CARNEGIE MELLON UNIVERSITY - AFRICA

# Kaggle Competition: Titanic- Machine Learning From Disaster

MUNYANEZA Kenny Roger

DATA, INFERENCE & APPLIED MACHINE LEARNING

(COURSE 18-785)

Professor Patrick McSharry

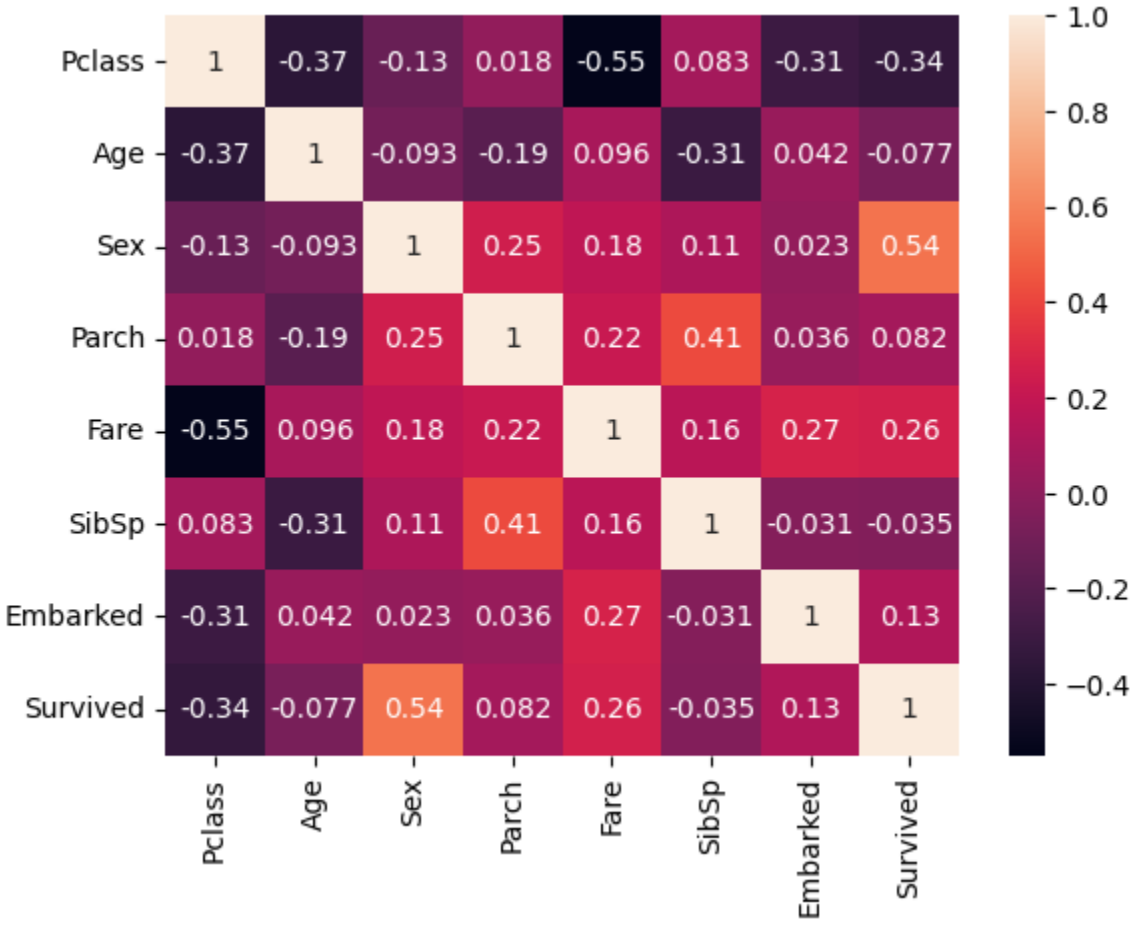
8th December 2023

# Introduction

Report showing the procedure I followed to complete the Titanic Kaggle competition.

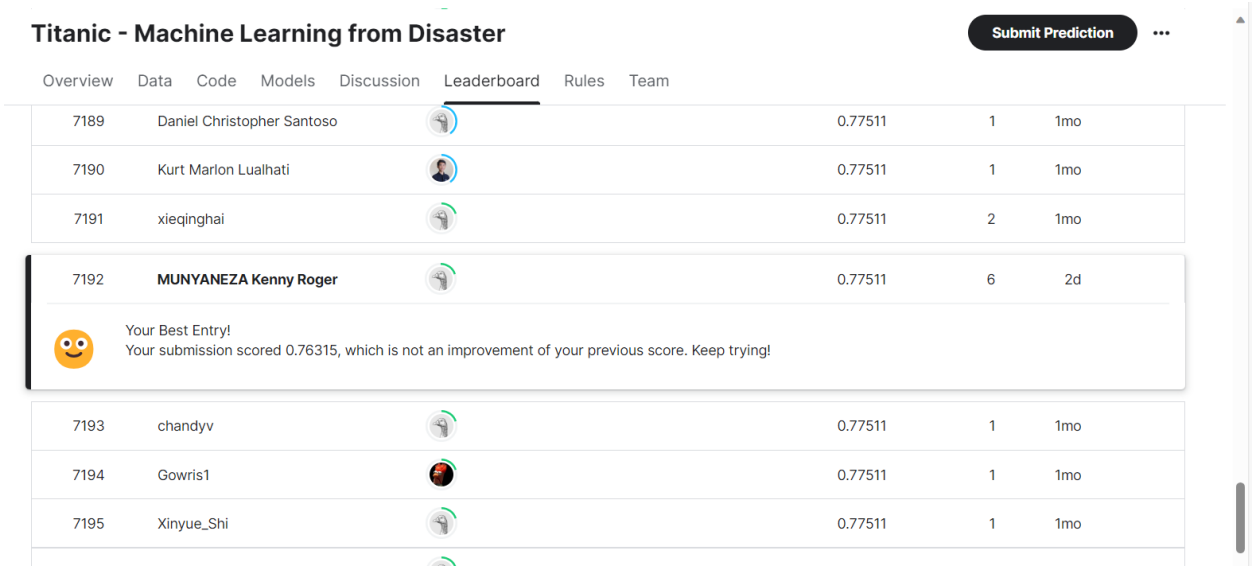
## Steps followed:

- Loading the training and testing datasets using the pandas framework.
- Using a function “data_clean_up()”, I cleaned the dataset this involved the following processes:
    - ✓ Converting the gender and embarked fields' value set from string values to binary values
    - ✓ Selecting the most important features using correlation with the “survived” field , which resulted in 'Pclass', 'Age', 'Sex', 'Parch', 'Fare', 'SibSp' and 'Embarked' as predictor variables.




    - ✓ Replacing the NaN values in columns like Age with the mean of the Age distribution.
- Then I fitted the train dataset using the Logistic Regression model.
- I used the test dataset to predict Survival statuses of the people in the test dataset.
- I generated the output as a csv file showing all individuals in the test dataset and their survival status.

## Score obtained in the competition




From the predictions I made, I was able to score 0.76 which earned me the 7192$^{nd}$ position on the leaderboard.