

Python for Data Analysis

- Documentation

Landslide after Rainfall

QIAN Victor - RAMPONT Martin



Landslides are natural disasters that cause significant loss of life and property worldwide 🌍. Understanding and predicting the impact of landslides is crucial for mitigating their devastating consequences. The **Global Landslide Catalog, provided by NASA 🚀**, offers a comprehensive repository of landslide events globally, containing crucial data that can be harnessed to predict the potential human toll of such disasters.

In this project, we delve into the realm of machine learning 🤖 using Python to analyze the Global Landslide Catalog Export dataset. **Our primary objective is to develop a predictive model that anticipates the number of fatalities resulting from landslides based on various parameters associated with these events.** By leveraging machine learning algorithms, we aim to uncover patterns and relationships 🧬 within the data to create a predictive framework that can assist in early warning systems and disaster management efforts.

Predicting the number of fatalities stemming from landslides is **not merely a statistical exercise**; it's a proactive step toward mitigating the impact of these catastrophic events. This project seeks to harness the power of data-driven insights 🌟 to potentially save lives and enhance preparedness measures in regions prone to landslides.

Dataset:

The Global Landslide Catalog Export dataset originates from the Global Landslide Catalog (GLC), a comprehensive initiative established by NASA's Goddard Space Flight Center in 2007. The GLC's primary objective is to systematically document rainfall-triggered landslide events worldwide, irrespective of their scale, geographical impact, or specific attributes. This dataset represents an

amalgamation of [various types of rainfall-induced mass movements](#) reported across media outlets, disaster databases, scientific publications, and other reliable resources.

Key Dataset Features:

source_name	News entity reporting the event
source_link	Website URL of the news source
event_id	Identification number for the event
event_date	Date of the landslide event
event_time	Time of the landslide event
event_title	Title of the news story
event_description	Description of the event
location_description	Description of the event's location
landslide_category	Categorization of the landslide
landslide_trigger	Cause of the landslide
landslide_size	Size classification of the landslide
landslide_setting	Environment where the landslide occurred
fatality_count	Number of fatalities
injury_count	Number of injuries
storm_name	Associated storm name (if applicable)
photo_link	URL link to related photos
notes	Additional notes about the event
country_name	Name of the country where the event occurred
country_code	Code representing the country
admin_division_name	Administrative division name
admin_division_population	Population of the administrative division
longitude	Longitude coordinate of the event location
latitude	Latitude coordinate of the event location
submitted_date	Date of submission of the event information
created_date	Date of dataset entry creation
last_edited_date	Date of the last edit made to the entry

Project Structure Overview:

The project is hosted on GitHub at [Python-data-analysis](#), involving a collaborative effort by the two of us, each managing an individual branch. The repository comprises several folders, files, and notebooks tailored to various functionalities and components essential for the Python data analysis project.

Folders:

1. **app:** This directory houses files central to the development of a Streamlit application. It includes:
 - **__init__.py:** Initialization file.

- **project_structure**: Description or documentation of the project's structure.
 - **app.py**: Primary file responsible for creating the application.
 - **data_analysis.py**: File dedicated to data analysis for the application.
 - **documentation.py**: File handling documentation aspects of the application.
 - **homepage.py**: File concerning the application's homepage.
 - **map.py**: File managing map-related functionalities within the application.
 - **ml_models.py**: File incorporating machine learning models for the application.
2. **data**: This directory contains the crucial Global Landslide Catalog (GLC) CSV file, which serves as the primary dataset for the analysis.
 3. **model**: This folder includes the **model.py** file, presumably housing code pertaining to machine learning model creation and training.
 4. **src/img**: This folder stores images utilized within the project.

Files:

- **.gitignore**: File specifying intentionally untracked files to be ignored by Git.
- **README.md**: The repository's main documentation or guide for users/contributors.
- **exploration.ipynb**: Jupyter Notebook showcasing exploration activities with potentially interactive elements.
- **landslide_prediction.ipynb**: Jupyter Notebook likely dedicated to landslide prediction, possibly showcasing model implementation and evaluation.
- **yesterday**: A file with an unclear purpose or content.
- **requirements.txt**: File specifying necessary dependencies or packages for the project.

Results:

