DATS 6202
GROUP# 1

RESEARCH PROPOSAL
EARTHQUAKE AND TSUNAMI PREDICTION

Jack McMorrow, Alex Khater, Alejandra Mejia

INTRODUCTION

In this project, we will utilize a seismic research dataset which contains earthquake information for the past 22 years in order to develop a classification model that has the capability to predict whether or not an earthquake poses a significant risk of a tsunami.

Why Tsunami Predictions?

Although tsunamis are generally infrequent, their unpredictable nature makes them a potentially devastating natural hazard. To be able to develop accurate methods to detect them and predict them quickly could become key to saving lives. Recent research has focused on the use of artificial intelligence (AI) algorithms (Cardiff University) and deep-learning models (Los Alamos National Laboratory) combined with real-time data. Our approach will focus on using historical data to test such predictions.

Dataset

The dataset we will use has been retrieved from Kaggle and contains data records of 782 different earthquakes. The amount of earthquakes recorded, should be large enough to perform machine learning techniques on. The variables contained in the database include: magnitude, date-time, estimated mmi\cmi, multiple columns of seismic and station data, country, continent, and a binary tsunami variable that we will use as our target.

Machine Learning Techniques and Metrics

We intend to use three different Machine Learning Models for our binary classification problem: MLP, KNN, and SVM. Since this is a relatively simple binary classification problem, we can use confusion matrices and F1 scores to compare these models' performances on the data. For now we intend to use these models in standard form, but some modifications for our data may be necessary further in. All of this should be done in Python using the *scikit-learn* package and *plotly* for visualization of geospatial data.

Reference Materials

We intend to use various scholarly articles about the exact metrics our dataset holds (such as this) to make informed decisions about our project and what metrics to include.

Preliminary Work Schedule

Activities/ Tasks		Timeline (March 2023- May 2023) Starting March 27				
		W1 3/27-4/2	W2 4/3-4/9	W3 4/10-4/17	W4 4/18-4/25	W5 4/26-5/1
Act 0. Inception						
Task 0.1	Github Set up					
Task 0.2	Dataset selection					
Act1. Proposal and data cleaning						
Task 1.1	Short proposal document		**D1**			
Task 1.2	Clean dataset null values					
Task 1.3	Create explanatory graphs					
Task 1.4	Perform One Hot Encoding and PCA					
Act.2 Training and Model Implementation						
Task 2.1	Split and train dataset					
Task 2.2	Implement MLP and KNN					
Task 2.3	Create confusion matrix					
Task 2.4	Start individual reports					
Act 3. Implementation additional model and data viz						
Task 3.1	Implement SVM (if time allows)					
Task 3.2	Create Final Data Visualizations					
Task 3.3	Layout Presentation Outline					
Task 3.4	Continue Work on Individual Reports					
Act 4. Project wrapup						
Task 4.1	Fully Make and Rehearse Presentation					
Task 4.2	Finish Individual Reports					**D2**
Task 4.3	Fully Prepare Github for submission					**D3**
Task 4.4	Give Presentation on 5/1					**D4**

Deliverables schedule		Timeline (March 2023- May 2023) Starting March 27				
		W1 3/27-4/2	W2 4/3-4/9	W3 4/10-4/17	W4 4/18-4/25	W5 4/26-5/1
D1	Short Proposal document		**D1**			
D2	Individual report submission					**D2**
D3	Group report and resources submissions					**D3**
D4	Group presentation					**D4**