

## *Tsunami Predictions*

### *Akshay Murali*

Earthquakes happen around the world very frequently. Luckily, modern technology has allowed scientists to use sensors to collect data about these earthquakes. Using this data, it is possible to make predictions and see trends regarding earthquakes and tsunamis. Our goal in this project is to see what factors influence whether or not a particular earthquake will cause a tsunami. A model that can make predictions would be useful to quickly determine whether or not a tsunami will form.

A GIS company named Esri publishes data about recent earthquakes to their platform ArcGIS (<https://hub.arcgis.com/datasets/esri2::recent-earthquakes?geometry=-136.055%2C-68.998%2C136.055%2C81.586&layer=0>). The dataset consists of earthquakes that occurred around the world since February 2021. Each row consists of roughly 20 statistics regarding the earthquake. These statistics include measurements about the severity, time, location, and significance of the earthquake.

For our analysis, we select magnitude, longitude, latitude, depth, significance, hoursOld, and tsunami. The tsunami column contains either a 1 or 0 depending on whether or not the earthquake caused a tsunami. In 3.6% of the rows, certain statistics were missing. To explore the dimensionality of the data, a Principal Component Analysis was necessary. The PCA would not be possible with incomplete rows. Because such a small proportion of rows were incomplete, each incomplete row was dropped from the dataset before the PCA.

The blue line of **Figure 1** shows the magnitudes of earthquakes that were followed by tsunamis in February and March 2021. The orange line shows the magnitudes of earthquakes that were not followed by tsunamis in the same time frame. The time frame excluded April because a great deal of information about earthquakes that caused tsunamis in April was unavailable. You can see in the figure that the magnitudes of earthquakes that caused tsunamis was typically higher. The average magnitude of earthquakes that caused tsunamis was approx. 6.97. The average magnitude of earthquakes that did not cause tsunamis was approx. 5.53. Out of the earthquakes analyzed in this figure, approx. 53% of them caused earthquakes. The magnitudes of the earthquakes that caused tsunamis ranged from 5.3 – 8.1. The magnitudes of the earthquakes that did not cause tsunamis ranged from 3.5 – 7.2.

**Figure 2** shows the result of the Principal Component Analysis. The blue line shows that just one column captures 91.7% of the variance. This can be due to different columns having various magnitudes. To solve this, standard scaling was performed on all columns. The orange line shows that after standard scaling, four columns capture 94.3% of the variance.

Another goal is to create a model to predict whether or not a certain earthquake will cause a tsunami. We create an sklearn pipeline that consists of a StandardScaler and LogisticRegression. We perform a 75/25 train/test split, stratifying on the “tsunami” column. The model achieves a 98.7% accuracy. The recall score is 54.5, meaning the model predicts a tsunami 54.5% of the time. The precision is 66.7%.

**Figure 3** shows the coefficient weights for each of the features used by the model. It is clear that the most important factor in predicting a tsunami is the magnitude of the earthquake (weight of 2.8). The latitude of the location of the earthquake can also play an important factor (weight of 0.8). The rest of the features were almost negligible.

In conclusion, the most important factor in predicting whether an earthquake will cause a tsunami is the magnitude of the earthquake. This information is useful for civilians living in areas that are at high risk of tsunamis because it can give them a better understanding of when they should take precautions.

Figures:

