

ORIE 4741 Project Proposal: Data Analysis Project

https://github.com/StavrosCherpelis/ORIE_4741_Project/tree/main

Stavros Cherpelis, sgc76

Vincent Moscarelli, vsm29

Predicting Earthquake Magnitude

Question:

Can we predict earthquake magnitude using historical seismic data provided by the National Earthquake Information Center (NEIC)?

Importance:

It is critical that we improve our understanding of earthquakes and their effects in order to better prevent injuries and fatalities caused by these events. By using qualities of past earthquakes like location, depth and travel time we hope to be able to predict the size of earthquakes, so that we can determine whether or not these future events will pose a significant threat to the public. In context with surroundings, earthquake magnitude can be used to qualitatively determine the potential damages and range of the earthquake. Being able to quickly discern earthquake magnitude is critical for government institutions to coordinate effective and efficient disaster relief. Our project will attempt to create a model that can provide accurate and rapid predictions of earthquake magnitude to better enable authorities to coordinate informed earthquake disaster response.

Dataset:

Our data will be provided directly from the National Earthquake Information Center (NEIC) database. The raw data includes information on a diverse set of variables that could be useful in determining earthquake magnitude. While the NEIC has millions of data points our dataset size is variable since the NEIC database lets us customize how we choose to pull data. As we move forward with the project we will determine the most effective amount of data to use. We decided to present the dataset grouped by our features and labels.

Features:

Table 1: Features of the the dataset

Data	Description	Data Type
Time	Time the earthquake occurs	Datetime64[ns]
Latitude	North/South location	float64

Longitude	East/West location	float64
Depth	Depth of the earthquake	float64
Magnitude Type	Which scale was used, i.e. Richter, Moment, etc.	String
nst	Number of seismic stations that recorded the earthquake	float64
gap	Largest gap between recording stations, in degrees	float64
dmin	Distance to the closest seismic station	float64
rms	Root-mean-square travel time	float64
Error types	Error measurements for different quantities	float64

Labels:

Table 2: Dataset Labels

Magnitude	Size of the earthquake on a logarithmic scale	float64
-----------	---	---------

We believe this dataset will help us to predict the magnitude of future earthquakes because many of these features are scientifically linked with magnitude. For example, the largest magnitude earthquakes are commonly very deep. By leveraging known scientific principles we are confident that we can predict earthquake magnitude with this dataset.

This dataset is also mostly numeric which should lend itself well to model construction. Any non numeric features can be easily feature engineered into usable data. We also have a large enough pool of data to drop any Nan values and possible outliers. Between having lots of data, a clear path towards cleaning the data and known scientific relationships we should be able to answer our question with this dataset.