Date	28 September 2023
Team Id	Proj_212172_team 1
Project name	Earthquake prediction model using python
Maximum marks	

Stage 1: Empathize—Research Your Users' Needs

It is well known that if a disaster is occurs in one region it is likely to happen again. Some region has frequent earthquake but this is only a comparative amount compared to other regions.

Problem statement:

This project aims to manage the magnitude and the probability of Earthquake occurring in a particular region using the historic data with various machine learning models to find which model is more accurate to accomplish the task.

Stage 2: Define—State Your Users' Needs and Problems

Requirements:

Define: the second phase of design thinking, where you define the problem statement in a human-centered manner.

An Earthquake prediction define three elements.

- A specified area or location
- A precise period,
- A defined magnitude range
- An exact probability of occurrence

Id	Date	Latitude	Longitude	Depth	Magnitude
iscgemsup907200	18/01/1930	-4.61	153.18	35	6.5
iscgem907212	02/02/1930	51.39	179.82	25	6.4
iscgem907224	14/02/1930	-21.87	-175.10	35	6.4
iscgem907259	06/03/1930	-33.29	-178.01	15	6.3
iscgem907286	26/03/1930	-7.74	125.81	10	7.0

This is the raw data which was sourced from United States Geological Survey(USGS)..

When a powerful earthquake is forecast, disaster control administrators must be alerted to take precautionary measures. In catastrophe preparedness, decisions and activities are centered on preventing losses.

Stage 3: Ideate—Challenge Assumptions and Create Ideas

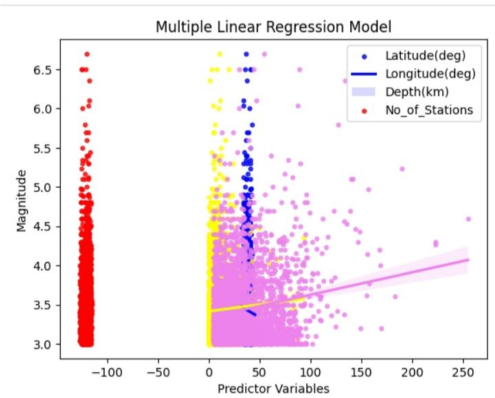
We will use four models in this project:

1) Linear regression
2) Support vector machine
3) NaiveBayes
4) Random forest

Linear Regression

Linear regression is a type of supervised machine learning algorithm that is used to model the linear relationship between a dependent variable (in this case, earthquake magnitude) and one or more independent variables (in this case, latitude, longitude, depth, and the number of seismic stations that recorded the earthquake).

In this situation, we have used multiple linear regression to model the relationship between earthquake magnitude and latitude, longitude, depth, and the number of seismic stations that recorded the earthquake.



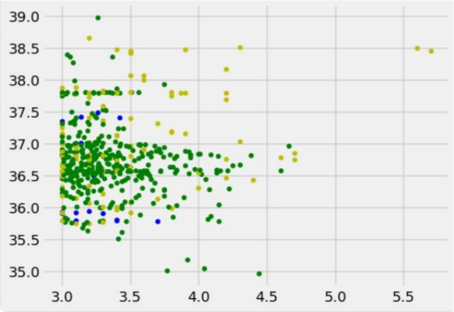
Magnitude = -0.6028 * Latitude + 1.2012 * Longitude - 0.0008 * Depth + 0.0239 * No_of_stations + 0.1573

The results we obtained from the linear regression model were as follows:

- Mean squared error (MSE): 0.17562
- R-squared (R2) score: 0.03498

SVM

Once the SVM model has been trained on the data, it can be used to predict the magnitude of a new earthquake given its features (latitude, longitude, depth, and number of seismic stations). This can be useful for predicting the magnitude of earthquakes in real-time and for better understanding the factors that contribute to earthquake occurrence.

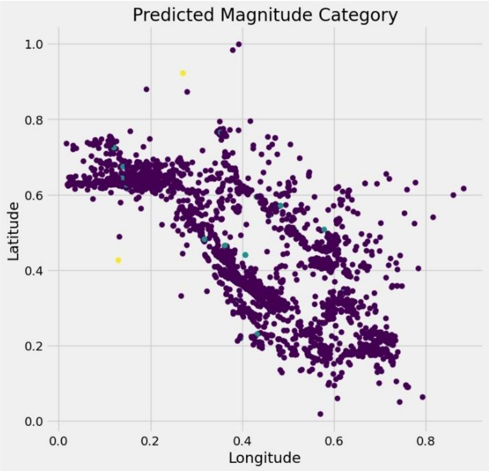


The predicted values from SVM model when evaluated using mse and r2 metrics:

- Mean squared error (MSE): 0.53166
- R-squared (R2) score: -1.92129

Naive Bayes

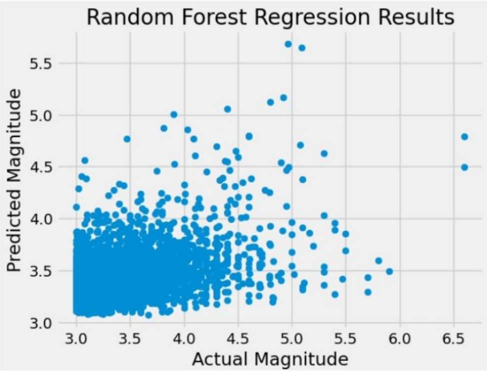
Naive Bayes classifier to predict the magnitude of earthquakes b6ased on their latitude, longitude and number of monitoring stations.



- Accuracy: 0.9853947125161767
- Confusion Matrix: [[5327 35 1] [38 3 1] [4 0 0]]

Random Forest

we used the random forest algorithm to predict the magnitude of earthquakes based on their latitude, longitude, depth, and number of monitoring stations



The results we obtained from the random forest model were as follows:

- Mean squared error (MSE): 0.15599
- R-squared (R2) score: 0.14288

Stage 4: Prototype—Start to Create Solutions

Prototype: the fourth phase of design thinking, where you identify the best possible solution.

The results from the models indicate that the random forest model was able to accurately predict the magnitude of earthquakes based on the given features. The low MSE and high R2 score indicate that the model was making accurate predictions, and was able to explain a large proportion of the variance in the target variable.

Overall, the random forest algorithm is a powerful tool for machine learning tasks, and can be used in a variety of applications, including finance, healthcare, and image recognition

Stage 5: Test—Try Your Solutions Out

Test: the fifth and final phase of the design thinking process, where you test solutions to derive a deep understanding of the product and its users.

When comparing two models, both the mean squared error (MSE) and R-squared (R2) score can be used to evaluate the performance of the models.

In general, a model with a lower MSE and a higher R2 score is considered a better model. This is because the MSE measures the average difference between the predicted and actual values, and a lower MSE indicates that the model is making more accurate predictions. The R2 score measures the proportion of the variance in the target variable that is explained by the model, and a higher R2 score indicates that the model is able to explain more of the variability in the target variable.

From the results of this project we can conclude that random forest is the most accurate model for predicting the magnitude of Earthquake compared to all other models used in this project.

However, it's important to keep in mind that the relative importance of MSE and R2 score may vary depending on the specific problem and the context in which the models are being used. For example, in some cases, minimizing the MSE may be more important than maximizing the R2 score, or vice versa. It's also possible that one model may perform better on one metric and worse on another, so it's important to consider both metrics together when evaluating the performance of the models.