ORIE 4741 Project Proposal

Vladia Trinh (vt95), Yoanna Efimova (yie3), Toshi Tokuyama (tt426)

Can we predict which buildings in Nepal were most susceptible to earthquake damage based on building makeup and location?

Earthquakes are one of the Earth's most destructive forces. They can destroy buildings, take lives, and cost tremendous amounts of money for loss and repair. According to the National Earthquake Information Center, there are an average of 20,000 earthquakes each year. One of the most severe earthquakes was the Gorkha earthquake in Nepal in 2015, which took around 9,000 lives. In addition, thousands of people were injured, and more than 600,000 structures in the city of Kathmandu and other nearby towns were damaged and destroyed.

After the earthquake in Nepal, the land was surveyed to see how much damage each building incurred. The features in our dataset include numerical, categorical, and binary values which give us many possibilities during the feature engineering and prediction stage of our project. Some numerical features include the age of the building, height, and the number of floors. Some categorical features are already one-hot encoded. For example, the superstructure of the building is composed of multiple binary features like `adobe_mud`, `mortar_stone`, `timber`, `bamboo`, etc. Using these features describing the building construction, location, and use case, we hope to predict the `damage_grade` of the building after the earthquake. The label we want to predict is an ordinal value ranging from grade 1 to grade 3, with each grade increasing in damage.

Our problem statement is significant because the final model can also be applied to other countries that have earthquakes frequently. Although the economic level may differ among the countries, the model derived from our problem statement can be used as a strong baseline measurement for approximating building damage caused by earthquakes. Furthermore, being able to predict building damages can prevent casualties which can encourage individuals and the government to provide further protection by renovating houses or creating regulations regarding building houses.

Another application of our model could be for home/earthquake insurance resource allocation. By predicting which building will likely be damaged in the event of an earthquake, insurance companies can predict the cost of claims and which locations to prioritize.

We are likely to be successful in this endeavor since the features in our dataset are informative and the large size of our dataset allows us to train more complicated models with more confident results on our test set.

Dataset: https://www.drivendata.org/competitions/57/nepal-earthquake/page/136/