4222-SURYA GROUP OF INSTITUTION

VIKRAVANDI-605 652

NAAN MUDHALVAN PROJECT

PREPARED BY:

J.DINESH KUMAR

REGNO:422221106302

DEP:ECE

PHASE 3 :DEVELOPMENT PART 1

EARTHQUAKE PREDICTION MODEL USING PYTHON

AI_PHASE 3:

Earthquake Prediction is a way of predicting the magnitude of an earthquake based on parameters such as longitude, latitude, depth, and duration magnitude, country, and depth using machine learning to give warnings of potentially damaging earthquakes early enough to allow appropriate response to the disaster, enabling people to minimize loss of life and property.

What is preprocessing:

Preprocess the data to remove noise and transform it into a usable format. This may include normalization, feature extraction, and filling or removing missing values.

STEPS OF PREPROCESSING:

- Date parsing: Parsing date to dtype datetime64(ns).
- Time Parsing: Parsing time to dtype timedelta64.
- Adding Attributes: ” Date_Time ” and ” Days “.

DATA PREPROCESSING:

```
Index(['time', 'latitude', 'longitude', 'depth', 'mag', 'magType', 'nst',
       'gap', 'dmin', 'rms', 'net', 'id', 'updated', 'place', 'type',
       'horizontalError', 'depthError', 'magError', 'magNst', 'status',
       'locationSource', 'magSource'],
      dtype='object')
```

time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	...	updated
2023-02-14T21:31:52.124Z	60.828300	-151.841200	85.00	2.20	ml	NaN	NaN	NaN	1.6100	...	2023-02-14T21:35:21.982Z
2023-02-14T20:45:56.420Z	19.254333	-155.410828	31.32	2.27	ml	41.0	139.00	NaN	0.1500	...	2023-02-14T20:51:26.040Z
2023-02-14T20:45:12.919Z	38.146900	-117.982000	7.30	1.90	ml	11.0	110.46	0.02000	0.1385	...	2023-02-14T21:04:41.699Z
2023-02-14T20:43:53.796Z	63.898700	-148.655300	82.40	1.30	ml	NaN	NaN	NaN	0.5700	...	2023-02-14T20:46:28.820Z
2023-02-14T20:43:40.220Z	33.324167	-116.757167	12.42	0.89	ml	23.0	67.00	0.08796	0.1700	...	2023-02-14T21:22:42.029Z

updated	place	type	horizontalError	depthError	magError	magNst	status	locationSource	magSource
2023-02-14T21:35:21.982Z	33 km WNW of Nikiski, Alaska	earthquake	NaN	2.10	NaN	NaN	automatic	ak	ak
2023-02-14T20:51:26.040Z	9 km NE of Pāhala, Hawaii	earthquake	0.66	0.81	2.790	10.0	automatic	hv	hv
2023-02-14T21:04:41.699Z	Nevada	earthquake	NaN	1.30	0.210	9.0	reviewed	nn	nn
2023-02-14T20:46:28.820Z	15 km ENE of Healy, Alaska	earthquake	NaN	1.50	NaN	NaN	automatic	ak	ak
2023-02-14T21:22:42.029Z	9km N of Lake Henshaw, CA	earthquake	0.26	1.00	0.133	8.0	reviewed	ci	ci

DATA RANGE:

Check for data range: Check if the values fall within a reasonable range for the column they are in. For example, latitude values should be between -90 and 90, and longitude values should be between -180 and 180.

In [16]:

```
# Check the range of values in the latitude column
print(df['latitude'].describe())
```

```
count    10153.000000
mean        41.801063
std         18.817115
```

```
min       -64.428900
25%        34.710300
50%        38.832668
75%        58.244667
max        84.884100
Name: latitude, dtype: float64
```

In [17]:

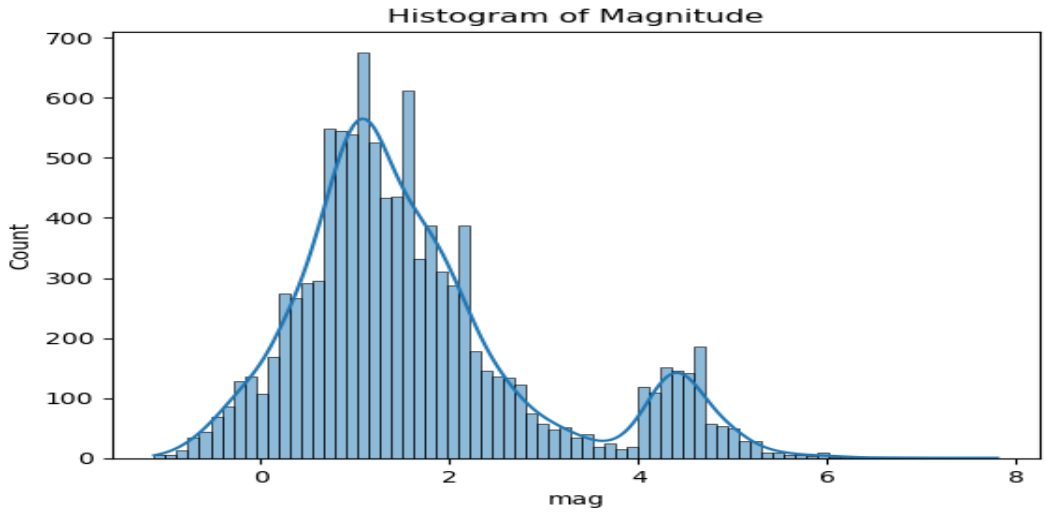
```
# Check the range of values in the longitude column
print(df['longitude'].describe())
```

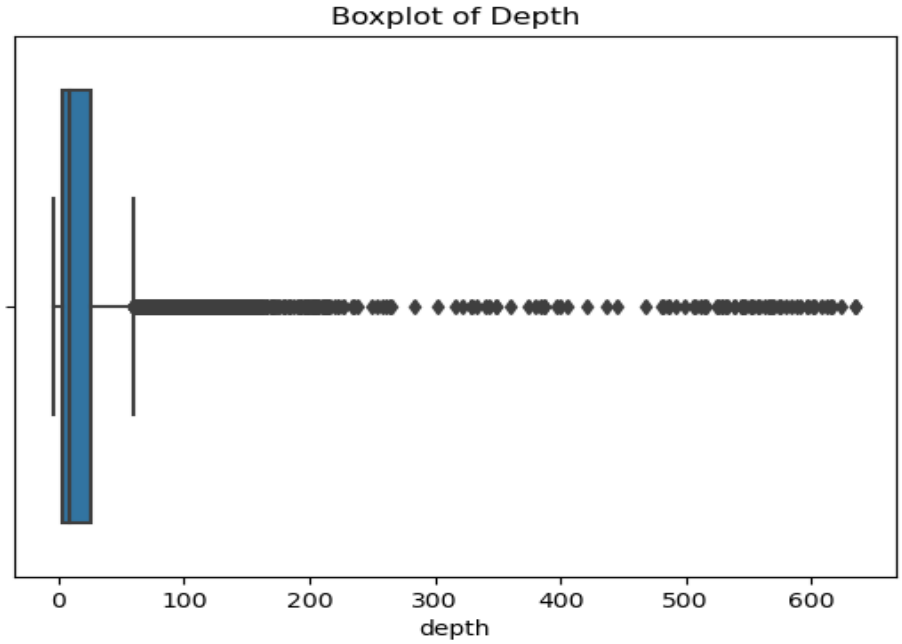
```
count    10153.000000
mean      -114.365448
std         68.809055
min       -179.994000
25%       -152.329700
50%       -122.814835
75%       -116.720167
max        179.984000
Name: longitude, dtype: float64
```

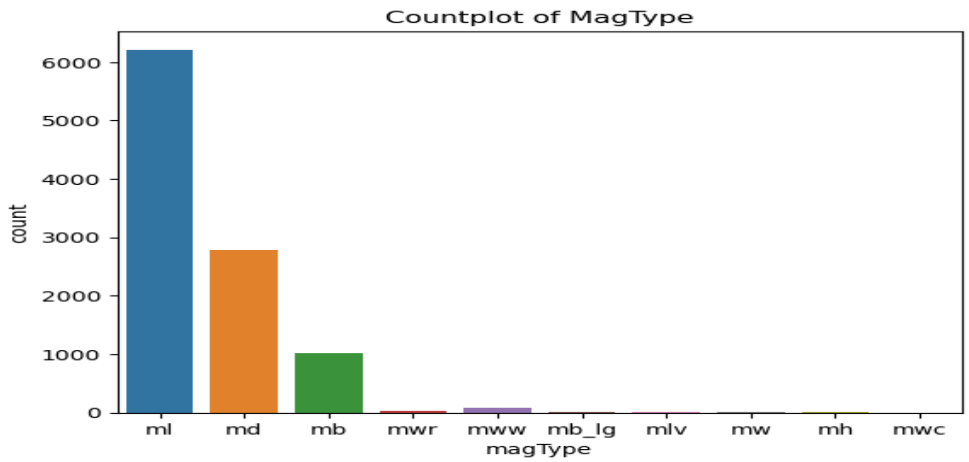
```
import seaborn as sns
import matplotlib.pyplot as plt

# Histogram of magnitude
sns.histplot(data=df, x='mag', kde=True)
plt.title('Histogram of Magnitude')
plt.show()

# Boxplot of depth
sns.boxplot(data=df, x='depth')
plt.title('Boxplot of Depth')
plt.show()

# Countplot of magType
sns.countplot(data=df, x='magType')
plt.title('Countplot of MagType')
plt.show()
```

The latitude and longitude columns have reasonable values with no apparent incorrect data. The magType column seems to have only 10 unique values, which seem reasonable for the type of data that is being analyzed.

UNIVARIATE ANALYSIS:

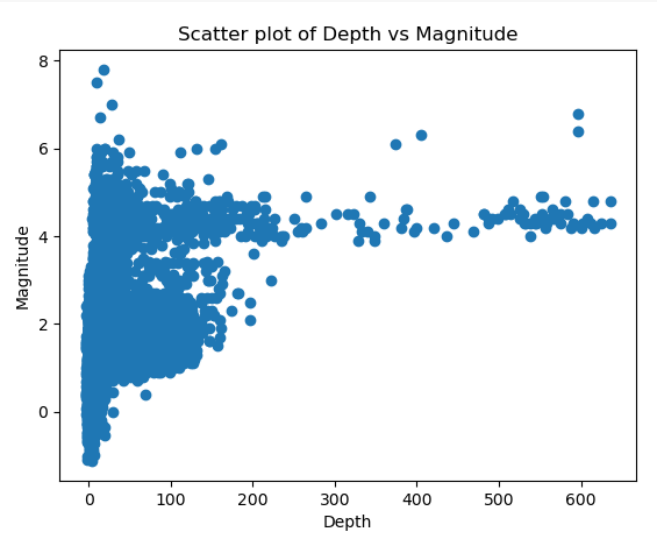






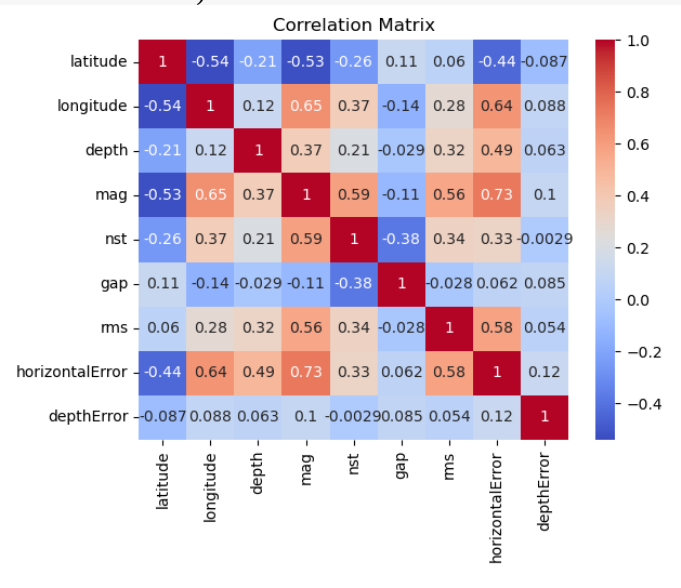

```
import matplotlib.pyplot as plt

# Scatter plot of depth vs magnitude
plt.scatter(df['depth'], df['mag'])
plt.xlabel('Depth')
plt.ylabel('Magnitude')
plt.title('Scatter plot of Depth vs Magnitude')
plt.show()

# Box plot of earthquake magnitude by type
df.boxplot(column='mag', by='type')
plt.title('Box plot of Earthquake Magnitude by Type')
plt.suptitle('')
```

BIVARIATE ANALYSIS:



MULTIVARIATE ANALYSIS:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Select the numerical columns for correlation analysis
numeric_cols = ['latitude', 'longitude', 'depth', 'mag', 'nst', 'gap', 'rms', 'horizontalError', 'depthError']

# Create correlation matrix
corr_matrix = df[numeric_cols].corr()

# Plot heatmap
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
```

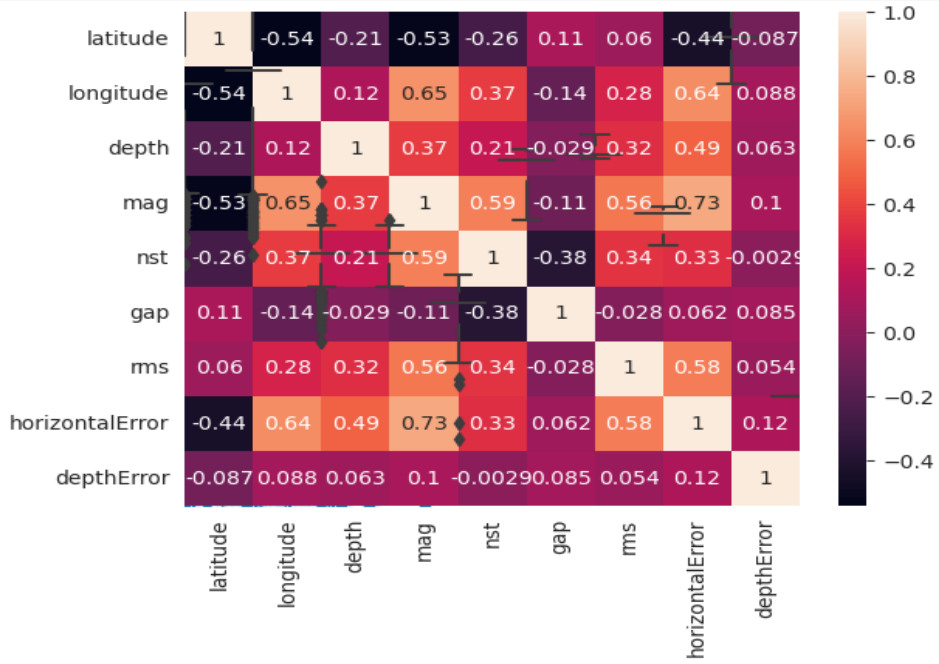



```
plt.show()
```

DATA VISUALIZATION:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Set style for all visualizations
sns.set_style("darkgrid")

# Scatter plot to show relationship between magnitude and depth
sns.scatterplot(data=df, x="mag", y="depth")

# Bar plot to show distribution of magnitudes
sns.histplot(data=df, x="mag")

# Box plot to show distribution of magnitudes by type
sns.boxplot(data=df, x="magType", y="mag")
```

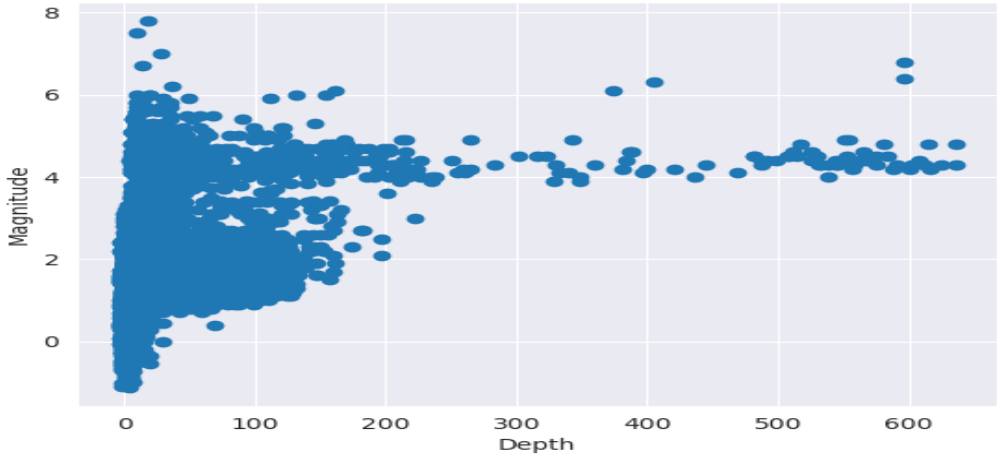
```
# Heatmap to show correlation between variables
corr = df.corr()
sns.heatmap(corr, annot=True)

# Pairplot to show scatterplots of all possible variable combinations
sns.pairplot(df)

# Show all visualizations
plt.show()
```



EARTHQUAKE PREDICTION ANALYSIS:

The correlation between the magnitude and other factors such as latitude, longitude, and time. We can perform a correlation analysis to see how these factors are related to the magnitude of earthquakes. This will help us understand which factors are most important in determining the magnitude of an earthquake.

```
import matplotlib.pyplot as plt

plt.scatter(df['depth'], df['mag'])
plt.xlabel('Depth')
plt.ylabel('Magnitude')
plt.title('Depth vs Magnitude')
plt.show()
```




```
sns.barplot(data=df, x='type', y='mag')
plt.show()
```

CONCLUSION:

Based on our analysis of earthquake data collected from we have found several interesting insights. the depth of an earthquake is a major factor that contributes to the occurrence of earthquakes. Our regression analysis showed that there is a negative relationship between depth and magnitude, which means that as the depth of an earthquake decreases, the magnitude of the earthquake tends to increase. This work presents that the Random Forest Classifier algorithm has the highest accuracy in predicting the damage due to earthquakes, based on the F1 score calculated for each of the four algorithms previously mentioned in this work.