Project Title – Sink or Swim? Titanic survivor prediction

Problem Definition – The goal of this project is to build a model that could predict whether a passenger survived or not. It is a binary classification problem

Data Source – I will be using the data available from Kaggle (<https://www.kaggle.com/datasets/yasserh/titanic-dataset>). It consists of 12 attributes, 3 predictors, and 1 target variable. The predictors include variables such as passenger class, age, and sex. The target variable is Survived which is binary and 0 means they didn’t make it while 1 means they survived.

Machine Learning Algorithm – I am planning to use Logistic Regression since it is a supervised algorithm

Metric of Assessment – Since this is a relatively trivial dataset for events that happened in the past (no lives are at stake on future Titanic) I will stick with accuracy as the benchmark for the performance score of my model. Accuracy will be compared with the work done (<https://www.kaggle.com/code/walidabdelhameed/apply-10-models-to-titanic-dataset?scriptVersionId=196557170&cellId=62>) in which their Test accuracy for Logistic Regression was 81.82% and F1 score of 80.39%.