Project Pitch

Problem Space

Problem:

Applying Machine Learning methods to predict future disaster risk, resource allocation, and financial impact for an optimized disaster relief.

Disaster Risk Prediction:

Accurately predicting the time, location and impact of disasters is a hard task.

Our ML model can use the historical EM-DAT data to identify patterns based on features such as the type of disaster, the time, and the region where it happened to predict future disasters.

Possible ML techniques that would help us solve this part of the problem would be the use of a time series analysis, and a classification model like a random forest or a gradient booster.

Resource Allocation:

Optimize the precision of resource allocation of areas affected by disasters based on the type of it, and the previous costs and resources used for them.

Our ML model can use the historical EM-DAT financial data, such as the reconstruction cost, the insured damaged cost, and the total damage cost to predict the required resources for different types of disasters.

Possible ML techniques that would help us solve this part of the problem would be the use of regression models to predict financial needs, and reinforcement learning and decision trees for more effective resource allocation.

Financial Impact:

Estimate the total financial impact of a disaster more accurately.

Our ML model can use the historical EM-DAT financial data, as well as other financial data to predict the financial impact of disasters based on the number of affected people, the region, and the number of homeless after the disaster would help create a better budgeting plan for disasters.

Possible ML techniques that would help us solve this part of the problem would be the use of regression models to predict financial needs, and Support Vector Machines for identifying outliers.

Data Set

Scope of the data:

We are using several outside financial data sets to support our model which is generally focused on the EM-DAT dataset. The EM-DAT dataset “is compiled from various sources, including UN agencies, non-governmental organizations, reinsurance companies, research institutes, and press agencies”. It contains 15,984 entries that classify and provide statistics for disasters around the world from the year 2000 to 2024.

Each entry has 46 different features which classify the disaster by group, type, and then move into important statistics such as total deaths, number affected. The dataset also contains features for aid provided by the Office of US Foreign Disaster Assistance (OFDA) or the Bureau of Humanitarian Assistance (BHA) and how much was contributed.

The rows below are those that we find most significant and will play the strongest role in our model creation. We intend to break down the different disaster types and analyze the impact that aid from the government could’ve made by looking at an aggregate of total affected, total deaths, number injured and number homeless, weighting death with the most priority as number affected is ambiguous. Ideally, we would like to create a model that would estimate the amount of spending necessary to save lives and reduce the total number of affected people around the world to better budget for potential disaster relief.

Disaster Group	Disaster Subgroup	Disaster Type	Disaster Subtype	External IDs	Event Name	ISO	Country	Su
Natural	Climatological	Drought	Drought			DJI	Djibouti	Su
Natural	Climatological	Drought	Drought			SDN	Sudan	Nu
Natural	Climatological	Drought	Drought			SOM	Somalia	Su
Technological	Transport	Road	Road			AGO	Angola	Su
Natural	Hydrological	Flood	Riverine flood			AGO	Angola	Su
Natural	Meteorological	Extreme temperature	Cold wave			BGD	Bangladesh	Su
Technological	Transport	Road	Road			BRA	Brazil	La

Fig 1. The disaster types and how they are categorized.

Total Deaths	No. Injured	No. Affected	No. Homeless	Total Affected
		100000		100000
		2000000		2000000
21		1200000		1200000
14	11			11
31		70000		70000
49				
42	71			71

Fig 2. Number affected by the disaster and statistics for death toll.

OFDA/BHA Response	Appeal	Declaration	AID Contribution ('000 US$)
Yes	No	No	1556
No	No	No	
No	No	No	
No	No	No	
No	No	Yes	
No	No	No	
No	No	No	

Fig 3. Response by the government and how much monetary aid was contributed.

Approach

We believe that these different models could work for our problem:

- Time series models
    - Analyze sequential features in order to identify patterns or trends. Forecasting when and where a type of disaster may occur based on the historical data.
- Regression Models
    - Compare two or more features and identify strong or weak correlation between the many features of the dataset.
    - Use them to predict the financial impact or need for resources based on features such as the type of disaster, region, and number of affected people.
- Neural Networks
    - Location-based predictions. Are certain locations experiencing higher frequencies of certain types of disasters? Can they be predicted efficiently?
    - Help identify any long-term dependencies between the features.
- Random Forests and Decision trees
    - Map out different features of past disasters to identify which feature gives the most information gained in order to predict outcomes like the amount of resources or financial support needed after a certain type of disaster.
- Support Vector Machines
    - Use this classification method to find the optimal boundaries to help identify outliers within the disaster dataset.

Comparison to other research done in the problem space:

Many current ideas in the research field involve applying machine learning and deep learning methods on predicting and assessing natural disaster risk. Some focus on subsets of natural disasters, such as this study below:

- https://www.research-collection.ethz.ch/handle/20.500.11850/398592

They applied various machine learning methods to calibrate tropical cyclone impacts, first defining regions based on socio-economic indicators through feature selection, dimensionality reduction and clustering algorithms. They then utilized Bayesian optimization methods to calibrate region specific damage functions to estimate damage from incoming/predicted storms. This is great for specific regions who experience this one type of disaster, but we would like to push it to a broader scale with all natural disasters that pose risks to various countries.

This research paper focuses on recent developments within pre-disaster management, utilizing the latest improvements in IoT, machine learning and big data:

- https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10100463&tag=1

This study mostly accumulates various approaches in working on pre-disaster management and the new technologies experimented with. They noticed the biggest challenge being implementing these new technologies in countries with poor technological infrastructure. Our goal for this project would be to not only predict future natural disasters but provide optimal resource allocation towards countries classified as more in need.

Additional research links:

- Leveraging machine learning algorithms for improved disaster preparedness and response through accurate weather pattern and natural disaster prediction
- The Disaster Relief Fund: Overview and Issues
- [PDF] Machine Learning in Disaster Management: Recent Developments in Methods and Applications | Semantic Scholar
- Investigating the non-linear impacts of seven types of natural disasters on inbound tourism: Insights from the EM-DAT database
- https://models-simsuite.hub.arcgis.com/ - model for estimating damages as a result of hurricanes and tropical storms

Stretch Goals:

- Integrating a dataset consisting of global GDP values for each country. We would then utilize this data and identify countries that aren't as financially well-off who are also experiencing an influx in disaster frequency and magnitude. With these patterns, our goal would be to allocate financial resources optimally, meaning countries who experience a higher frequency of disasters or a greater magnitude of damage while also being lower on the GDP scale would receive a greater portion of the collected funds versus a country experiencing little disasters with a higher GDP.

- Utilizing another dataset with various non-profit entities currently fundraising for disaster relief. We would then compare these contributions to privatized entities' monetary contributions to see which are more effective. This would help determine where money would best be spent and donated by the general public and other groups of interest.

- Examining climate change impacts on disaster frequency, determining if there is an increase in the frequency of a certain type of disaster and if there is a correlation between the frequency of that disaster and damages inflicted by it.