", "STON/O2. 3101282", "113803", "37~
## $ Fare        <dbl> 7.2500, 71.2833, 7.9250, 53.1000, 8.0500, 51.8625, 21.0750~
## $ Cabin       <chr> "", "C85", "", "C123", "", "E46", "", "", "", "G6", "C103"~
## $ Embarked    <chr> "S", "C", "S", "S", "S", "S", "S", "S", "C", "S", "S", "S"~
```

## 1. Split DATA

```
set.seed(10)
n <- nrow(titanic_train)
id <- sample(1:n, size = n*0.7) ## 70% train 30% test
train_data <- titanic_train[id, ]
test_data <- titanic_train[-id, ]
```

## 2. Train Model

```
model_train <- glm(Survived ~ Pclass + Age + Sex, data = train_data, family = "binomial")
summary(model_train)
```

```
##
## Call:
## glm(formula = Survived ~ Pclass + Age + Sex, family = "binomial",
##     data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7761  -0.7050  -0.3826   0.7038   2.4686
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   5.2269     0.5990   8.726  < 2e-16 ***
## Pclass       -1.3140     0.1668  -7.875 3.41e-15 ***
## Age          -0.0406     0.0090  -4.511 6.44e-06 ***
## Sexmale      -2.4970     0.2473 -10.097  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 679.83  on 498  degrees of freedom
## Residual deviance: 465.44  on 495  degrees of freedom
## AIC: 473.44
##
## Number of Fisher Scoring iterations: 5
```

## 3. Predict and Evaluate Model

```
train_data$prob_survived <- predict(model_train, type = "response")
train_data$pred_survived <- ifelse(train_data$prob_survived >= 0.5, 1, 0)
```

### 4. Confusion Matrix of Train Model

```
conM <- table(train_data$pred_survived, train_data$Survived,
                      dnn = c("Predicted", "Actual"))

acc_train <- (conM[1, 1] + conM[2, 2]) / sum(conM)
prec_train <- conM[2, 2] / (conM[2, 1] + conM[2, 2])
rec_train <- conM[2, 2] / (conM[1, 2] + conM[2, 2])

f1_train <- 2*((prec_train * rec_train) / (prec_train + rec_train))

cat("Accuracy : ", acc_train, "\nPrecision : ", prec_train, "\nRecall : ", rec_train, "\nF1 Score : ",
```

```
## Accuracy :  0.7995992
## Precision :  0.7655502
## Recall :  0.7582938
## F1 Score :  0.7619048
```