Prediction of Survival on Titanic

Tensorflow 2.0

11/02/2020

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.

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, 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, 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 the real people that designed, built and sailed on the RMS Titanic.

Glimpse of Data

Now, lets have a quick glimpse of our dataset.

```
## Rows: 1,310
## Columns: 14
## $ pclass    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ survived  <dbl> 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, ...
## $ name      <chr> "Allen, Miss. Elisabeth Walton", "Allison, Master. Hudson...
## $ sex       <chr> "female", "male", "female", "male", "female", "male", "fe...
## $ age       <dbl> 29.0000, 0.9167, 2.0000, 30.0000, 25.0000, 48.0000, 63.00...
## $ sibsp     <dbl> 0, 1, 1, 1, 1, 0, 1, 0, 2, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, ...
## $ parch     <dbl> 0, 2, 2, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, ...
## $ ticket    <chr> "24160", "113781", "113781", "113781", "113781", "19952",...
## $ fare      <dbl> 211.3375, 151.5500, 151.5500, 151.5500, 151.5500, 26.5500...
## $ cabin     <chr> "B5", "C22 C26", "C22 C26", "C22 C26", "C22 C26", "E12", ...
## $ embarked  <chr> "S", "S", "S", "S", "S", "S", "S", "S", "S", "C", "C", "C"...
## $ boat      <chr> "2", "11", NA, NA, NA, "3", "10", NA, "D", NA, NA, "4", "...
## $ body      <dbl> NA, NA, NA, 135, NA, NA, NA, NA, NA, 22, 124, NA, NA, NA,...
## $ home.dest <chr> "St Louis, MO", "Montreal, PQ / Chesterville, ON", "Montr...
```