Titanic: Machine Learning from Disaster

Description

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

In this challenge, we ask you to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc).

Approach

The problem was divided into several steps:

Data Collection : Data was collected from the Kaggle website. Data were having train, test and sample csv.

Data Wrangling: The datasets were uploaded to a dataframe and explored. Null values were filled in wherever appropriate and polluted values were discarded or wrangled.

EDA: Extensive data visualisation and summary statistics were used to extract insights and pattern from the various datasets. The history, facts and trivia behind passengers survival.

Machine Learning: Gradient Boosting Classifier,RandomForest, LogisticRegression, KNN were trained on our feature engineered dataset to predict passenger survival. Their feature importances were noted to gain insights into what factors influence the survival.

Final Result

Gradient Boosting Classifier,RandomForest, LogisticRegression, KNN were trained on our feature engineered dataset to predict passenger survival. Here is a comparison matrix.

Model	Score
GradientBoost	80.65
RandomForest	79.64
LogisticRegression	79.42
KNN	76.72

Repository Structure

Titanic_Survivor_Analysis.ipynb : The Jupyter notebook that contains the EDA and narrates the Story of Titanic Survivors.

Code : https://www.kaggle.com/navesharma9/titanic-survivor-analysis