Problem Definition:

To Develop a machine learning model for earthquake magnitude using a dataset from kaggle.The goal is to build a neural network model that can accurately predict the magnitude of earthquakes .This predictive model will assist in assessing the potential impact of earthquakes in different regions.

Design Thinking:

1. **Data Source**: Choosing a suitable Kaggle dataset containing earthquake data with features like date, time, latitude, longitude, depth, and magnitude. The kaggle dataset **Significant Earthquakes, 1965-2016** provided by the National Earthquake Information Center (NEIC) has been chosen. (https://www.kaggle.com/datasets/usgs/earthquake-database)
2. **Feature Selection:** Identifying which features are most relevant for earthquake magnitude prediction. Features like latitude, longitude, depth, and historical earthquake data are crucial.
3. **Feature Transformation**: Performing transformations such as scaling, encoding categorical variables, and extracting additional features like seasonality or proximity to fault lines.
4. **Visualization**: Creating a world map visualization to display the distribution of earthquake frequencies. Colour coding or markers to represent the magnitude of earthquakes at different geographical locations. This visualization can help to easily identify earthquake-prone regions.
5. **Data Splitting**: Splitting the dataset into a training set and a test set. Typically, an 80-20 or 70-30 split is used, where the majority of data is allocated for training and a smaller portion for testing.
6. **Model Development:** Building a neural network model for earthquake magnitude prediction. Libraries like TensorFlow and PyTorch are used to create the model. Also the input features and output features(earthquake magnitude) are defined .
7. **Training and Evaluation**: Training the neural network model on the training set. Also the training process is monitored and adjustments are made as required. Then the model's performance is evaluated by using the test set. For earthquake

prediction, metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or R-squared are used to assess how well the model predicts earthquake magnitudes.

Innovation

Data preprocessing steps, and feature exploration techniques.

- **Steps involved in KNN:**
  - Choosing an appropriate value for k, which represents the number of nearest neighbors to consider. We also use cross-validation techniques to find the optimal k value.
  - Implementing the KNN algorithm using a machine learning library like scikit-learn in Python.
  - Training the KNN model using the training dataset, with the earthquake locations (latitude and longitude) as input features and potentially other relevant features.
  - Evaluating the model's performance using metrics such as accuracy, precision, recall, F1-score, and ROC AUC on the testing dataset.

- **Steps involved in finding validity of data/ confidence of data**:
  - Certain features like Magnitude Type, Magnitude Error, Azimuthal Gap, Horizontal Distance, Horizontal Error, Root Mean Square, Source will help to detect the validity levels of the magnitude and depth information used in KNN.
  - Hence this can be found by detecting the outliers using **Support Vector Machine** or by providing confidence values that help the ensemble model decide.

- **Data Preprocessing:**
  - To guarantee the quality and appropriateness of the data for training the model, data preparation is an essential step. The tasks are broken out as follows:
    - Managing Missing Values:The dataset contains missing values like in magnitude error. In order to solve this, we will fill in the missing data points using the proper imputation techniques, such as mean or median imputation.
    - Normalization and Standardization: We will normalize the data to make sure that all features have comparable scales. This

procedure stops the model from being dominated by one feature while it is being trained.
        - Data division: The dataset will be split into two sections: around 70% for training and 30% for testing. This allows us to train the model on one subset and evaluate its performance on another, ensuring that the model can generalize well to unseen data.

- **The ensemble model:**
    - We are planning to use models like AdaBoosting that can combine multiple weak learners and provide a result that predicts the magnitude of the earthquake along with the confidence.

Approaches used during the development

- Prediction of earthquake methods generally consists of two methods.
    - Precursor method
    - Trend or Pattern

- Precursor method - identifying distinctive *precursors* to earthquakes.
        - **Dilatancy–diffusion**
        - **Changes in V_p/V_s**
        - **Radon emissions**
        - **Electromagnetic anomalies**

- Trend or Pattern - identifying some kind of geophysical *trend* or pattern in seismicity that might precede a large earthquake.
        - Date
        - Time
        - Latitude
        - Longitude
- Since we are using a kaggle dataset which consists of data that corresponds with the second method(location ,time) we are going to use that method.
- Using the location, time and date we are planning to create a machine learning model that will predict earthquakes.
- We are planning to use Ensemble model with multiple ML models that include KNN,SVM .

Dataset Used:

Significant Earthquakes, 1965-2016
https://www.kaggle.com/datasets/usgs/earthquake-database

- **Context**

  The National Earthquake Information Center (NEIC) determines the location and size of all significant earthquakes that occur worldwide and disseminates this information immediately to national and international agencies, scientists, critical facilities, and the general public. The NEIC compiles and provides to scientists and to the public an extensive seismic database that serves as a foundation for scientific research through the operation of modern digital national and global seismograph networks and cooperative international agreements. The NEIC is the national data center and archive for earthquake information.

- **Content**

  This dataset includes a record of the date, time, location, depth, magnitude, and source of every earthquake with a reported magnitude 5.5 or higher since 1965