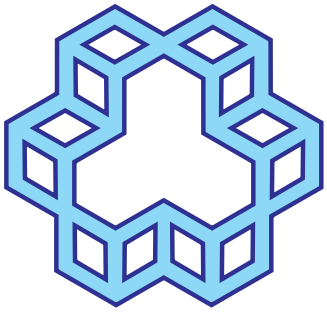

K. N. Toosi University of Technology

Faculty of Physics
Educational Group of
Atomic-Molecular and Astronomy

Machine Learning Projects
(Project 7)

by
Ali Bagheri

Teacher
Dr. Mohammad Hossein Zhoolideh

Academic Year 1401-1402
(First Semester)

Titanic Ship

Titanic is one of the most famous shipwrecks in history. On April 15, 1912, the Titanic sank after hitting an iceberg. Unfortunately, there were not enough lifeboats for everyone on board, resulting in the deaths of 1502 of the 2224 passengers and crew. While some element of luck played a role in survival, it appears that some groups of people were more likely to survive than others.

In this project, we ask you to build a predictive model using the passenger's data (ie name, age, gender, socioeconomic class, etc.) that answers the question: What kind of people are most likely to survive?

You will have access to a dataset that include the passenger's information such as name, age, gender, socio-economic class, etc. You can download the dataset needed to answer this question from this link.[1] The information provided to you is as follows:

survival Survival

- **0** No
- **1** Yes

pclass Ticket class (A proxy for socio-economic status (SES))

- **1** 1st (Upper)
- **2** 2nd (Middle)
- **3** 3rd (Lower)

sex Sex

Age Age in years (Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5)

sibsp Number of siblings / spouses aboard the Titanic. The dataset defines family relations in this way:

- **Sibling** brother, sister, stepbrother, stepsister
- **Spouse** husband, wife (mistresses and fiances were ignored)

[1] To save the dataset, you need to press Ctrl+S on the opened page and save the .csv file

parch Number of parents / children aboard the Titanic. The dataset defines family relations in this way:

Parent mother, father

Child daughter, son, stepdaughter, stepson

***** Some children travelled only with a nanny, therefore parch=0 for them.

ticket Ticket number

fare Passenger fare

cabin Cabin number

embarked Port of Embarkation

C Cherbourg

Q Queenstown

S Southampton

Note: The given data is raw. To answer this question, you must first preprocess the data using the Pandas package.

Important Points

Be sure to

- Leave appropriate comments for different parts of your code.
- Completely explain about the algorithm(s) you use to answer this question.
- Measure your model performance using model evaluation metrics and interpret the obtained result(s).

A part of your score will be allocated to these items.

***** You should write all the steps of your project in the **Jupyter notebook** and upload it as a file with the **.ipynb** extension on the vc site.