This is an attempt to capture my approach to the kaggle competition, Titanic’s Machine learning from disaster. First of all, in this problem, I wanted to explore the whole data to understand the factor impacting the survival and see whether I could predict the survival using machine learning models:

The first thing I did that gathered information about the whole features/variables in the dataset, to check the datatypes of the dataset and also the missing values in the dataset.

Afterward I did the data visualization:

1. The percentage of the survival rate in the titanic
2. Does gender have any bearing on the survival of the passenger?
3. Does Pclass has any types of relation with survival of the passenger?
4. Does travelling with Parch improve the chance of survival?
5. Does port of embarkation impact the survival of the passenger?

The sinking of the titanic is one of the most infamous shipwrecks in history. some of the reason that the shipwreck led to such a loss of life would may be that there were not enough lifeboats for the passenger and crew. Some groups were more likely to survive than others such as women, children and the upper class.

In this dataset some missing values were also found in some of the variable such as embarked, age, cabin. There were lot of missing values in the cabin of the train and test data.so, I decided to drop out this column. Because no conclusion was found with this column. After that, with the Age variable I filled the missing value with their median value.

Model Building: Supervised Learning Technique

Logistic regression model

Random forest

These two algorithm I used for the model building.