Titanic

The task: The sinking of the Titanic is one of the most infamous shipwrecks in history. Unfortunately, in that incident there weren’t enough lifeboats for everyone on board, resulting in the death of 1502 out of 2224 passengers and crew. Our aim in Titanic Dataset is to predict the survival status of passengers. What type of problem: Since we have to find the given passenger survived or not, thus this is a categorical problem which goes to classification.

Algorithm: Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for a given set of features (or inputs), X.

Why This Algorithm: There are a lot of predictive modeling algorithms to choose from. Here, our problem is a classification and regression problem. We want to check the relationship between output (Survived or NOT Survived) with other variables or features like (Gender, Age, Class, etc). So I have Logistic regression.

Result: I use 4 features of dataset to train on model, which is Pclass, Sex, Age and Embarked, and train on logistic regression model. The model can achieved accuracy with 0.82 score in Kaggle scoreboard which is almost the same accuracy as validation dataset (with 0.7 split ratio on training dataset).

Accuracy : 82%