

Data Science < PREDINA >

Prepared by

| Student name | Student ID |
|-------------------------|------------|
| Nouf AlMansour | 444200525 |
| Sara AlOqiel | 444203016 |
| Shahad AlMutairi | 444200935 |

Supervised by:
Dr. Mashael AlDayel

1 Introduction

Natural disasters occur unpredictably, often leaving significant devastation in their wake. Leveraging data-driven approaches in disaster mitigation can enhance preparedness, improve early warning systems, and reduce the impact on affected communities. While these events may not always be preventable, early warning indicators can play a crucial role in mitigating losses and enhancing protective measures. This study seeks to provide a deeper understanding of natural disasters, emphasizing proactive strategies to improve preparedness.

This research addresses the question: "What are the regional patterns, financial impacts, and evolving trends of natural disasters over the past century, and what insights do these provide for future risk management and preparedness?"

In the following section, we will discuss the data sources utilized for this study.

2 Data sources

- **EM-DAT: The International Disaster Database**

The EM-DAT Public Table is a global disaster database maintained by CRED, tracking natural and technological disasters. It includes data on fatalities, affected populations, and economic damages, and is used for research and disaster management.

- **Kaggle Dataset: ALL NATURAL DISASTERS 1900-2021 / EOSDIS**

This dataset, hosted on Kaggle, provides a record of natural disasters worldwide from 1900 to 2021, sourced from NASA's Earth Observing System Data and Information System (EOSDIS). It includes details such as disaster type, location, dates, and impacts (e.g., fatalities, affected populations, and economic damages), making it useful for analyzing historical disaster trends and impacts.

- **Global Disaster Alert and Coordination System (GDACS)**

The GDACS API provides real-time alerts on natural disasters such as earthquakes, tsunamis, and storms, offering data on disaster type, location, magnitude, and impact. It is useful for monitoring, coordinating disaster response efforts, and supplementing data for building predictive models.

3 Data Collection and Integration

- **Observations:** The integrated dataset encompasses disaster records from 1900 to 2025, potentially exceeding 100,000 observations due to its extensive historical coverage.
- **Features and Data Types:**
 - **Categorical:** Disaster type, country, affected region
 - **Numerical:** Fatalities, affected population, economic damages, magnitude (for earthquakes, storms, etc.)
 - **Temporal:** Date of occurrence, duration
 - **Geospatial:** Latitude, longitude of the disaster even

3.1 Evaluation of Potential Biases

1. Representation:

- Some disasters may be underreported, particularly in low-income regions where reporting infrastructure is limited.
- Certain regions and disaster types may be overrepresented due to better monitoring systems in developed countries.

2. Measurement Bias:

- The datasets aggregate data from multiple sources, reducing the likelihood of systematic measurement bias.
- Variations in economic damage estimation methodologies across countries may introduce inconsistencies.
- Differences in data collection methods over time and across sources may affect comparability but are mitigated through standardized reporting frameworks.

3. Historical:

- Older records may be incomplete or inconsistent due to technological and administrative limitations in past decades.
- The dataset may reflect historical inequalities, as wealthier nations tend to have more comprehensive records and faster response mechanisms, potentially skewing impact analysis.

In the next section, we will outline the objectives of this study.

4 Objectives

This study aims to uncover the following:

- What factors influence the intensity and impact of natural disasters in specific regions?
- Which regions are most affected by specific types of natural disasters?
- What are the estimated losses associated with natural disasters based on current data and forecasts?
- Which regions have a greater chance in expecting earthquakes?

In the next section, we will outline the objectives of this study.

5 Methodology

To uncover patterns and trends in natural disasters, we need to apply a combination of data science techniques.

- Analyzing correlation and feature importance will help determine the factors influencing the intensity and impact of natural disasters.
- Geospatial analysis will be employed to identify regions most affected by specific natural disasters.
- Regression analysis will be utilized to estimate financial losses based on various factors such as region and disaster type.
- Exploratory Data Analysis (EDA) will be conducted to assess which regions have a higher probability of experiencing earthquakes, as historical earthquake data serves as a key forecasting tool.

These methods will facilitate the extraction of meaningful insights from the data.

In the next section, we will examine challenges in data collection and provide recommendations for improvement.

6 Challenges in Data Collection and Recommendations

6.1 Challenges in Data Collection:

- **Missing or Incomplete Data:**

Some regions, particularly low-income countries, may have gaps in historical records due to weak reporting systems.

- **Standardization Issues:**

Variability in measurement units (e.g., currency for economic damages, population reporting methods) may introduce inconsistencies.

- **Access Restrictions:**

Some high-quality disaster datasets may require paid access or institutional affiliations, limiting data availability.

6.2 Recommendations for Future Data Collection and Analysis:

- **Adopt Standardized Reporting Protocols:**

Encourage the use of internationally recognized frameworks, such as EM-DAT classification criteria, to ensure consistency across sources.

- **Integrate Remote Sensing Data:**

Utilize satellite imagery and other geospatial data sources to supplement on-the-ground disaster reports and improve coverage in underreported regions.

- **Promote Open Data Initiatives:**

Advocate for greater transparency and accessibility of disaster-related datasets to enhance data availability and completeness.

This concludes our discussion on natural disasters, their data sources, methodologies, and challenges. Future research should focus on refining predictive models and improving data integration for more accurate disaster preparedness strategies.

