# **Machine Learning to Determine Titanic Survivors**

# **Data Set Column Descriptions**

* ****pclass:**** Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
* ****survived:**** Survival (0 = No; 1 = Yes)
* ****name:**** Name
* ****sex:**** Sex
* ****age:**** Age
* ****sibsp:**** Number of siblings/spouses aboard
* ****parch:**** Number of parents/children aboard
* ****fare:**** Passenger fare (British pound)
* ****embarked:**** Port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
* ****adult\_male:****A male 18 or older (0 = No, 1=Yes)
* ****deck:****Deck of the ship
* ****who:****man (18+), woman (18+), child (<18)
* ****alive:**** Yes, no
* ****embarked\_town:**** Port of embarkation ( Cherbourg, Queenstown, Southampton)
* ****class:**** Passenger class (1st; 2nd; 3rd)
* ****alone:****1= alone, 0= not alone ( you have at least 1 sibling, spouse, parent or child on board)

STEPS TO PREDICTING TITANIC SURVIVAL PREDICTION:

* Import the packages/libraries to make it easier to write the program
* load the data from csv file to Pandas DataFrame
* check the number of missing values in each column
* Data Analysis
* Data Visualization
* Encoding the Categorical Columns
* Separating features & Target
* Splitting the data into training data & Test data
* Model Training
* Model Evaluation

Procedure in the predicting the titanic survival prediction:

* There is two variable that is independent and dependent variable.
* X is for independent variable and Y for dependent variable
* In the train dataset the cabin having more missing values that’s why are dropping the column for clear data , getting better prediction output.
* We are taking mean and mode values for Age and Embarked columns for getting better output prediction values.
* We are using Logistic Regression in this model.
* Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. In this we predicting whether people survived or not.
* The Y values are the Survived column. The targeted values.
* We are analyzing the dataset using seaborn by plotting it.
* We split the dataset into train data and test data.
* After spliting the data we are fit the train data in it.
* And after that we are evaluating the predicting how are survived and how many are not survived in the titanic ship.
* We are finally checking the accuracy of our predicted model.