

Project Proposal

Title : Survival Prediction on Titanic

Authors: Shreehar Joshi, Manav Bhagat, Ashish Poudel, and Russel Kennedy

Date: 11/12/2020

Section 1. Introduction

Titanic, in full Royal Mail Ship (RMS) Titanic, was a British Passenger Luxury Liner that sank on April 14-15, 1912 during its maiden voyage, en route to New York City from Southampton, England killing about 1500 passengers and ship personnel. Being one of the most famous tragedies in modern history, it has inspired numerous stories, several films, and a musical and has been the subject of much scholarship and scientific speculation.

In this project, we aim to explore the presence of factors, if any, that could have predicted the likelihood of any passenger to survive in the shipwreck. We hypothesize that the female and passengers from upper-class were more likely to survive the shipwreck.

Different python libraries like pandas, numpy, matplotlib, seaborn and scikit-learn will be applied in this project for effective data preparation, visualization, and building models that can most accurately predict the survival of different passengers in the shipwreck.

Section 2. Data Description

The dataset consists of the information of 1310 passengers that traveled in the Titanic. The 14 columns included in the dataset are pclass, survived, name, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat, body, and home.dest.

Here are the information of each of the columns in detail:

pclass: The class on which the passenger was traveling.

survived: The survival status of the passenger.

name: The name of the passenger.

sex: The sex of the passenger.

age: The age of the passenger.

sibsp: The number of siblings/spouses the passenger had aboard.

parch: The number of parents/children the passenger had aboard.

ticket: The ticket number of the passenger.

fare: The price of the ticket of the passenger.

cabin: The cabin address of the passenger.

embarked: The port from which the passenger had embarked.

boat: The lifeboat number of the passenger (if survived).

body: The body number of the passenger (if not survived and body recovered).

home.dest: The home/destination of the passenger.

While we retrieved the dataset from data.world(<https://data.world/nrippner/titanic-disaster-dataset/workspace/file?filename=titanic.csv>), the principal source for data about Titanic passengers is the Encyclopedia Titanica (by the method of web-scraping) which consists of the facts and information of thereal people that designed, built and sailed on the RMS Titanic.

Section 3. Glimpse of Data

Now, lets have a quick glimpse of our dataset.

```
In [7]: import pandas as pd
titanic = pd.read_csv('titanic.csv')
titanic
```

Out[7]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	en
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	
...
1305	3.0	0.0	Zabour, Miss. Thamine	female	NaN	1.0	0.0	2665	14.4542	NaN	
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5000	0.0	0.0	2656	7.2250	NaN	
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0000	0.0	0.0	2670	7.2250	NaN	
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0000	0.0	0.0	315082	7.8750	NaN	
1309	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

1310 rows × 14 columns

