Genki Hirayama, Hamdi Kucukengin, Maddy Lok, Brian Duckworth

MSDS 422 | October 27, 2024

**Final Project Midpoint: Estimating Economic Losses from Disasters**

**Executive Summary**

This project developed a predictive model to estimate economic losses from disasters, addressing the urgent need for accurate forecasts to aid resource planning and disaster response. Using machine learning, the project focused on key variables—disaster type, location, magnitude, date, and fatalities—to create a model that enhances decision-making for organizations facing financial impacts of catastrophic events. After evaluating multiple models, linear regression proved the most effective based on R-squared and RMSE metrics, offering both accuracy and interpretability. The resulting tool enables stakeholders to make proactive, data-driven decisions, supporting better preparedness and resource management in the face of disasters.

**Problem Statement**

Disasters cause substantial economic damage, but accurately predicting these losses remains challenging. Traditional methods often overlook critical factors like disaster type, location resilience, magnitude, and fatalities, limiting preparedness and response efficiency. This study seeks to leverage machine learning to identify the key factors that most significantly impact fatalities and economic damage in order to decide which areas are subject to facing more negative impacts and would need a higher allocation of resources.

**Research Objectives**

This project aims to develop a predictive model to estimate disaster-related economic losses based on the following factors: disaster type, location, magnitude, date, and fatalities. By accurately modeling these variables, the tool will provide actionable insights to enhance disaster preparedness and resource planning.

**Exploratory Data Analysis**

This EDA was primarily focused on gaining insights into how disaster characteristics influence certain aspects such as fatalities and economic factors. We first started by analyzing the types of disasters and identified the fatalities and economic loss broken down by disaster types in order to display which disaster type tends to cause the most severe outcomes. We also developed a time analysis on the occurrences and frequencies of disasters to display which months each disaster type helps to occur more frequently. This allows for predictive analysis to determine when certain resources should be utilized. We also explored the relationships between key variables such as disaster magnitudes. With certain disasters, it is easier to predict the impact on economic loss or fatalities based on magnitude.

**Data Preparation & Feature Engineering**

In our analysis, we first identified any inconsistencies or missing values to prepare the data for analysis. Then, we created a feature representing the frequency of each disaster type by location and another feature that calculates the ratio of economic loss to fatalities. We also performed a log transformation to reduce the skewness and normalize the data, making it easier for the model to interpret. We also standardized key features to maintain consistency and improve model performance to ensure that features contribute equally to model decisions without being disproportionately influenced by scale discrepancies. We also conducted variance inflation factor analysis to identify multicollinearity among predictors. We utilized tree-based methods to determine which predictors have the strongest association with the target variables. All of these processes are important for ensuring the dataset is optimized for modeling and accuracy.

**Resources**

Forecasting Disaster Management in 2024 <https://www.kaggle.com/datasets/umeradnaan/prediction-of-disaster-management-in-2024/data>