

Project-Earthquake Prediction Model using Python

Understanding the Problem Statement :

- The problem is to develop a machine learning model for earthquake prediction using a dataset from Kaggle's repository.
- The primary objective is to explore and understand key features in earthquake data, create visualizations for global earthquake distribution, split the data into training and testing sets, and build a neural network model to predict earthquake magnitudes based on given features.

Goal:

- The primary goal of this project is to predict the probability of occurrence of earthquake to mitigate the ill effects caused by the earthquake.
- It can be achieved by our neural network model.

Understanding the Design Thinking :

Data Source Selection:

- The first step is to import the earthquake dataset downloaded from Kaggle .
- The dataset contains features such as date, time, latitude, longitude, depth, and magnitude.

Data Preprocessing:

- Handle Missing Data: If missing data values are present in the dataset , then try to remove or impute it .
- Data Formatting: Convert data types as needed, especially date and time features, which should be converted into datetime objects for analysis.
- Outlier Handling: Identify outliers in the dataset, which could adversely affect model performance.

Feature Exploration:

- Exploratory Data Analysis (EDA) should be conducted to understand the distribution, central tendencies, and variability of each feature.
- Identification of target variable in our dataset .
- Calculate and visualize correlations between features and the target variable (earthquake magnitude) to identify relationships.

Visualization:

- Data visualization libraries such as matplotlib and seaborn is used to build histograms, scatter plots and correlation matrices to provide clearer understanding of the features in the dataset.
- A world map visualization depicting the frequency distribution of earthquakes globally is useful for identifying earthquake prone regions visually.

Data Splitting:

- The dataset is split into training and testing sets.
- A common practice is to allocate 80% of the data for training and 20% for testing .

Model Development:

- **Neural Networks** machine learning model is used to predict the earthquake magnitudes .
- The neural network architecture should be designed by specifying the number of hidden layers, units, activation functions, and any regularization techniques (e.g., dropout) to be used.

Training and Evaluation:

- Train the neural network model using the training data and set suitable hyperparameters.
- Monitor the training process, track metrics (e.g., mean squared error , accuracy , precision ,correlation matrix), and visualize training/validation loss to check for overfitting.
- Evaluate the performance of the model using appropriate evaluation metrics such as Mean Squared Error and R-squared.

Flowchart:

