Titanic: Machine Learning from Disaster

(end to end project using python)

**Table of Contents:**

* Introduction
* The RMS Titanic
* Import Libraries
* Getting the Data
* Data Exploration/Analysis
* Data Preprocessing
  + Missing Data
  + Converting Features
  + Creating Categories
  + Creating new Features
* Building Machine Learning Models
  + Training 8 different models
  + Which is the best model ?
  + K-Fold Cross Validation
* Random Forest
  + What is Random Forest ?
  + Feature importance
  + Hyperparameter Tuning
* Further Evaluation
  + Confusion Matrix
  + Precision and Recall
  + F-Score
  + Precision Recall Curve
  + ROC AUC Curve
  + ROC AUC Score
* Submission
* Summary

**Introduction**

In this kernel I will go through the whole process of creating a machine learning model on the famous Titanic dataset, which is used by many people all over the world. It provides information on the fate of passengers on the Titanic, summarized according to economic status (class), sex, age and survival. In this challenge, we are asked to predict whether a passenger on the titanic would have been survived or not.

**The RMS Titanic**

RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg during its maiden voyage from Southampton to New York City. There were an estimated 2,224 passengers and crew aboard the ship, and more than 1,500 died, making it one of the deadliest commercial peacetime maritime disasters in modern history. The RMS Titanic was the largest ship afloat at the time it entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. The Titanic was built by the Harland and Wolff shipyard in Belfast. Thomas Andrews, her architect, died in the disaster.