

NAAN MUDHALVAN

PROJECT : Earthquake Prediction Model using Python

PHASE 2

TEAM MEMBERS:

Srijeet Goswami - 2021504044

Akash R - 2021504501

Gokul E – 2021504514

Balamurugan M – 2021504507

PROBLEM DEFINITION:

The problem is to develop an earthquake prediction model using a Kaggle dataset. The objective is to explore and understand the key features of earthquake data, visualize the data on a world map for a global overview, split the data for training and testing, and build a neural network model to predict earthquake magnitudes based on the given features.

DESIGN APPROACH:

1. Choosing appropriate data source
2. Feature selection
3. Visualization
4. Data splitting
5. Model development
6. Model training
7. Model evaluation

8. Iterative refinement

9. Deployment

1. DATA SOURCE:

Choosing a suitable Kaggle dataset containing earthquake data with features like date, time, latitude, longitude, depth, and magnitude.

Dataset:

<https://www.kaggle.com/datasets/gustavobmgm/earthquakes-for-ml-prediction-new-version>

Context:

All earthquakes from 1973 until 2021, worldwide are available. Contains basic information such as location, time and magnitude. The model to be deployed is to be trained using this data, following instructions from here.

Acknowledgements: The United States Geological Survey (USGS)

2. KEY FEATURES REQUIRED:

- **Tectonic Plate Boundaries:** The majority of earthquakes occur along tectonic plate boundaries. Studying the interactions between these plates, such as convergent, divergent, and transform boundaries, is fundamental for understanding earthquake potential.
- **Geological History and Paleo seismology:** Studying the geological history of an area and identifying evidence of past seismic events through paleo seismology is essential for assessing the recurrence interval of earthquakes.
- Tabulating the places having a history of encountering earthquakes till available records and recording their geographic coordinates like latitude, longitude, altitude etc.

3. VISUALIZATION:

Create a world map visualization to display earthquake frequency distribution.

4. DATA SPLITTING:

In machine learning, datasets are often divided into two subsets: a training set and a testing (or validation) set. The training set is used to train a machine learning model, while the testing set is used to assess the model's performance and generalization to unseen data.

5. MODEL DEVELOPMENT:

Build a neural network model for earthquake magnitude prediction. Building a neural network involves designing its architecture, defining the number of layers, neurons, and activation functions. Data is pre-processed, normalized, and divided into training and testing sets.

6. MODEL TRAINING:

The model is trained using optimization algorithms to adjust weights and minimize the loss function. Regularization techniques are applied to prevent overfitting.

7. MODEL EVALUATION:

Post-training, the model's performance is assessed on the testing set.

8.ITERATIVE REFINEMENT:

Iterate through the process, refining features and tuning hyperparameters to improve the model's accuracy and reliability. This iterative process underpins neural network applications in machine learning, image and speech recognition, and natural language processing, where models learn patterns and relationships in data inspired by the human brain's interconnected neurons.

9. DEPLOYMENT AND MONITORING:

Once we achieved a model with great accuracy, we will deploy it for real-time applications. The model is updated by monitoring its performance over time.

FLOWCHART:

