

Earthquake prediction model using python

Artificial intelligence phase-3 [development]

Team member

210221104039 : M.THIRISHA

OBJECTIVE : To Load and Preprocess our dataset and begin our building of the Earthquake prediction model using python

Data Source: Kaggle dataset containing earthquake data with features like date, time, latitude, longitude, depth, and Magnitude

Dataset Link: <https://www.kaggle.com/datasets/usgs/earthquake-database>

To load and preprocess a dataset for building an Earthquake magnitude prediction model in Python, you can follow these steps:

1. Data Loading:

Load the earthquake dataset into your Python environment. You can use the Pandas library to read data from a CSV file. For example:

```
import pandas as pd
#Load the dataset from a CSV file
Earthquake_data=pd.read_csv(' earthquake_data.csv' )
```

2. Data Exploration:

Explore the dataset to understand its structure and quality. Check for missing values, outliers, and get a sense of the data by using Pandas functions like ``head()`, `info()`, and `describe()``.

3. Feature Engineering:

Create relevant features for earthquake magnitude prediction. These features could include seismic sensor readings, geographic information, and historical earthquake data. This often requires domain knowledge in geophysics.

4. Data Preprocessing:

Prepare the data for modeling by performing preprocessing steps. Here are some common preprocessing tasks:

a. missing values:

```
Earthquake_data = Earthquake_data.dropna() #Remove rows with missing values
```

b. Scaling and normalization (if necessary):

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
Earthquake_data[['feature1', 'feature2']] = scaler.fit_transform(earthquake_data[['feature1', 'feature2']])
```

c. categorical variables (if applicable):

```
Earthquake_data = pd.get_dummies(earthquake_data, columns=['categorical_feature'])
```

d. selection (select the most relevant features):

```
from sklearn.feature_selection import SelectKBest, f_regression
```

```
X = earthquake_data.drop('magnitude', axis=1)
Y = earthquake_data['magnitude']
```

```
Selector = SelectKBest(score_func=f_regression, k=5) #Select top 5 features
X_new = selector.fit_transform(X, Y)
```

5. Train-Test Split:

Split the data into training and testing sets to evaluate the model's performance. You can use Scikit-learn for this:

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X_new, y, test_size=0.2, random_state=42)
```

Conclusion : *Thereby we loaded our dataset which we have gotten from kaggle using pandas library and preprocessed the dataset for precise prediction of Earthquake when we build machine learning models like linear regression, random forest etc,.*