

Stefan Bobick
Alberto Espinosa
Michael Sexton

Team Name: Team Disaster

Problem Space

We want to use past historic information to predict at-risk areas and approximate spending in disaster areas for resource allocation in the US and other countries around the world. We would aim to approximate the amount of funding that different countries would need annually and assist in government spending in order to better budget for disaster relief. We will also take into account how climate change affects the data over time to better budget for the future as different types of disasters may have become more prevalent than they used to be in the past.

This is an important question to answer as the climate begins to change around the world. Disasters will only become more frequent and a machine learning algorithm and model could greatly assist in how countries spend their relief budget. This idea arose when browsing different data sets and wondering what would be a real pressing question that can be solved using data.

There are a couple of reasons why it would be useful to solve this problem using machine learning. First, there are several groups and organizations like the UN, the Red Cross, Research Institutes, and press agencies that have been recording data from disasters for a long time. There are organizations like the Centre for Research on the Epidemiology of Disasters (CRED) that have gathered such data which would allow us to create a ML model to try to predict resource allocation more efficiently. Secondly, due to the sheer amount and complexity of data that has been gathered throughout the years created a perfect opportunity to create an ML model that can interpret said data more efficiently.

Data & Data Plan

- Critical features within the data set include: date & location of natural disaster, total people affected, injured or homeless, number of deaths, total damage induced (USD), reconstruction cost, and emergency declarations
 - Planning to exclude features such as: external IDs, in-depth locational data such as latitude/longitude, entry dates & update dates, etc.
- Outcome variables/expected results: We are looking to gauge different levels of impact and their relationship to certain types of natural disasters. With this information, we are

looking to predict accurate amounts of funding needed to aid struck communities and alleviate the total impact put forth on said communities by these disasters.

- This would be supervised learning, since we are finding patterns within fully observed data in the past few decades, with the goal of predicting partially observed data - funding needed, magnitude of damage, total impact, etc.
- How many samples?
 - Roughly 16,000 recorded disasters from 1999 until 2024. More samples utilized as we go farther back in time.
 - The samples gathered meet one of the following criteria:
 1. At least ten deaths (including dead and missing).
 2. At least 100 affected (people affected, injured, or homeless).
 3. A call for international assistance or an emergency declaration.
- Our current data set:
 - The data set we are using is the EM-DAT, an open-source database compiled by the CRED that “contains data on the occurrence and impacts of over 26,000 mass disasters worldwide from 1900 to the present day.” (EMDAT). The database is continuously revised and updated.
 - <https://public.emdat.be/>