

Global Disaster Response Analysis (2018–2024)

Evaluating Recovery Time Using Disaster Characteristics and Aid Factors

Authors:

Group 20 — Lois, Eric, Jisu, Vania

1. Introduction

Natural disasters vary widely in severity, response efficiency, and recovery duration. Understanding which factors are associated with faster or slower recovery is important for evaluating disaster response effectiveness and informing future aid allocation.

In this project, we analyze a global disaster dataset covering events from 2018 to 2024. We focus on the following research question:

How well can recovery time be predicted using disaster severity, response characteristics, and aid-related variables?

2. Data

We use the *Global Disaster Response Analysis (2018–2024)* dataset from Kaggle. The dataset contains approximately 50,000 disaster events across 20 countries and 10 disaster types.

Key variables include:

- Disaster severity and casualties
- Economic loss and aid amount
- Response time and response efficiency
- Recovery duration (target variable)

The data were split temporally:

- **Training set:** January 2018 – December 2022
- **Test set:** January 2023 – December 2024

This temporal split avoids information leakage and reflects real-world forecasting conditions.

3. Methods

We model recovery time (in days) as a regression problem. The following models were evaluated:

- Baseline model (predicting the training mean)
- Ridge regression
- Lasso regression

- Random Forest regression

Numerical variables were standardized, and categorical variables (country and disaster type) were one-hot encoded using a preprocessing pipeline. Model performance was evaluated on a held-out test set.

```
import pandas as pd
metrics = pd.read_csv("outputs/ml/metrics.csv")
metrics

      model      RMSE      MAE      R2
0    ridge    4.978634  3.970246  0.939425
1  random_forest  5.093356  4.059898  0.936602
2     lasso    5.099404  4.062804  0.936451
3 baseline_mean  20.228813 16.355676 -0.000025

metrics = metrics.rename(columns={"R2": "R^2"})
metrics

      model      RMSE      MAE      R^2
0    ridge    4.978634  3.970246  0.939425
1  random_forest  5.093356  4.059898  0.936602
2     lasso    5.099404  4.062804  0.936451
3 baseline_mean  20.228813 16.355676 -0.000025
```

4. Results

Model performance was evaluated using RMSE, MAE, and R^2 . Lower RMSE and MAE indicate better predictive accuracy.

Among the evaluated models, **Ridge regression achieved the lowest RMSE**, suggesting that recovery time exhibits largely linear relationships with the available predictors.

The baseline model performed substantially worse, indicating that the models capture meaningful signal beyond the average recovery duration.

Compared to the baseline mean predictor (RMSE around 20.23 days), Ridge reduces RMSE to around 4.98 days, indicating substantial predictive improvement.

5. Discussion

Although Ridge regression performed best, the RMSE is about 5 days. Given a median recovery time of approximately 50 days, this corresponds to an average error on the order of about one week, indicating moderate uncertainty.

This suggests that while the included variables provide partial explanatory power, recovery duration is influenced by additional unobserved factors such as policy decisions, infrastructure resilience, and governance capacity.

The results indicate that recovery dynamics are complex and only partially captured by the available disaster and response metrics.

6. Limitations and Future Work

This analysis relies on aggregated disaster-level data and does not account for country-specific institutional differences or long-term recovery policies.

Future work could incorporate:

- Country-level socioeconomic indicators
- Time-varying policy responses
- More granular spatial or temporal data

More advanced modeling approaches could also be explored, though interpretability remains an important consideration.

Author Contributions

- **Lois:** Data acquisition, exploratory data analysis, research question formulation
- **Eric:** Model selection, machine learning implementation
- **Jisu:** Model evaluation, reproducibility support, results interpretation
- **Vania:** Repository structure, documentation, website deployment