Prediction of Intensity of Earthquakes

Team:
- **Jaya Prakash Lavdyavath (YD00257)**
- **Prashanthi Ponakalla (FH16094)**
- **Saivaraprasad Settipalli (EP53338)**

Abstract:

The object of this work is to predict the intensity of Earthquakes that occurred between the period of 1906 to 2023 globally. To predict the intensity of earthquake damage in advance and improve the effectiveness of earthquake emergency measures, this paper proposes the activity of performing multiple Machine learning models for the prediction of the trend of ground motion intensity using the factors like Magnitude and Depth. The input data is taken from a dataset that is already existing and contains real-time facts on earthquakes. Using the factors such as Latitude, Longitude, and depth of earthquake's occurrence, the magnitude of earthquakes is predicted. Moreover, performing the various models using the ML algorithm gives the best method that provides an accurate value as a prediction.

Introduction:

One of the most disastrous natural occurrences is an earthquake. After the occurrence of an earthquake, gathering information on building damage and other effects allows for efficient recovery. This research identifies cutting-edge data sources for estimating building damage and offers suggestions for more effective data collection[1](*Earthquake Hazards Program | U.S. Geological Survey*, 2023). Depending on the questions that these data are to answer, different criteria must be considered while choosing a certain data source or data gathering method for seismic reconnaissance. Several data sources should support one another, confirm gathered data, or systematically assess the damage. As evidenced by the recent rise in the number of crowdsourcing and SM platforms used to collect seismic reconnaissance data, this data source is likely to become more and more significant[2](*World Earthquake Data From 1906-2022*, 2023). Between 1902 and 2011, these occurrences were responsible for more than 23 million fatalities in addition to significant physical, social, economic, institutional, cultural, and environmental losses. Fieldwork or ground surveys, omnidirectional imaging (OD), terrestrial laser scanning (TLS), remote sensing (RS), crowdsourcing platforms, social media (SM), and closed-circuit television films are among the current data collection techniques that have been divided into seven categories (CCTV)[3](*Stanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI*, 2019). To prepare statistics, calibrate and evaluate engineering models, and detect the design flaws that result in subpar structural performance, earthquake reconnaissance permits the collection of perishable data on building performance. Five closely connected topics are involved in damage detection and characterization: structural health monitoring (SHM), condition monitoring (CM), non-destructive evaluation (NDE), statistical process control (SPC), and damage prognosis(DP). To understand the characteristics of the earthquake and its impact on the social, economic, institutional, cultural, and environmental dimensions, researchers could thus collect useful data from citizens, avoiding knowledge gaps that would otherwise allow for the spread of rumors and false information.

As mentioned earlier, one of the most devastating natural disasters, earthquakes occur suddenly and shake the surface of the earth. Buildings and infrastructure are both harmed by earthquakes, which have an impact on daily living[4](*Contreras et al., 2021*). Machine learning can be quite useful for making early predictions about how an earthquake will affect a region, and this necessitates the originality of the work. Six different machine learning classifiers—Artificial Neural Network, Random Tree, CHAID, Discriminant, XGBoost Tree, and TreeAS—were used on six datasets from various regions of India to classify this perception. It has been noted that the occurrence of large earthquakes typically occurs in cycles, with a long aseismic interval—which may continue for decades or even centuries—occurring between two seismic occurrences that follow one another. Seismographs can record the movement of the free surface during the seismic period, which only lasts for a few seconds or a few minutes at most. Seismic waves of various sorts caused by earthquakes cause significant disruptions in the area. It has been noted that in seismically active places, absorption levels of deliquescing minerals and the gassy integrant of subsurface water stay practically constant during the seismically passive time. Except for the earthquake dataset from Gujarat, XGBoost performed quite well in each dataset[5]*(Jozinović et al., 2020*). Artificial Neural Network and Tree-AS both performed well and provided results for applying this model to predict the seismic impact of upcoming earthquakes, together with XGBoost. Random tree and discriminant analysis didn't always work properly. The Indian earthquake can be predicted in large part thanks to machine learning. With the use of earlier seismic datasets, an ML model has been created for detecting the magnitude range. This analysis predicts a magnitude value. Together with the prediction curve fitting between the magnitude and depth of the Andaman & Nicobar dataset, magnitude, and depth are essential features that should be considered if the earthquake occurs.

Methods:

To retrieve the best value of magnitude depending on the values of latitude and longitude, initially, the activity of data collecting, data cleaning, and transformation of data is performed to make the data error-free and make the dataset ready for modeling. In addition, machine learning models such as Linear Regression, OLS, and Random Forest Regression are performed to test in all the possible ways and retrieve the best model for the dataset.

a) **Linear Regression:**

Linear regression is a type of supervised learning algorithm used for predicting a continuous target variable based on one or more input variables. In the context of predicting earthquake data, linear regression can be used to model the relationship between the magnitude of an earthquake and its corresponding ground motion intensity. The model would then be trained to find the best-fit line that represents the relationship between earthquake magnitude and ground motion intensity. This line would be based on the training data and would be used to make predictions on the test data.

Once the model is trained, it can be used to predict the ground motion intensity of future earthquakes based on their magnitude. However, it's important to note that linear regression is only effective if there is a strong linear relationship between the input variables and the target variable, and it may not be the most accurate method for predicting earthquake data.

b) OLS(Ordinary Least Squares):

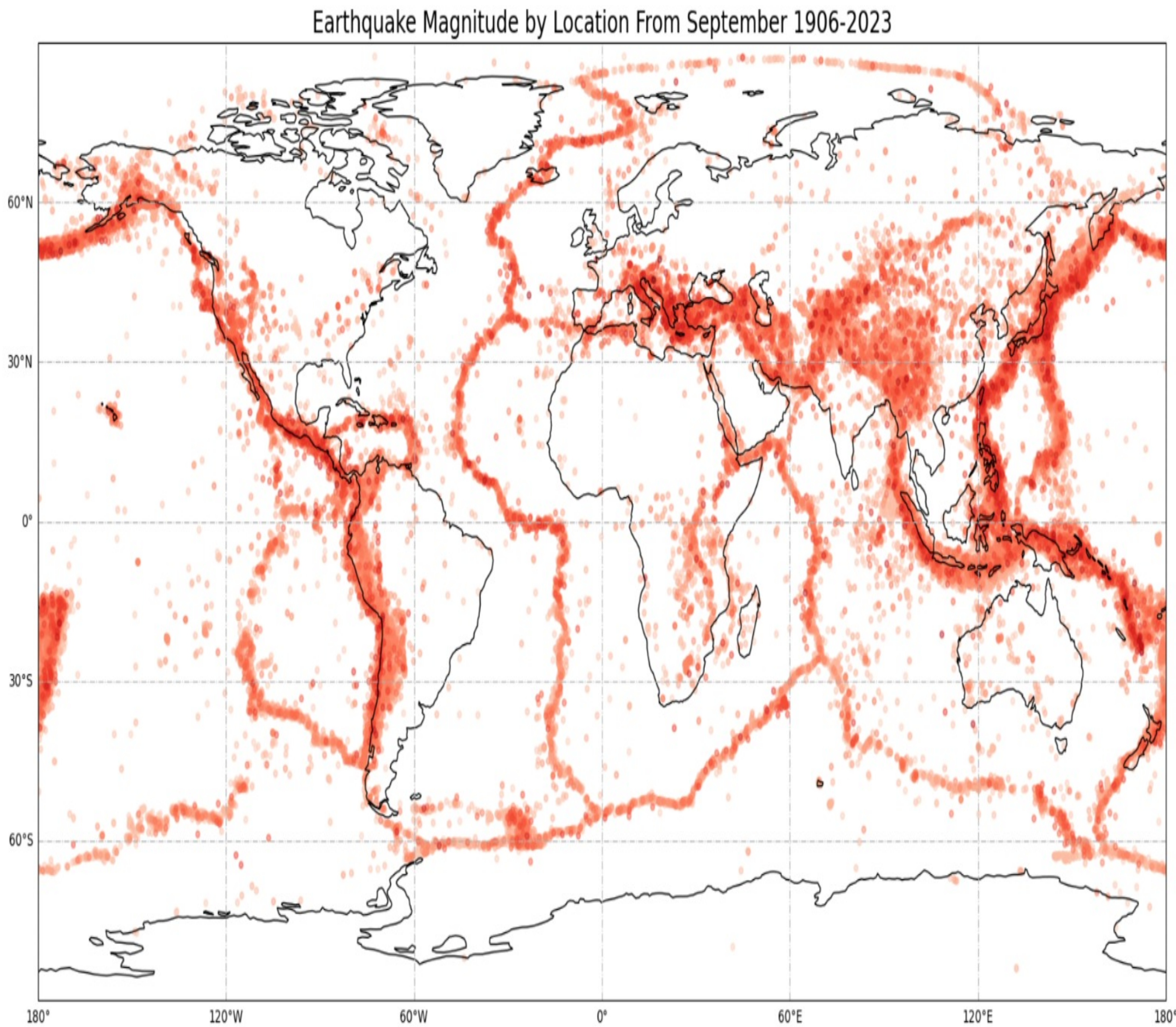
OLS stands for Ordinary Least Squares, which is a method used in linear regression analysis to find the best-fit line that describes the relationship between two variables, such as earthquake magnitude and ground motion intensity. In the context of earthquake data prediction, OLS is used to estimate the parameters of a linear regression model that can be used to predict ground motion intensity based on earthquake magnitude.

To use OLS to predict earthquake data, we would first collect a dataset containing information on earthquake magnitude and ground motion intensity, split the data into training and testing sets, and then fit a linear regression model to the training data using the OLS method.

c) Random Forest Regression:

Random Forest Regression is a machine learning algorithm that uses decision trees to predict a continuous target variable, such as ground motion intensity in earthquake prediction. It works by creating a large number of decision trees and then combining their predictions to make a final prediction. To use Random Forest Regression to predict earthquake data, we would first collect a dataset containing information on both earthquake magnitude and ground motion intensity. The data would be split into training and testing sets, and then a Random Forest Regression model would be trained on the training data.

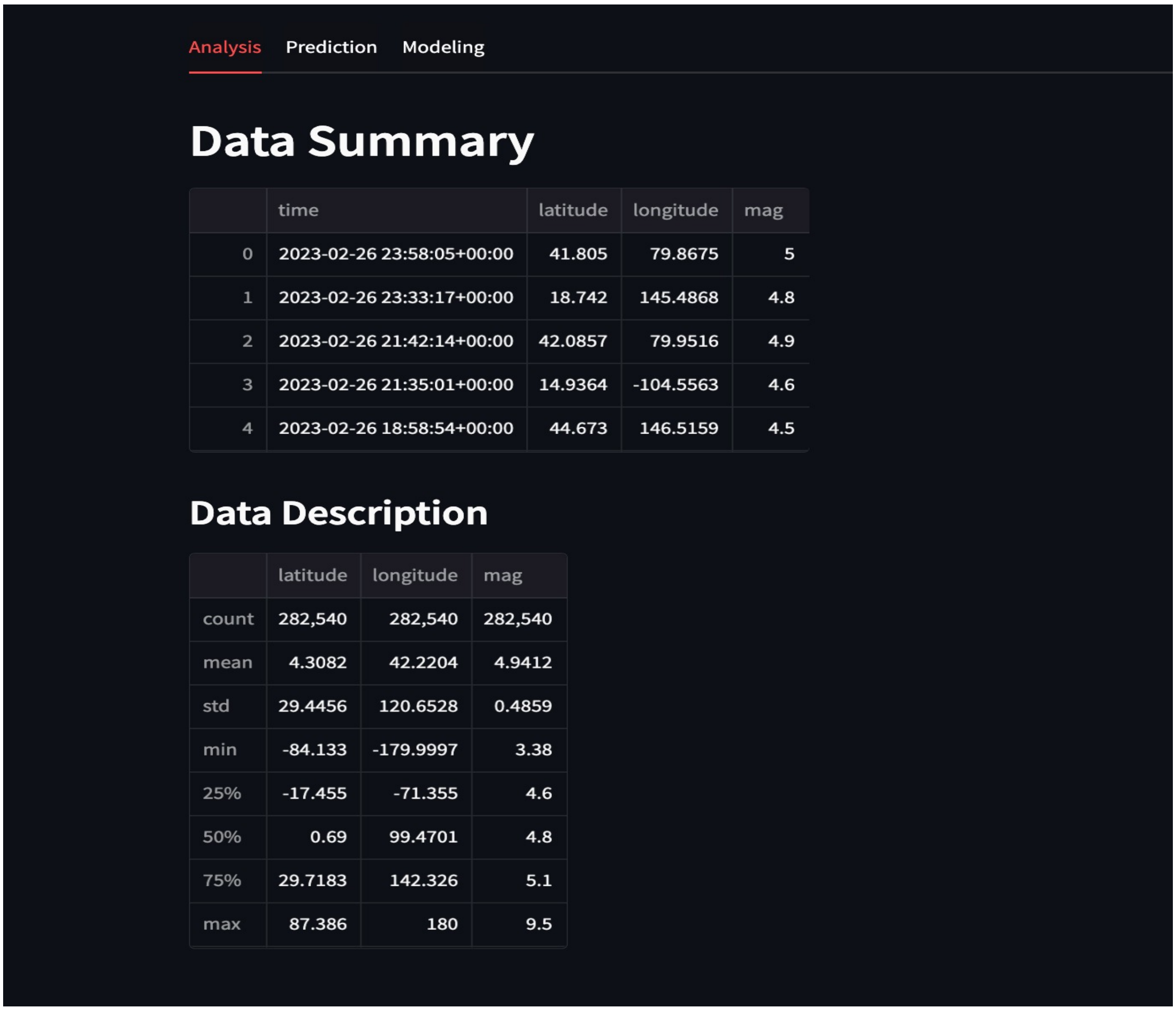
The Random Forest Regression model works by creating a large number of decision trees, each trained on a different subset of the training data, and using them to make predictions on the testing data. The final prediction is made by averaging the predictions of all the decision trees.



Results & Analysis:

After performing a sequence of activities on the dataset to retrieve the best output, the following is the result and its analysis:

1) **Analysis of numerical data:**



Data Summary

	time	latitude	longitude	mag
0	2023-02-26 23:58:05+00:00	41.805	79.8675	5
1	2023-02-26 23:33:17+00:00	18.742	145.4868	4.8
2	2023-02-26 21:42:14+00:00	42.0857	79.9516	4.9
3	2023-02-26 21:35:01+00:00	14.9364	-104.5563	4.6
4	2023-02-26 18:58:54+00:00	44.673	146.5159	4.5

Data Description

	latitude	longitude	mag
count	282,540	282,540	282,540
mean	4.3082	42.2204	4.9412
std	29.4456	120.6528	0.4859
min	-84.133	-179.9997	3.38
25%	-17.455	-71.355	4.6
50%	0.69	99.4701	4.8
75%	29.7183	142.326	5.1
max	87.386	180	9.5

The above tables show the analysis of numerical data such as latitude, longitude, and magnitude. It also describes the values of these factors with different variations.

2) Prediction of values magnitude with latitude and longitude as shown in the below tables.

Analysis Prediction Modeling

	0
time	datetime64[ns, UTC]
latitude	float64
longitude	float64
mag	float64

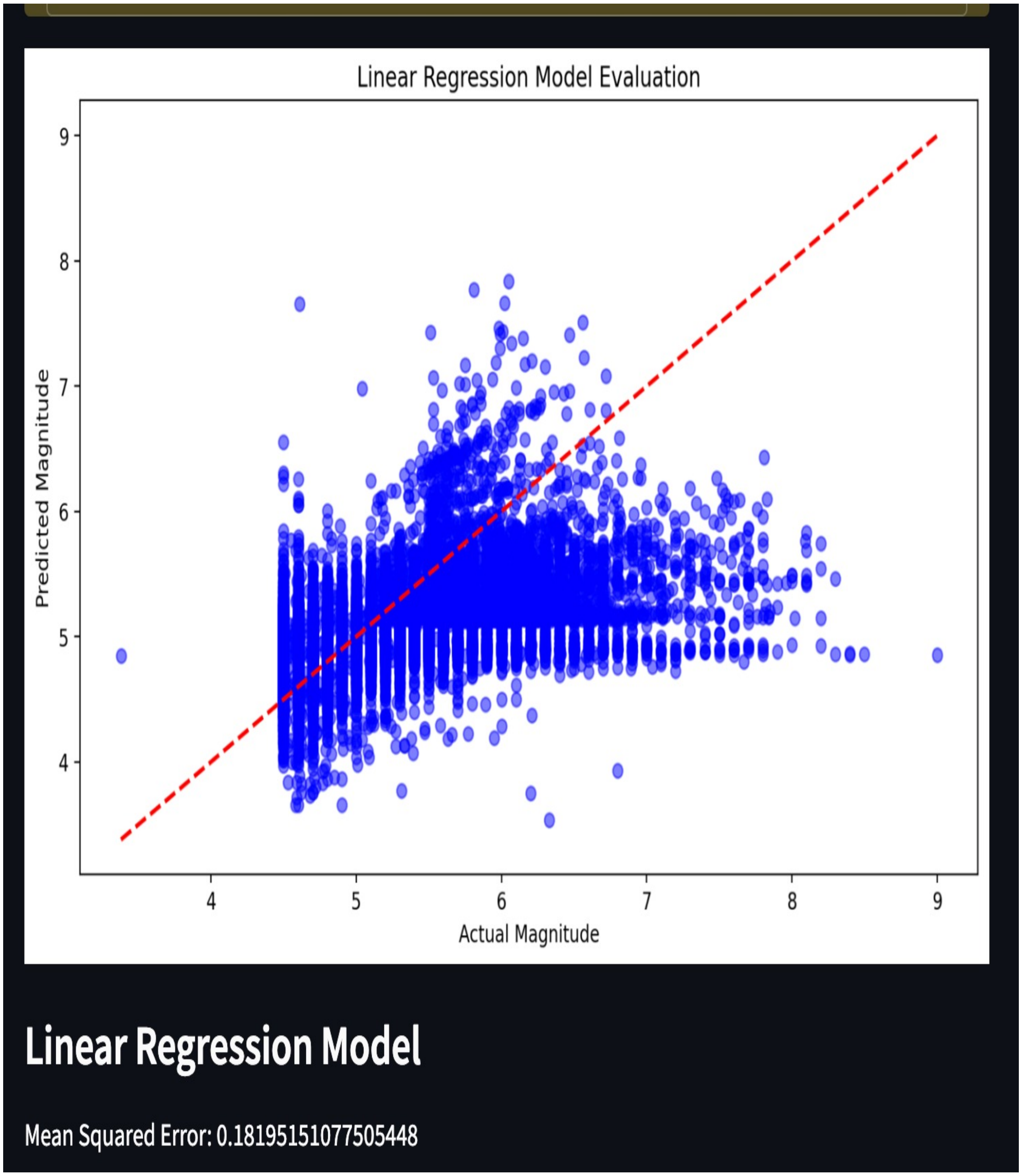
	272373
time	1960-05-22 19:11:20+00:00
latitude	-38.143
longitude	-73.407
mag	9.5

```
(97, 4)
```

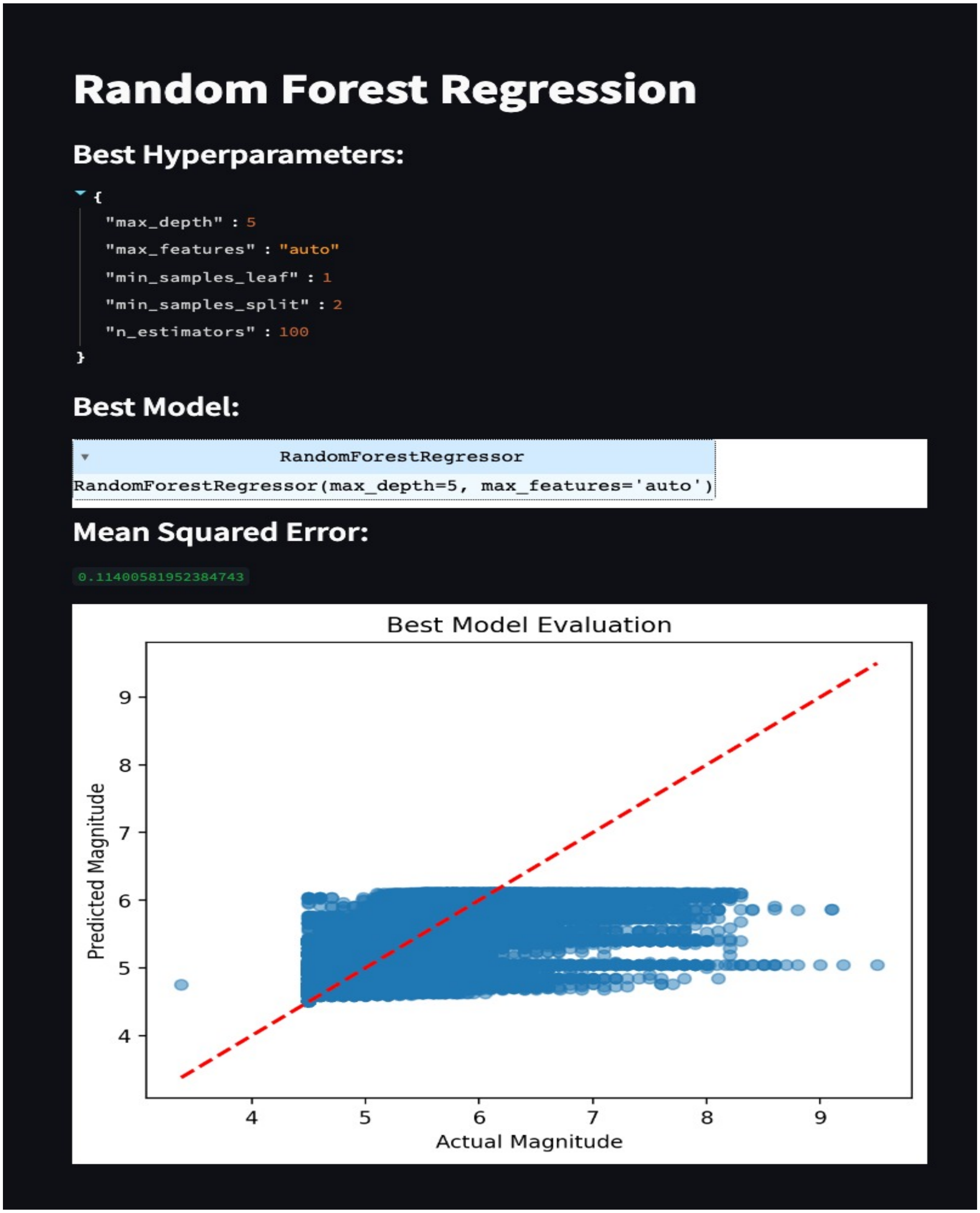
```
282540
```

```
97
```

3) Linear Regression model on the data represented as shown in the below figure.



4) Representation of the Random forest model on the dataset as the best model as shown in the below figure.



5) Testing the model to predict the magnitude using the factors such as latitude, longitude, depth, etc as shown below:

Testing the Model

In [41]:
```
# dict(X.loc[0])
```

In [42]:
```
data = {'const': [1.0],
 'latitude': [41.805],
 'longitude': [79.8675],
 'depth': [10.0],
 'nst': [46.0],
 'gap': [91.0],
 'dmin': [1.293],
 'rms':[0.8],
 'horizontalError': [6.59],
 'depthError': [1.897],
 'magError': [0.078],
 'magNst': [52.0]}
```

In [43]:
```
# Get the best estimator from the grid search
best_model = grid_search.best_estimator_

# Define the custom data
custom_data = pd.DataFrame(data)

# Make predictions on the custom data using the best model
predictions = best_model.predict(custom_data)

# Print the predicted magnitude
print('Predicted Magnitude:', predictions[0])
```

```
Predicted Magnitude: 4.69693598387009
```

Conclusion:

To predict the magnitude, the factors such as latitude, longitude, and depth are mainly used from the dataset. To create the best model, several ML algorithms are performed in which Random Forest made more impact and gave the best output. Moreover, based on the magnitude in the test model, the severity of the earthquake can be predicted in a particular location using latitude and longitude. Hence, these locations can be warned and required action can be taken to avoid the effects of earthquakes.

References:

1. *Earthquake Hazards | U.S. Geological Survey*. (2023, March 7), accessed March 09, 2023, <https://www.usgs.gov/programs/earthquake-hazards>

2. *World Earthquake Data From 1906-2022.* (2023, February 27), Kaggle, accessed March 09, 2023,<https://www.kaggle.com/datasets/garrickhague/world-earthquake-data-from-1906-2022?resource=download>

3. *Stanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI*. (2019). IEEE Journals & Magazine | IEEE Xplore, accessed March 09, 2023,<https://ieeexplore.ieee.org/abstract/document/8871127>

4. *Contreras, D., Wilkinson, S., & James, P. (2021). Earthquake Reconnaissance Data Sources, a Literature Review. Earth, 2(4), 1006–103,* accessed March 09, 2023 .< https://doi.org/10.3390/earth2040060>

5. Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2020). Rapid prediction of the earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophysical Journal International*, *222*(2), 1379–1389. https://doi.org/10.1093/gji/ggaa233

Github Link:

https://github.com/settipallis18/Capstone606