**AI LAB 1 2023**

**Assignment DUE by 30 Jan 23:45 H**

**Instructor : Moinul Zaber**

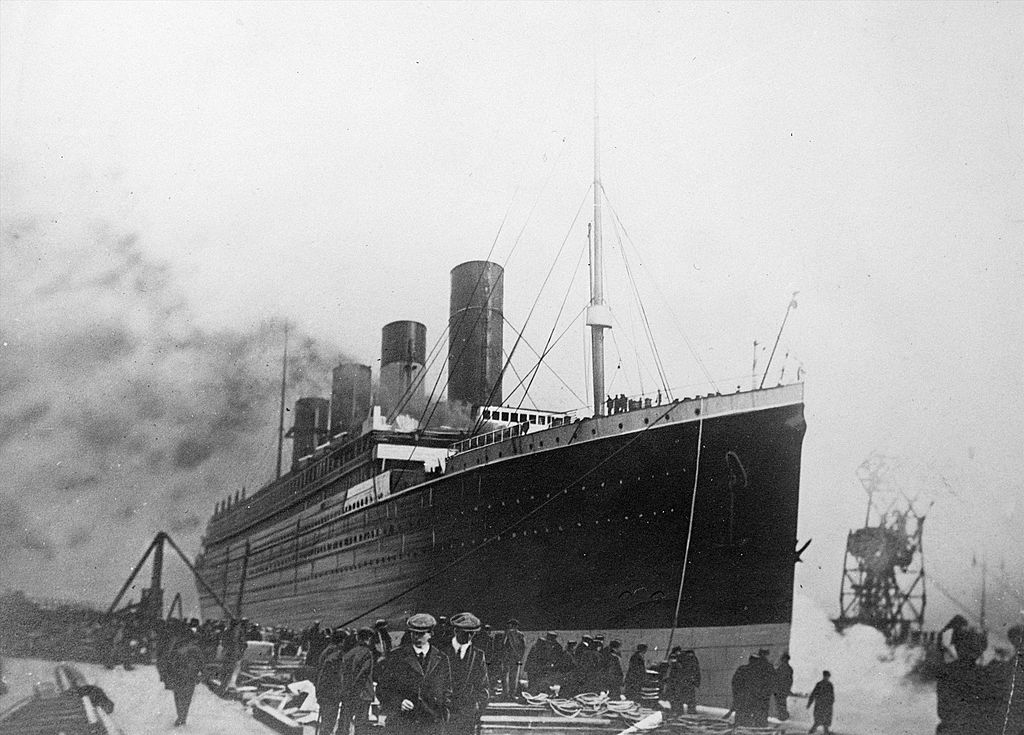


Image source : https://time.com/5767262/titanic-treaty/

The Titanic was a ship carrying 2,224 passengers with crewmates during her maiden voyage on April 15, 1912. The ship was considered unsinkable until it hit an iceberg and sank. Due to not having enough lifeboats, 1502 deaths were recorded.

Problem: Now, using the passenger records such as age, sex, socio-economic class, etc., we want to train a model that will predict how likely a person was to survive and use the model parameters to “infer” about the features in the dataset. We can use a logistic regression since the output is Boolean (Y=1, if the person survives or Y=0, if he/she dies).

Objective:

* Learning to code ML/AI with Python
* Learning to search API documentation and online resources

Evaluation: (total score: 100)

* Coding basic ML pipeline
  + Data preparation: input, output, and cleaning (30)
  + Feature engineering: preparing raw data for ML (20)
  + Train a [logistic regression](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html) model and evaluate model score (20)
* Documenting your code (20)
  + Code description in simple English
  + The reasoning behind code logic
  + Inference from model parameters
* Bonus marks for achieving a score >= 80th percentile in the leaderboard (10)

Steps:

1. Join the [competition](https://www.kaggle.com/c/titanic), read the description, data dictionary and download the data.
2. Prepare a [google colab](https://colab.research.google.com/) notebook with the pipeline and documentation
   1. Use python [pandas](https://pandas.pydata.org/docs/getting_started/index.html) and sci-kit learn ([sklearn](https://scikit-learn.org/stable/getting_started.html)). These are likely to be installed in colab.
3. Submit predictions to the competition and check score (iteratively improve pipeline if necessary)
4. Submit the following information via [form](https://forms.gle/fz27QqV8VNr9dPz26)
   1. ID
   2. Name
   3. URL to your notebook (please remember to share with code output)
   4. Kaggle username
   5. Highest submission score

Tip: You can find many online resources to learn about the tools and Kaggle. Here are a few videos from the top of my google search:

* [Corey Schafer, Pandas Tutorial](https://www.youtube.com/watch?v=ZyhVh-qRZPA),
* [Kunal Naik, sklearn logistic regression tutorial](https://www.youtube.com/watch?v=tI_Pco7snZw)
* [DS Dojo, submitting predictions in Kaggle](https://www.youtube.com/watch?v=68l47Zu4Yxg).

Feel free to look up different tutorials, resources, and titanic competition kernels (starter notebooks) to learn more. Good luck!