

Experiment 8

April 21, 2025

```
[1]: library(tidyverse)
library(dplyr)
```

Attaching core tidyverse packages

```
tidyverse 2.0.0
dplyr      1.1.4      readr      2.1.5
forcats    1.0.0      stringr    1.5.1
ggplot2    3.5.2      tibble     3.2.1
lubridate  1.9.4      tidyr      1.3.1
purrr      1.0.4
Conflicts
```

```
tidyverse_conflicts()
dplyr::filter() masks stats::filter()
dplyr::lag()     masks stats::lag()
Use the conflicted package
(<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

```
[2]: # Basic way to load a CSV
train <- read.csv("/home/asus/content/Notes/Semester 4/FDN Lab/Experiments/
↳Experiment 8/titanic/train.csv")

test <- read.csv("/home/asus/content/Notes/Semester 4/FDN Lab/Experiments/
↳Experiment 8/titanic/test.csv")
```

```
[3]: # Removing Unessacry Cols
train <- train %>% select(-one_of("Cabin", "Ticket", "Name", "Embarked"))
test <- test %>% select(-one_of("Cabin", "Ticket", "Name", "Embarked"))
```

```
[4]: train <- train %>% fill(everything(), .direction = "down")
test <- test %>% fill(everything(), .direction = "down")
```

```
[5]: X_train <- train %>% select(-Survived)
Y_train <- train %>% select(Survived)
```

```

[6]: train_df <- as_tibble(train) %>%
      mutate(Survived = train)

[7]: # Train the model
      logit_model <- glm(Survived ~ .,
                        data = train,
                        family = binomial)

[8]: predictions <- predict(logit_model, newdata = train, type = "response")

[9]: predicted_classes <- ifelse(predictions > 0.5, 1, 0)

[10]: ground_truth <- train$Survived

[11]: conf_matrix <- table(Predicted = predicted_classes, Actual = ground_truth)

[12]: print(conf_matrix)
# Actual
# Predicted    0    1
#              0 472 110
#              1   77 232

      Actual
Predicted  0    1
          0 472 110
          1   77 232

[13]: accuracy <- sum(diag(conf_matrix))/sum(conf_matrix)
      precision <- conf_matrix[2,2]/sum(conf_matrix[2,])
      recall <- conf_matrix[2,2]/sum(conf_matrix[,2])
      f1_score <- 2 * (precision * recall) / (precision + recall)

[14]: summary(logit_model)
      print(conf_matrix)
      cat("\nAccuracy:", round(accuracy, 3))
      cat("\nPrecision:", round(precision, 3))
      cat("\nRecall/Sensitivity:", round(recall, 3))
      cat("\nF1 Score:", round(f1_score, 3))
      cat("\nSpecificity:", round(conf_matrix[1,1]/sum(conf_matrix[,1]), 3))

```

Call:

```
glm(formula = Survived ~ ., family = binomial, data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6513	-0.6196	-0.4077	0.6269	2.6737

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.6705255	0.5301556	8.810	< 2e-16 ***
PassengerId	0.0000850	0.0003468	0.245	0.80639
Pclass	-1.0457412	0.1374434	-7.609	2.77e-14 ***
Sexmale	-2.8025118	0.2007943	-13.957	< 2e-16 ***
Age	-0.0338376	0.0067970	-4.978	6.42e-07 ***
SibSp	-0.3422950	0.1094887	-3.126	0.00177 **
Parch	-0.1195347	0.1171594	-1.020	0.30760
Fare	0.0031898	0.0023918	1.334	0.18233

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1186.66 on 890 degrees of freedom
Residual deviance: 790.18 on 883 degrees of freedom
AIC: 806.18

Number of Fisher Scoring iterations: 5

	Actual	
Predicted	0	1
0	472	110
1	77	232

Accuracy: 0.79
Precision: 0.751
Recall/Sensitivity: 0.678
F1 Score: 0.713
Specificity: 0.86