Shikha Goel & Muriel Pokol
November 18, 2019

Project 2 Proposal: Earthquake Damage Score

Motivation

On April 25th, 2015, the deadliest earthquake of the area in the past 85 years struck Nepal. It had a magnitude of 7.8 and was catastrophic for the area. Multi-story buildings fell in Kathmandu, and there were over 22,000 injured and almost 9,000 fatalities.[1] 75 percent of earthquake-related deaths are the product of collapsing structures.[2] Earthquakes occur frequently across the globe and an understanding of how to assess and manage their risks is critical.

We were particularly interested in looking at how to minimize the damage caused by earthquakes because of their unpredictability. With the current technology, it is not possible to anticipate when a major earthquake will occur, unlike most other natural disasters. This prompted us to look into how one might minimize injuries and fatalities when there is no warning.

By predicting the amount of damage to a structure by its features, we believe that the value added would be potentially minimizing injuries and lives lost by allowing for better safety regulations in high-risk buildings, once those buildings are identified. This would be particularly helpful in places such as Kathmandu, a seismically active area where there is a large array of building types.

Earthquake Dataset

Our dataset is from DataDriven.com and contains information about approximately 260,000 building structures and the amount of damage that they faced during the 2015 earthquake in Nepal. The damage is represented by an ordinal scale of 1 to 3, with 1 representing minimal damage, 2 representing a moderate amount of damage, and 3 representing almost total destruction of the building. Buildings are denoted by a unique identifier, and features include:

- Geographic information
- Number of floors
- Age of the building in years
- Area and Height of the building footprint
- Surface condition of the land where the building is located
- Foundation and Roof types
- Ground floor type, Floor type for floors used in higher levels
- Position of building
- Building plan configuration
- Binary features to denote the type of structure (mortar, cement, brick, timber, etc.)
- Legal ownership status
- Number of families that live in the building
- Building use (Government, school, etc)

---

1. https://www.worldvision.org/disaster-relief-news-stories/2015-nepal-earthquake-facts
2. http://www.iitk.ac.in/nicee/wcee/article/10_vol10_5989.pdf

Project Goal and Plan

The goals of our project are to:

- Identify the important building features that contribute to its robustness/fragility in the event of an earthquake.
- Given a set of property features, predict the extent of earthquake damage.

Project Plan:

- This is a classification task. The data already has labels indicating the extent of damage and falls under the realm of supervised learning.
- Our **task (T)** would be the classification of building damage into low damage (scores of 1) or higher damage (scores of 2s and 3s).This could be a binary or multi-class classification problem. We hope to explore different types of classification models such as Logistic Regression, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis(QDA), Naive Bayes, and Random Forests.
- The **experience (E)** - Dividing the dataset into medium and large training, validation and test datasets that will allow us to pick the best model and then improve P on the selected model using the larger training set.
- The **performance (P) -** We want our classifier to be able to identify the buildings at risk of high damage (class 3) with high accuracy. We care more about false negatives being low (where negative = low damage, positive = high damage) where buildings are being incorrectly classified as being low risk when their true score would be higher damage during an earthquake. A high false negative rate would lead to high losses in real life. However, a false positive would likely result in a building facing stricter safety procedures and, although might be an inconvenience, should not result in any unnecessary losses of life.
- (Note - It can be argued that from a home-builder/seller's perspective, a higher false positive would drive up the cost or reduce property prices. If the scope and duration of the project allow, we would like to explore this angle and find a weighted performance metric for multi-class classification and assign appropriate weights to different classes based on their importance.)

Business Value

The outcome of this analysis is important to multiple stakeholders:

- Government - Knowing which buildings are at risk in the event of an earthquake will allow the government to take proactive measures to impose regulatory standards on those buildings. This will allow the government to save more lives and reduce the reactive disaster relief expenditure due to reduced damage.
- Consumers buying houses - Information on the importance of various property characteristics can allow a consumer to take more informed decisions while purchasing property, especially in earthquake-prone countries.
- Websites like Zillow, 99accres etc can use this model to assign an earthquake safety score to the listings on their website thereby increasing the website's appeal to consumers in real estate.