EARTHQUAKE PREDICTION USING PYTHON

TEAM MEMBERS:

SHOBADEVI M

SOWNDHARYA R

YUVASRI M

RAMYA S

VYSHALI M

AIM:

To preprocess the given dataset by the kaggle and perform the loading of data , preprocessing and analysing.

Loading of data set:

import pandas as pd

data = pd.read\_csv('database.csv')

print(data.head())

Preprocessing of dataset:

print(data.isnull().sum())

data = data.drop(['Depth Error', 'Depth Seismic Stations', 'Magnitude Error', 'Magnitude Seismic Stations', 'Azimuthal Gap', 'Horizontal Distance', 'Horizontal Error', 'Root Mean Square'], axis=1)

data['Magnitude Type'].fillna(data['Magnitude Type'].mode()[0], inplace=True)

data['Type'].fillna(data['Type'].mode()[0], inplace=True)

data['Magnitude Source'].fillna(data['Magnitude Source'].mode()[0], inplace=True)

data['Status'].fillna(data['Status'].mode()[0], inplace=True)

data['Location Source'].fillna(data['Location Source'].mode()[0], inplace=True)

data['Magnitude'].fillna(data['Magnitude'].mean(), inplace=True)

Output:

Date 0

Time 0

Latitude 0

Longitude 0

Type 0

Depth 0

Magnitude 0

Magnitude Type 0

ID 0

Source 0

Location Source 0

Magnitude Source 0

Status 0

dtype: int64

label\_encoders = {}

categorical\_cols = ['Type', 'Magnitude Type', 'Source', 'Location Source', 'Magnitude Source', 'Status']

for col in categorical\_cols:

le = LabelEncoder()

data[col] = le.fit\_transform(data[col])

label\_encoders[col] = le

numeric\_cols = ['Latitude', 'Longitude', 'Depth', 'Magnitude']

scaler = StandardScaler()

data[numeric\_cols] = scaler.fit\_transform(data[numeric\_cols])

data.to\_csv('preprocessed\_earthquake\_data.csv', index=False)

Analysis:

Descriptive analysis

descriptive\_stats = data.describe()

print(descriptive\_stats)

Magnitude analysis:

import seaborn as sns

import matplotlib.pyplot as plt

# Histogram of Magnitudes

plt.hist(data['Magnitude'])

plt.xlabel('Magnitude')

plt.ylabel('Frequency')

plt.title('Distribution of Earthquake Magnitudes')

plt.show()

# Count of Magnitude Types

magnitude\_counts = data['Magnitude Type'].value\_counts()

print(magnitude\_counts)

Scatter Analysis:

# Scatter plot of Latitude and Longitude

plt.scatter(data['Longitude'], data['Latitude'], c=data['Magnitude'], cmap='viridis')

plt.xlabel('Longitude')

plt.ylabel('Latitude')

plt.title('Earthquake Locations')

plt.colorbar(label='Magnitude')

plt.show()

Status Analysis:

# Count of different Status values

status\_counts = data['Status'].value\_counts()

print(status\_counts)

# Box plot comparing Magnitudes for different Status

plt.figure(figsize=(8, 6))

sns.boxplot(x='Status', y='Magnitude', data=data)

plt.xlabel('Status')

plt.ylabel('Magnitude')

plt.title('Magnitude Distribution by Status')

plt.show()