Tsunami behavior modeling.

data science bootcamp

Nouha Osama O. Alsaleh

Questions of this project
- Can we predict tsunami from the historical tsunami dataset?
- What can we say about the tsunami behavior from this dataset?

Data Description

This project will use the historical tsunami event dataset provided the national centers for environmental information (NOAA), pulling data from 1900-2021(current date) of all oceans. This dataset contains the following features (columns):
- Day and time
- Tsunami validity
- Tsunami cause code
- Earthquake magnitude
- Volume
- Country
- Latitude
- Longitude
- Max water height
- Source location
- Number of roundups
- Tsunami intensity
- Deaths

- Missing, injuries and damage

Modeling: The target is predicting the probability of tsunami, predicting the tsunami behavior is a regression problem.

Tools
I will conduct this project using the following:
- Environment: Jupyter Notebook, Anaconda.
- Programming Language: Python
- Libraries: Matplotlib, Pandas, NumPy, Seaborn, Scikit Learn, Keras, TensorFlow, OpenCV.
- Files: TSV, CSV, NumPy arrays.

MVP Goal
- Setting the environment
- Install all packages and libraries
- Convert Tsunami.tsv to .csv and upload the data
- Analyze the data and visualize it
- Study the dependencies and relationship and exclude unrelated features (feature engineering)
- Build machine learning model
- Prepare the model input
- Train and test the model
- Measure performance and retrain if needed