123

by Arshad Khan

Submission date: 18-May-2021 09:14AM (UTC+0530)
Submission ID: 1588507304
File name: MajorProject_BatchD08.pdf (1.95M)
Word count: 2127
Character count: 11034

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION AND BACKGROUND

There are various reasons linked to the eruption of a volcano. When two tectonic plates collide, the denser plate pushes underneath the lighter plate. The molten rock, seawater and sediments also move along with it and get melted. Over time they fill the magman chamber. When the pressure in the chamber increases, a volcanic eruption occurs. The type, size of eruption depends on the formation of the volcano and various other factors. While reading a research paper titled "Understanding the timing of eruption end using a machine learning approach to classification of seismic time series" (published in the Journal of Volcanology and Geothermal Research), we realized that earthquakes and volcanic eruptions could be linked. Only an earthquake that has a magnitude greater than six can trigger an eruption. Papers published in (2002) by Warner Marzocchi, Michael Manga1 and Emily Brodsky (2006), Sebastian Watt and David M Pyle (2009), M. S. Bebbington W. Marzocchi (2007), Ken'Ichiro Yamashina and Kazuaki Nakamura (1978) also support the linking of earthquakes and volcanic eruptions.

Volcanic eruptions are magnificent and sometimes the deadliest natural events on Earth. Predicting a volcanic eruption is a challenging task. Various factors trigger a volcanic eruption. It is also believed that sometimes tectonic earthquakes can also cause volcanic eruptions. It is important to remember that only earthquakes above 6 can trigger an eruption. The first writer to link earthquakes and volcanic eruptions was a very famous English naturalist, geologist and also a biologist: Charles Darwin. In 1835 an earthquake having a magnitude of 8.8 occurred in Chile. Charles Darwin was also a witness to this earthquake. In the following weeks, he started investigating the reasons for the earthquake and its effects. Using his observations and the observations of the locals, he discovered that three volcanoes erupted along the Chilean coast at the same time as the earthquake.

At present, approximately 800 million people live around active volcanos. Understanding these earthquake-volcano interactions can help us create hazard management systems that can save these people's lives. There have been various events in history when earthquakes triggered a volcanic eruption. Eg. The eruption of Mount Pinatubo (volcano in the Zambales Mountains, Philippines) in 1991 is considered the most significant and destructive eruption of the 20th century. Mount Pinatubo erupted approximately one year after a 7.8 Magnitude earthquake (16 July 1990) hit the Philippines. Another event like this also took place in Japan in 1707. An 8.7 Magnitude earthquake was followed by a volcanic eruption approximately 47 days later.

Richter scale of earthquake magnitude

magnitude level	category	effects	earthquakes per year
less than 1.0 to 2.9	micro	generally not felt by people, though recorded on local instruments	more than 100,000
3.0–3.9	minor	felt by many people; no damage	12,000–100,000
4.0–4.9	light	felt by all; minor breakage of objects	2,000–12,000
5.0–5.9	moderate	some damage to weak structures	200–2,000
6.0–6.9	strong	moderate damage in populated areas	20–200
7.0–7.9	major	serious damage over large areas; loss of life	3–20
8.0 and higher	great	severe destruction and loss of life over large areas	fewer than 3

Fig1.1: Only earthquakes above six can cause a volcanic eruption

Our primary focus is on Japan because it is the only country that receives maximum earthquakes every year. It is also the fourth country with the maximum number of volcanoes. Japan is located along the pacific ring of fire, the most active earthquake belts globally. Most of the worlds earthquakes and volcanic eruptions occur here. Moreover, Japan sits on the boundary of four tectonic plates: the Pacific plate, the North American Plate, the Eurasian Plate, and the Filipino plate. These reasons make Japan an earthquake-prone zone.

1.1.1 FORMATION OF JAPANESE ISLAND ARC:

- The concept of Ocean-Ocean Convergence helps us understand the formation of the Japanese Island Arc. In Ocean-Ocean Convergence, a denser oceanic plate subducts below a less dense oceanic plate forming a trench along the boundary.

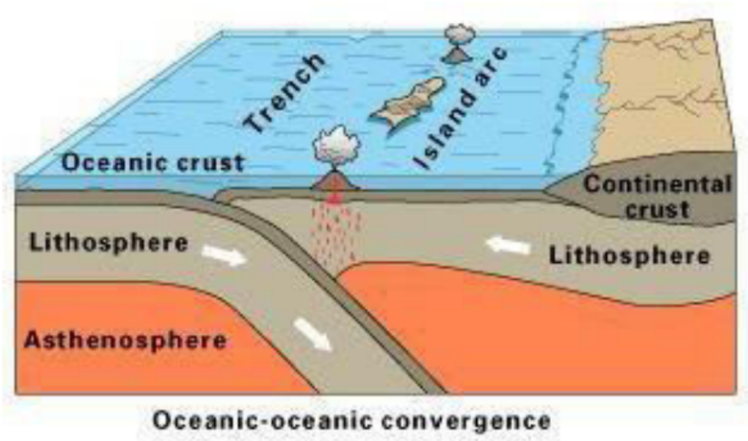


Fig.1.2 Ocean – Ocean Convergence (via Wikimedia Commons)

Under high pressure and temperature, the rocks in the subduction zone metamorphose (change in the composition or structure of a rock) as the ocean floor crust (oceanic plate) subducts into the asthenosphere. Magma (metamorphosed sediments and melted subducting plate) has a lower density and is under high pressure. It rises as a result of the buoyant force provided by the surrounding denser medium.The magma then rises to the surface and constant volcanic eruptions on the ocean floor are caused by the continuous upward movement of magma.

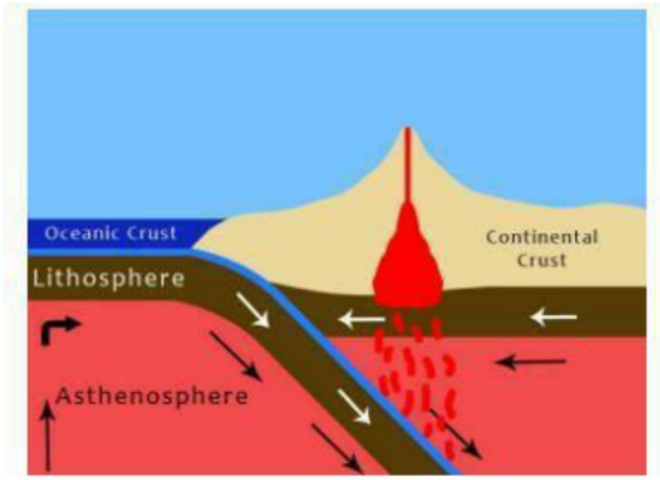


Fig 1.3. Subduction Zone Illustration (Eround1, via Wikimedia Commons)

This process continues over millions of years and a volcanic landform, which in some cases rises above the ocean, is formed. Along the boundary, such volcanic landforms form a chain of volcanic islands known as Island Arcs (Indonesian Island Arc or Indonesian Archipelago, Philippine Island Arc, Japanese Island Arc, and so on).Orogenesis (mountain formation) is the process by which continental crust is formed by replacing oceanic crust (this occurs much later). New islands, for example, appear around Japan every few years.

CHAPTER 2: LITERATURE SURVEY

Warner Marzocchi (2002): An italian volcanologist Warner Marzocchi published research paper in 2002 "Remote seismic influence on large explosive eruptions", **in** Journal Of Geophysical Research. According to him: the physical process governing the most powerful explosive eruptions has a very high degree of freedom; identifying any non-random pattern can significantly improve knowledge of the process The main aim of their paper was to find out if the tetonoc earthquakes can chance or impact the occurrence of a volcanic event. They found out that there is a significant relation between earthquakes and the greatest volcanic eruptions that occurred in the previous centuries.

Michael Manga1 and Emily Brodsky (2006): In 2006 Michael Manga and Emily E. Brodsky published a paper "Seismic Triggering of Eruptions in the Far Field: Volcanoes and Geysers". according to them Their paper provided more quantitative analysis. They found out that the occurrence of an eruption after an earthquake was 0.3 to 0.5 percent.

M. S. Bebbington W. Marzocchi (2007): In 2007 Begington (Professor in Geostatistics) and Marozocchi published a paper "Stochastic models for earthquake triggering of volcanic eruptions and Geysers." Acoording to their finding earthquakes caused by the inflation of the volcanic surface in a particular region can inturn trigger a volcanic eruption.

Sebastian Watt and David M Pyle (2009): Sebastian Watt (volcanologist and senior lecturer in Earth Sciences at the University of Birmingham, UK) and David Pyle(a volcanologist at the University of Oxford) published a paper "The influence of great earthquakes on volcanic eruption rate along the Chilean subduction zone. They found out that earthquakes which were greater than 8 on the richter scale could successfully trigger a volcanic eruption . They had analysed all the data between 1906 and 1960.

Ken'Ichiro Yamashina and Kazuaki Nakamura ,1978: In 1978 Yamashina and Nakamura published a research paper "Correlation between tectonic earthquakes and volcanic activity of izu-oshima volcano, japan". They studied the coorelation between the earthquakes and Izu-Oshima volcano from (1921-1975) .According to their analysis starins caused by earthquakes have a verey stron relation to the volcanic eruptions.

CHAPTER 3: OBJECTIVE OF THE PROJECT

3.1 OBJECTIVE:

Our project aims to predict the possibility of a volcanic eruption if an earthquake occurs. We want to link earthquakes and volcanic eruptions.

3.2 FEATURES

We are focusing on the following features:

1. Latitude
2. Longitude
3. Magnitude

Among all the features available we chose these three features because of the following reasons:

1. Latitude and longitude can tell us the location of the epicentre or the nearest location where an earthquake has originated at a particular place in the map.

1	Date	Time	Latitude	Longitude
2	01-02-1965	13:44:18	19.246	145.616
3	01-04-1965	11:29:49	1.863	127.352
4	01-05-1965	18:05:58	-20.579	-173.972
5	01-08-1965	18:49:43	-59.076	-23.557
6	01-09-1965	13:32:50	11.938	126.427
7	01-10-1965	13:36:32	-13.405	166.629
8	01-12-1965	13:32:25	27.357	87.867
9	01/15/1965	23:17:42	-13.309	166.212
10	01/16/1965	11:32:37	-56.452	-27.043
11	01/17/1965	10:43:17	-24.563	178.487
12	01/17/1965	20:57:41	-6.807	108.988
13	01/24/1965	00:11:17	-2.608	125.952
14	01/29/1965	09:35:30	54.636	161.703
15	02-01-1965	05:27:06	-18.697	-177.864

Fig.3.1dataset1

In the above figure the latitudes and longitudes are mentioned in decimal degrees. The –ve latitude represents south of the equator and –ve longitude represents the west of the prime meridian(underlined in red) . Positive latitudes represent north of the equator and positive longitude represents west of the prime meridian(underlined in yellow).

2. Magnitude represents the intensity of the earthquake. Only earthquakes which have a value greater than 6 on the richter scale have to potential to trigger an eruption.

Fig.3.2 Epicentre

In the Fig.3.2 the latitude -56.452 and longitude -27.04 pont the epicentre of the earthquake. which is in the South Atlantic Ocean.

CH 4: SYSTEM METHODOLOGY

4.1 TECHNICAL OR SOFTWARE REQUIREMENTS

For the implementation of the project one needs to have the following software requirements:

- **Python idel 3.7 version (or)**
- **Anaconda 3.7 (or)**
- **Jupiter (or)**
- **Google colab**

4.2 HARDWARE REQUIREMENTS

The hardware requirements entire dependent of the type of software which is being developed.A system which has the following requirements is enough for the project execution.

- **Operating system: Windows, Linux**
- **Processor : minimum intel i3**
- **Ram : minimum 4 gb**
- **Hard disk : minimum 250gb**

4.1 UML DIAGRAMS

4.3.1 Use Case Diagram

Use case diagrams are behaviour diagrams that are used to describe the use cases. The set of actions that the subject should or can perform in collaboration with one or more actors who are present external users. It represents the user's interaction in the use case diagram by visualizing the relationship between the user and various use cases in which the user is involved.

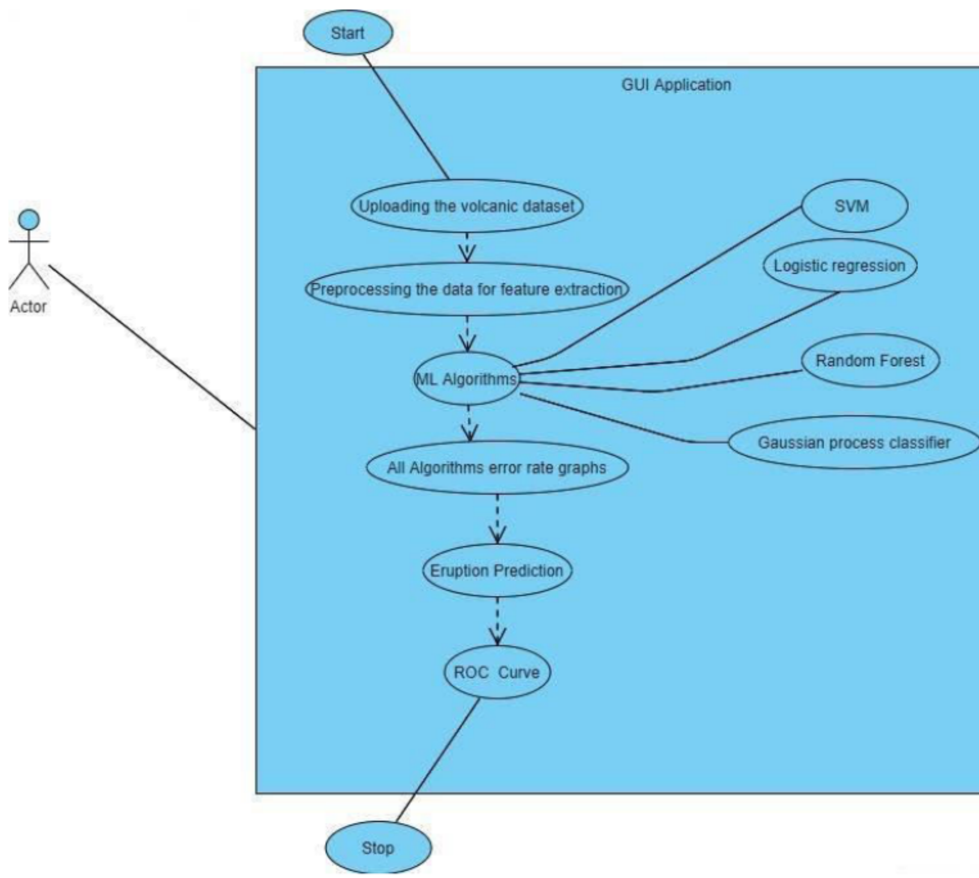


Fig.4.1 Use case diagram

4.3.1 Sequence Diagram

A sequence diagram is an interaction diagram that shows how the objects in the diagram interact with one another and in what order (stepwise). Because it is made up of messages, it is also known as a message sequence chart. It will display object interactions between the objects that are present in a time sequence for the flow of functionality.

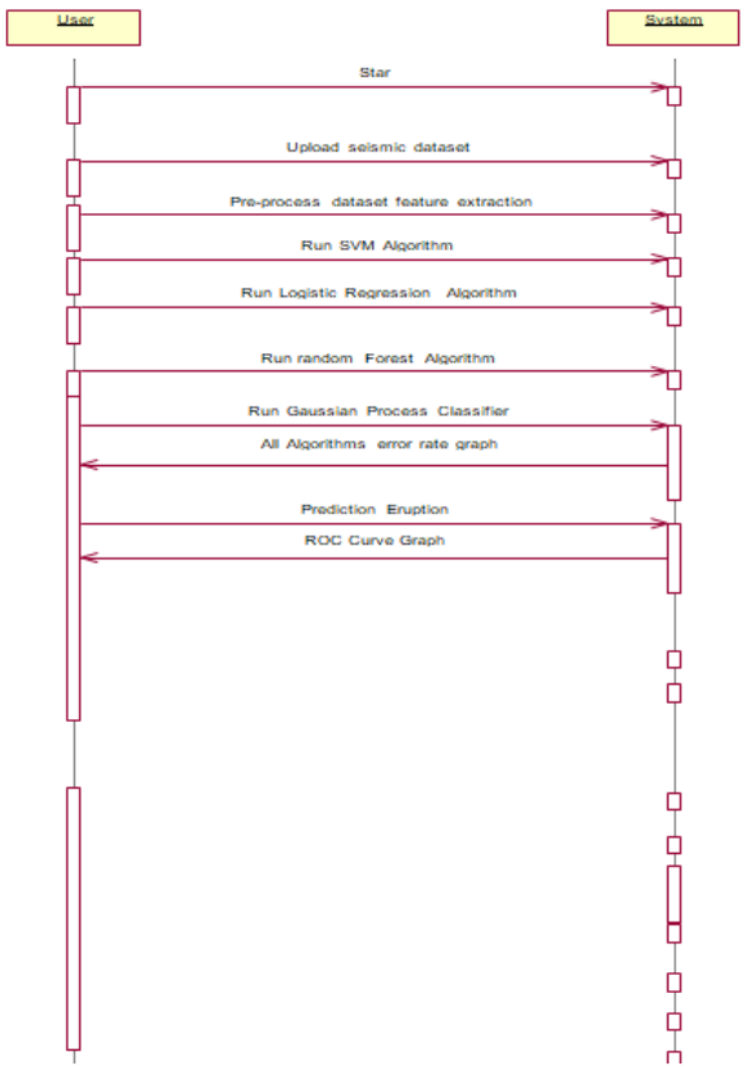


Fig. 4.2 Sequence diagram

4.1 ALGORITHMS

4.4.1. SVM ALGORITHM

The Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification or regression tasks. It is, however, primarily used in classification problems. In the SVM algorithm, each data item is plotted as a point in n-dimensional space (where n is the number of features). The value of each feature is the value of a specific coordinate. Then, we perform classification by locating the hyperplane that best distinguishes the two classes.SVMs, which are based on statistical learning frameworks or the VC theory proposed by Vapnik (1982, 1995) and Chervonenkis, are among the most robust prediction methods (1974). Given a set of training examples, each labelled as belonging to one of two categories; an SVM training algorithm constructs a model , that assigns new examples to one of the two categories, resulting in a non-probabilistic binary linear classifier. (Although methods such as Platt scaling exist to use SVMin a probabilistic classification setting).

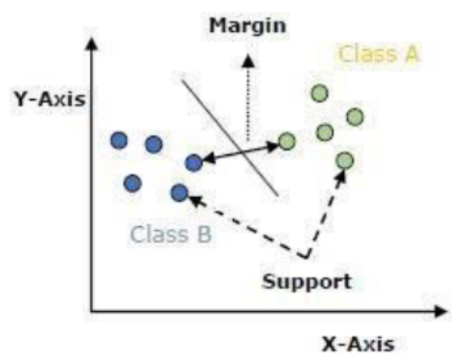


Fig.4.3 SVM

```
def runSVM():

global  svm_acc

global cls1

text.delete('1.0', END)

cls = svm.SVC(C=1.5,gamma='scale')

cls.fit(X, Y)

prediction_data = cls.predict(X_test)

svm_acc = accuracy_score(y_test,prediction_data)*100

text.insert(END,"SVM Accuracy on Eruption Dataset : "+str(svm_acc)+"\n")

cls1 = cls
```

The following are key concepts in SVM:

Datapoints: that are closest to the hyperplane are referred to as support vectors. These data points will be used to define a separating line.

Hyperplane: As shown in the diagram above (Fig.5.3), a hyperplane is a decision plane or space that divides a set of objects of different classes.

Margin: It is s defined as the difference between two lines on the closest data points of different classes. A large margin is considered a good margin, and a small margin is considered an insufficient margin.

Applications of SVM:

- Face detection entails classifying parts of an image as face or non-face and drawing a square boundary around the face.
- Text and hypertext categorization — SVMs support text and hypertext categorization in both inductive and transductive models. They classify documents into different categories using training data. It categorizes based on the generated score and then compares it to the threshold value.

- Image classification — The use of SVMs improves search accuracy for image classification. It outperforms traditional query-based searching techniques in terms of accuracy.

4.4.2.LOGISTIC REGRESSION

This Algorithm helps in determining the probability of an intended variable. This ML algorithm shows the probability as either 1 which represents success and 0 which represents failure. It has a variety of applications eg. Cancer detection, span detection etc.

```
def runLR():
    global lr_acc
    global cls2
    cls = LogisticRegression()
    cls.fit(X, Y)
    prediction_data = cls.predict(X_test)
    lr_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"Logistic Regression Accuracy on Eruption Dataset : "+str(lr_acc)+"\n")
    cls2 = cls
```

Logistic regression is classified into the following types:

Multinimoal:

The term "multinomial" refers to the fact that there. The dependent variable in such a classification can have three or more possible unordered types or types with no quantitative significance. These variables could, for example, represent "Type A," "Type B," or "Type C."

Ordinary

In this type of classification, the dependent variable can have three or more possible ordered types or types with quantitative significance. For example, these variables could represent "poor" or "good," "very good," or "Excellent," and each category could have a score of 0, 1, 2, or 3.

Binomial or binary:

In this type of classification, a dependent variable will only have two possible values: 1 or 0. These variables could, for example, represent success or failure, yes or no, win or loss, and so on.

4.4.3.RANDOM FOREST ALGORITHM

This ML algorithm is applicable for bothe regression and classification related problems. Here multiple classifiers are used to solve a problem. It has seven decision trees and instead of depending on a single tree it takes the predictions from every single tree, it averages to improve the accuracy of the dataset..

merges them to produce a more accurate and stable prediction. Random forest has a significant advantage in that it can be used for both classification and regression problems, which comprise most current machine learning systems. Let us look at the random forest in classification because classification is sometimes thought to be the foundation of machine learning.

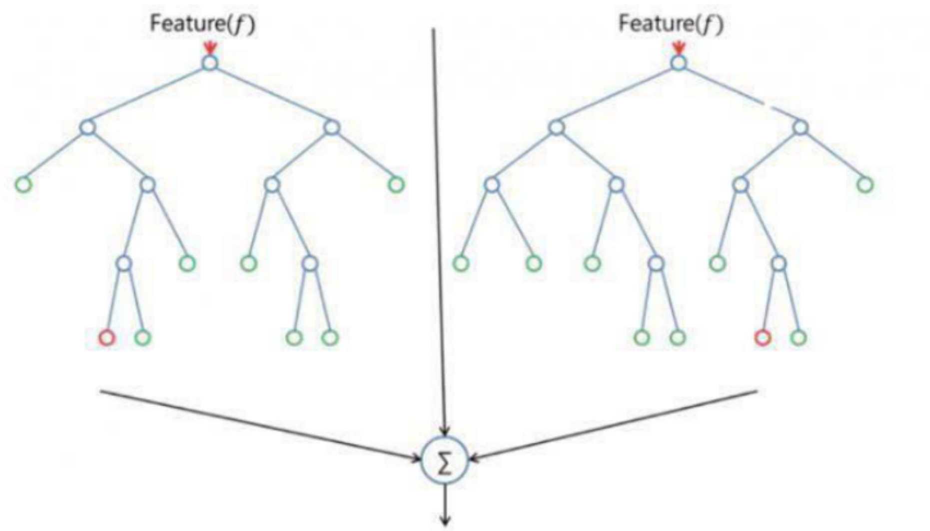


Fig:4.4 Random forest with two trees

```
def runRandomForest():
    global model
    global cls3
    global rf_acc
    cls = RandomForestClassifier(n_estimators=20, random_state=0)
    cls.fit(X, Y)
    prediction_data = cls.predict(X_test)
    rf_acc = accuracy_score(y_test,prediction_data)*100
```

```
text.insert(END,"Random Forest Accuracy on Eruption Dataset : "+str(rf_acc)+"\n")

model = cls

cls3 = cls
```

Random Forest Assumptions

- Because the random forest combines multiple trees to predict the dataset's class, some decision trees may predict the correct output while others may not. However, when all of the trees are combined, they predict the correct outcome. As a result, the following are two assumptions for a better Random forest classifier:
- There should be some actual values in the dataset's feature variable so that the classifier can predict accurate results rather than guesses.
- Each tree's predictions must have very low correlations.

Random Forest Assumptions

- Because the random forest combines multiple trees to predict the dataset's class, some decision trees may predict the correct output while others may not. However, when all of the trees are combined, they predict the correct outcome. As a result, the following are two assumptions for a better Random forest classifier:
- There should be some actual values in the dataset's feature variable so that the classifier can predict accurate results rather than guesses.
- Each tree's predictions must have very low correlations.

4.4.4. GAUSSIAN PROCESS CLASSIFIER

This algorithms main practical advantage is that it can provide a reliable estimate of its own uncertainty. By the end of this high-level, math-free post, I hope to have given you an intuitive understanding of what a Gaussian process is and what distinguishes it from other algorithms. It uses a lazy learning approach.

Applications of Gaussian Process Classifier:

- Environmental science
- Hydrogeology
- Real estate valuation
- Analysis and Optimization of Integrated Circuits

```
def runGaussian():
    global cls4
    global gaussian_acc
    cls = GaussianProcessClassifier()
    cls.fit(X_test, y_test)
    prediction_data = cls.predict(X_test)
    gaussian_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"Gaussian Process Classifier Accuracy on Eruption Dataset : "+str(gaussian_acc)+"\n")
    cls4 = cls
```

CHAPTER 5 : IMPLEMENTATION

5.1 PYTHON PACKAGES

- **NumPy :** NumPy includes tools for creating multidimensional arrays and performing calculations on the data contained within them. You can perform common statistical operations, solve algebraic formulas, and much more.
- **Sklearn:** Sklearn is a Python machine learning library that includes algorithms such as logistic regression, decision trees, support vector machines, random forests, and many others. These are the Python packages we use, which can be used to build by using numpy, matplotlib, and scipy. We use these critical packages to ensure that the project runs smoothly and efficiently. These packages are ineffective for reading data, manipulating data, or summarising data.
- **pandas:** It has some functions for analyzing, cleaning, exploring, and manipulating data. Pandas enables us to draw conclusions and analyze large amounts of data using statistical theories. We can use pandas to clean up messy data sets and make them relevant and readable.
- **Tkinter:** Tkinter is the most important and widely used framework for creating graphical user interfaces. It connects Python to the TK GUI toolkit, which runs on almost every modern operating system.

5.1 DATASETS

This dataset contains the seismic magnitude of many volcanoes. If its value is greater than 6, then it will be considered that a volcano is about to erupt. We had used 80% of dataset records to train Machine Learning algorithms and 20% of the records to calculate its classification accuracy. We have used two datasets: The first dataset consists of earthquakes across the globe from 1965-2016. This dataset consists of 23413 rows. The second dataset is of earthquakes in Japan from 2001-2018. This dataset has 14093 rows. Based on the features(latitude, longitude and magnitude), these are stored in a .csv file.

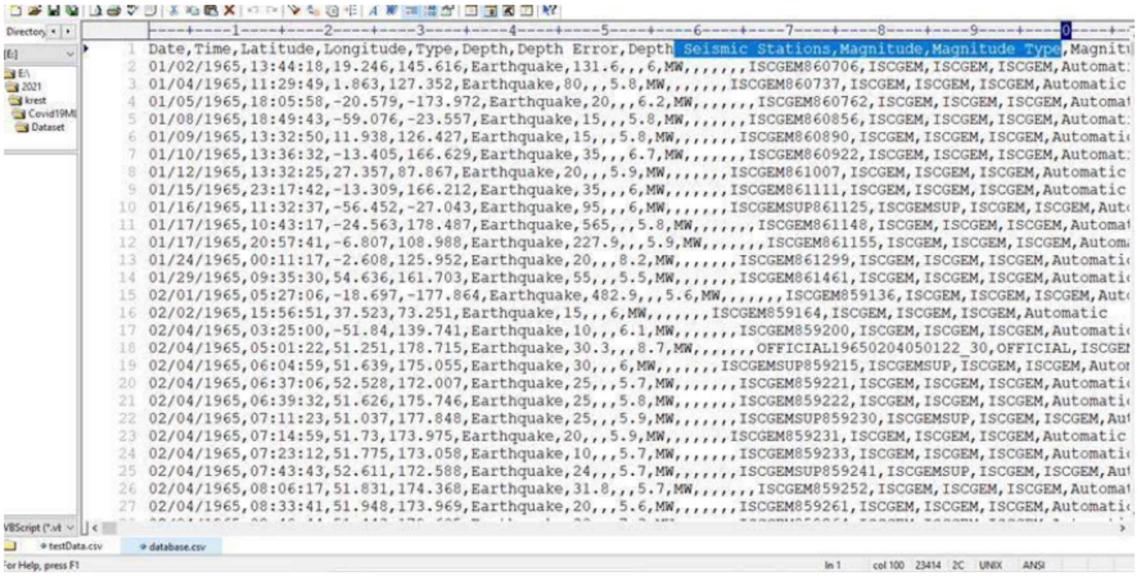


Fig.5.1 Dataset 1 earthquakes across the globe from (1965-2016)

time	latitude	longitude	depth	mag	magType
2018-11-2	48.378	154.962	35	4.9	mb
2018-11-2	36.0733	139.783	48.82	4.8	mww
2018-11-2	38.8576	141.8384	50.56	4.5	mb
2018-11-2	50.0727	156.142	66.34	4.6	mb
2018-11-2	33.95	134.4942	38.19	4.6	mb
2018-11-2	48.4158	155.0325	35	4.6	mb
2018-11-2	37.1821	141.1721	46.76	5.2	mb
2018-11-2	29.3424	142.3121	10	4.7	mb
2018-11-2	44.4524	148.0753	101.46	4.7	mww
2018-11-2	30.4087	130.0687	123	5.5	mww
2018-11-2	42.0009	142.7654	73.55	4.5	mb
2018-11-1	24.1937	125.2046	30.25	4.5	mwr
2018-11-1	30.7822	141.9762	10	4.8	mb
2018-11-1	25.3931	141.0321	115.25	4.8	mb
2018-11-1	26.3407	143.8582	36.24	4.8	mb
2018-11-1	39.6014	141.937	49.23	4.5	mb
2018-11-1	42.6765	141.9846	35	4.8	mb
2018-11-1	34.7663	139.9805	105.5	5	mb
2018-11-1	45.2679	150.481	54.13	4.8	mb
2018-11-1	42.9554	139.2679	8.66	4.5	mb
2018-11-1	31.4232	141.8762	10	4.9	mb
2018-11-1	49.5372	155.8279	42.08	4.6	mb
2018-11-0	31.4465	141.757	10	4.7	mb
2018-11-0	31.4081	141.6164	7.65	4.9	mb
2018-11-0	44.7104	145.6708	22.35	4.7	mb
2018-11-0	40.7338	142.5527	46.17	4.5	mb

Fig.5.2 Dataset 2 Earthquakes in Japan

The GUI looks like this: The user can click on the "Upload seismic data" button to upload the data set.

Fig.5.3 GUI

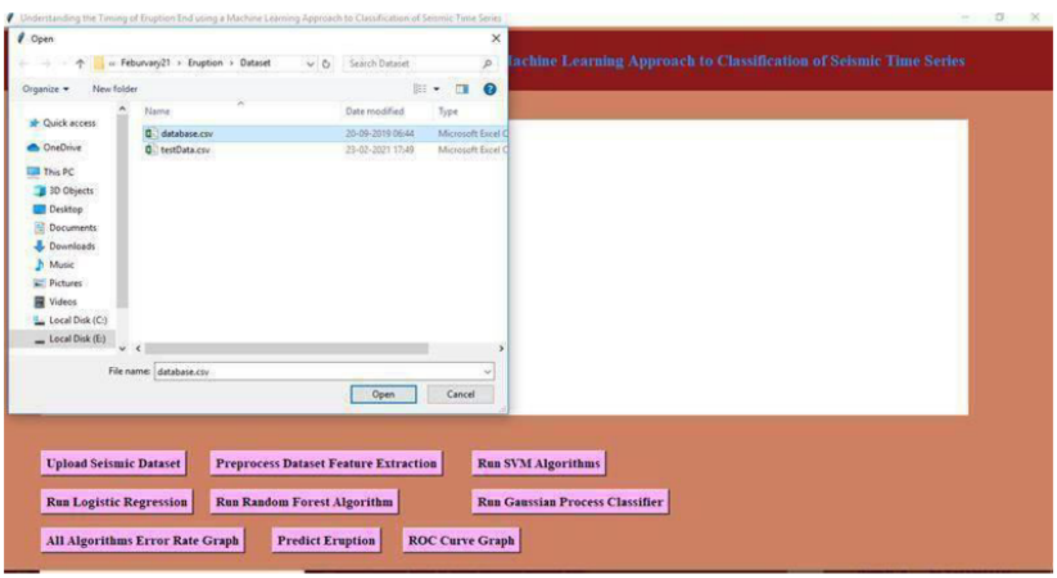


Fig:5.4 Upload seismic data

We uploaded the 'database.csv' file and then clicked on the 'Open' button to load the dataset.

In the screen below(Fig:5.5), the dataset is loaded and is displaying certain records. We can see string values, and we need to replace the string values with numeric values and then replace missing values with 0. Next we click on the "Preprocess Dataset Feature Extraction" button to convert the dataset into normalized format.

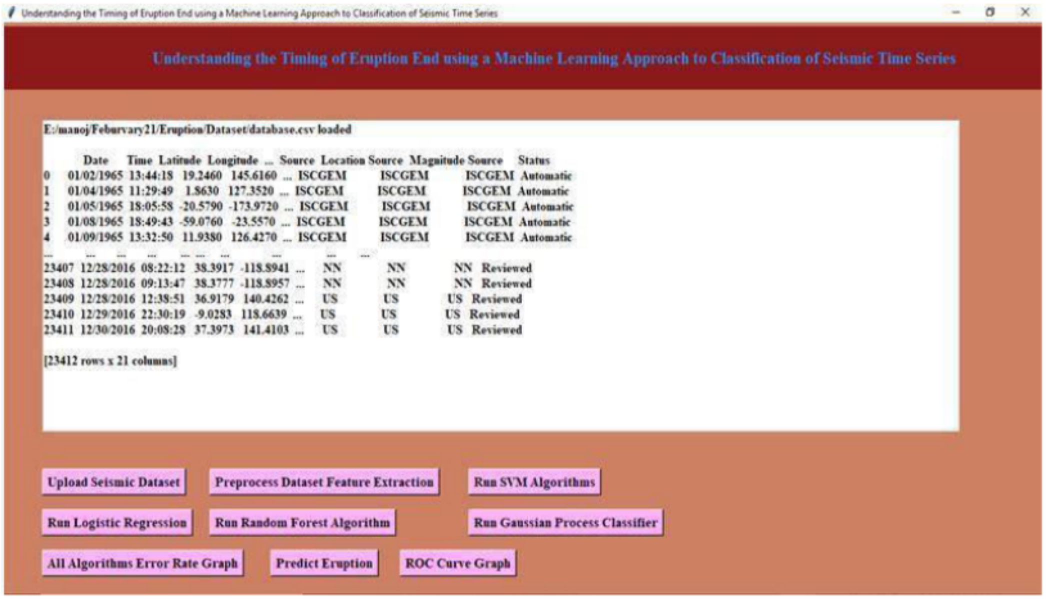


Fig:5.5 Dataset Loaded

In the screen below(Fig:5.6), all records were converted to numeric values. We can see that the application contains a total of 23412 records, and the application uses 18729 records to train machine learning algorithms and 4683 records to test them. Since both train and test data are ready, now we run the four ML classification methods:We click on the "Run SVM Algorithm' button to train the SVM model with the previous dataset.

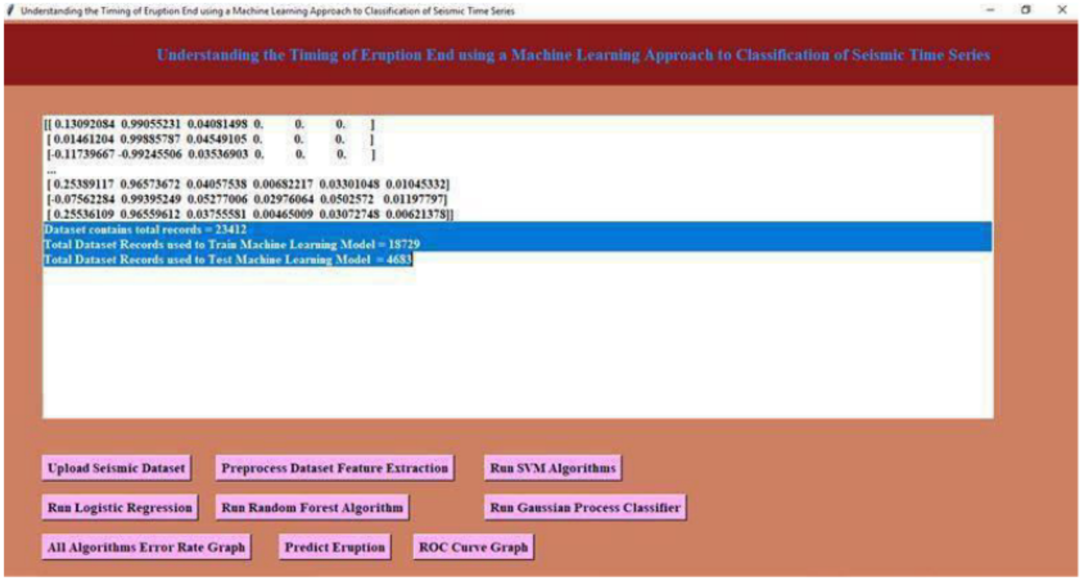


Fig:5.6 Records converted to numeric values

In the below screen(Fig.5.7), we trained the SVM model, and its accuracy is 54%. Next, we click on the 'Run Logistic Regression' button to get its accuracy.

Fig:5.7 Trainig the SVM model

In the below screen(Fig.5.8), we can see that logistic regression got an accuracy of 55%. We move on to 'Run Random Forest Algorithm' button to get its accuracy.

Fig:5.8 Running Logistic regression Algorithm

In the below screen(Fig.5.9), we can see that the random forest algorithm got 99.74% classification accuracy. We now click on the 'Run Gaussian Process Classifier' button to get its accuracy.

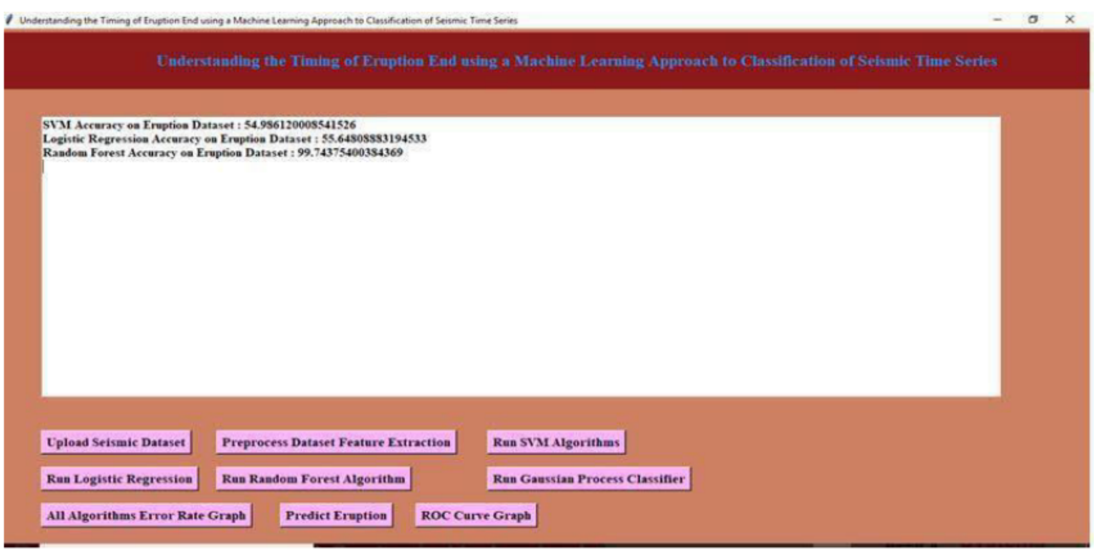


Fig.5.9 Running random forest algorithm

In the below screen(Fig. 6.0), we can see that the Gaussian process Classifier has an accuracy rate of 55%.

Fig:6.0 Gaussian process classifier

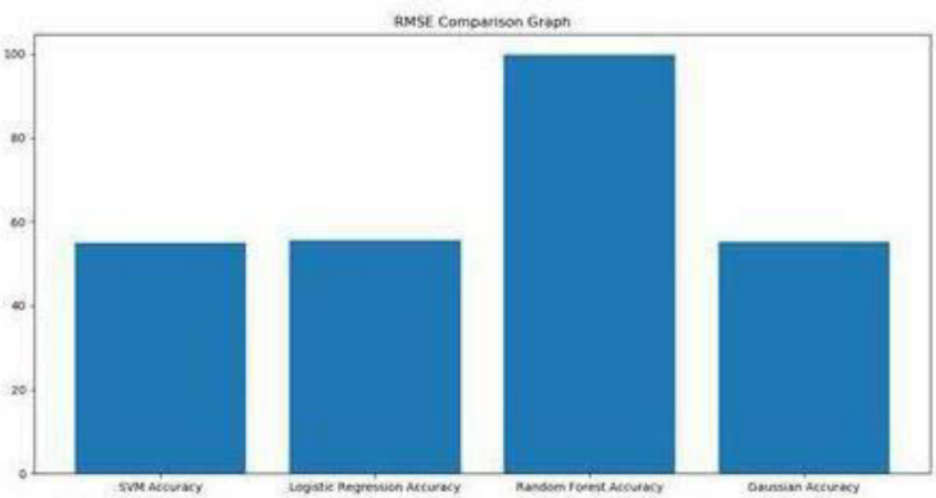


Fig. 6.1 All algorithms error rate graph

In the below screen(Fig:6.2), we upload the 'testData.csv' file. We then get the below result. We are stating whether an eruption will occur or not.

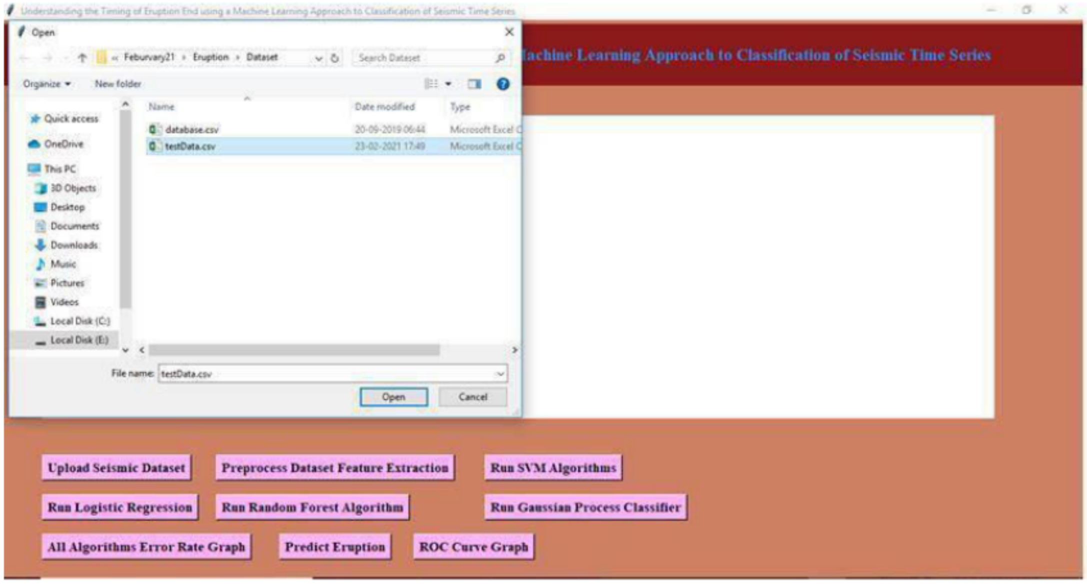


Fig:6.2 Test data

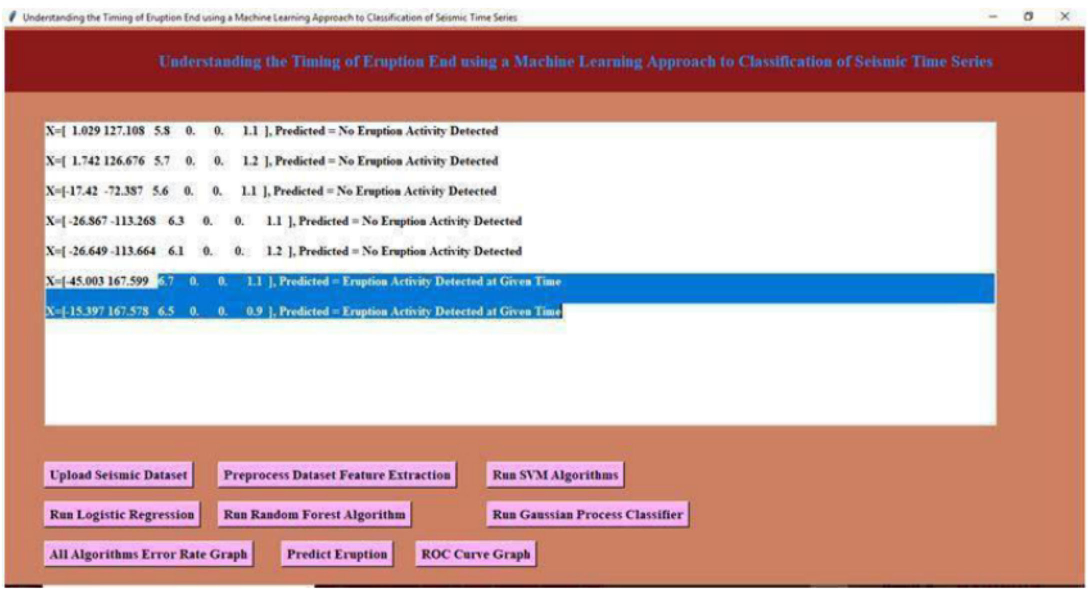


Fig: 6.3 Prediction

Here (Fig.6.3), in square brackets, we can see the test data. After the square bracket, we can see predicted results => as 'No eruption detected' or 'eruption detected'. In the above screen, we can see, whenever the classifier sees a magnitude value >=6.5, it classes that record time as 'eruption activity detected'.

CODE :

eruption.py

```
from tkinter import messagebox
from tkinter import *
from tkinter import simpledialog
import tkinter
from tkinter import filedialog
from tkinter.filedialog import askopenfilename
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import normalize
from sklearn.metrics import accuracy_score
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.gaussian_process import GaussianProcessClassifier
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score
main = tkinter.Tk()
main.title("Understanding the Timing of Eruption End using a Machine Learning Approacl
to Classification of Seismic Time Series") #designing main screen
main.geometry("1300x1200")

global filename
global svm_acc,lr_acc,rf_acc,gaussian_acc
global X, Y
```

```
global X_train, X_test, y_train, y_test
global dataset
global model
global cls1,cls2,cls3,cls4

def upload(): #function to upload tweeter profile
    global filename
    global dataset
    filename = filedialog.askopenfilename(initialdir="Dataset")
    text.delete('1.0', END)
    text.insert(END,filename+" loaded\n\n");
    dataset = pd.read_csv(filename)
    text.insert(END,str(dataset))

def preprocess():
    global X, Y
    global X_train, X_test, y_train, y_test
    global dataset
    text.delete('1.0', END)
    dataset.fillna(0, inplace = True)
    dataset = dataset[['Latitude','Longitude','Magnitude','Horizontal Distance','Horizontal
Error','Root Mean Square']]
    X = dataset.values
    Y = []
for i in range(len(X)):
    m = X[i,2]
    if m < 5.8:
        Y.append(0)
    else:
        Y.append(1)

Y = np.asarray(Y)
X = normalize(X)
text.insert(END,str(X)+"\n")
indices = np.arange(X.shape[0])
np.random.shuffle(indices)
```

```
    X = X[indices]
    Y = Y[indices]
    X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2)
    text.insert(END,"Dataset contains total records = "+str(len(X))+"\n")
    text.insert(END,"Total Dataset Records used to Train Machine Learning Model =
"+str(X_train.shape[0])+"\n")
    text.insert(END,"Total Dataset Records used to Test Machine Learning Model =
"+str(X_test.shape[0])+"\n")

def runSVM():
    global svm_acc
    global cls1
    text.delete('1.0', END)
    cls = svm.SVC(C=1.5,gamma='scale')
    cls.fit(X, Y)
    prediction_data = cls.predict(X_test)
    svm_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"SVM Accuracy on Eruption Dataset : "+str(svm_acc)+"\n")
    cls1 = cls
```

```
def runLR():
    global lr_acc
    global cls2
    cls = LogisticRegression()
    cls.fit(X, Y)
    prediction_data = cls.predict(X_test)
    lr_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"Logistic Regression Accuracy on Eruption Dataset : "+str(lr_acc)+"\n")
    cls2 = cls

def runRandomForest():
    global model
    global cls3
global rf_acc
    cls = RandomForestClassifier(n_estimators=20, random_state=0)
    cls.fit(X, Y)
    prediction_data = cls.predict(X_test)
    rf_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"Random Forest Accuracy on Eruption Dataset : "+str(rf_acc)+"\n")
    model = cls
    cls3 = cls

def runGaussian():
    global cls4
    global gaussian_acc
    cls = GaussianProcessClassifier()
    cls.fit(X_test, y_test)
    prediction_data = cls.predict(X_test)
    gaussian_acc = accuracy_score(y_test,prediction_data)*100
    text.insert(END,"Gaussian Process Classifier Accuracy on Eruption Dataset :
"+str(gaussian_acc)+"\n")
    cls4 = cls
```

```
def graph():
  height = [svm_acc,lr_acc,rf_acc,gaussian_acc]
  bars = ('SVM Accuracy','Logistic Regression Accuracy','Random Forest
Accuracy','Gaussian Accuracy')
  y_pos = np.arange(len(bars))
  plt.bar(y_pos, height)
  plt.xticks(y_pos, bars)
  plt.title('Accuracy Comparison Graph')
  plt.show()

def predict():
  text.delete('1.0', END)
  name = filedialog.askopenfilename(initialdir = "Dataset")
  test = pd.read_csv(name)
  test.fillna(0, inplace = True)
  test = test[['Latitude','Longitude','Magnitude','Horizontal Distance','Horizontal Error','Root
Mean Square']]
  test = test.values
  print(test.shape)
  y_pred = model.predict(test)
  print(y_pred)

for i in range(len(test)):
  if str(y_pred[i]) == '0':
    text.insert(END,"X=%s, Predicted = %s" % (test[i], 'No Eruption Activity
Detected')+"\n\n")
    else:
      text.insert(END,"X=%s, Predicted = %s" % (test[i], 'Eruption Activity Detected at
Given Time')+"\n\n")

def rocGraph():
  predict = cls1.predict(X_test)
  svm_fpr, svm_tpr, _ = roc_curve(y_test, predict)
  predict = cls2.predict(X_test)
```

```
lr_fpr, lr_tpr, _ = roc_curve(y_test, predict)

predict = cls3.predict(X_test)
rf_fpr, rf_tpr, _ = roc_curve(y_test, predict)

predict = cls4.predict(X_test)
g_fpr, g_tpr, _ = roc_curve(y_test, predict)

plt.plot(svm_fpr, svm_tpr, linestyle='--', label='SVM')
plt.plot(lr_fpr, lr_tpr, linestyle='--', label='Logistic Regression')
plt.plot(rf_fpr, rf_tpr, linestyle='--', label='Random Forest')
plt.plot(g_fpr, g_tpr, linestyle='--', label='Gaussian Process')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.legend()
plt.show()
font = ('times', 16, 'bold')
title = Label(main, text='Understanding the Timing of Eruption End using a Machine Learning Approach to Classification of Seismic Time Series')
title.config(bg='firebrick4', fg='dodger blue')
title.config(font=font)
title.config(height=3, width=120)
title.place(x=0,y=5)

font1 = ('times', 12, 'bold')
text=Text(main,height=20,width=150)
scroll=Scrollbar(text)
text.configure(yscrollcommand=scroll.set)
text.place(x=50,y=120)
text.config(font=font1)
font1 = ('times', 13, 'bold')
uploadButton = Button(main, text="Upload Seismic Dataset", command=upload, bg='#ffb3fe')
```

```
uploadButton.place(x=50,y=550)
uploadButton.config(font=font1)

processButton = Button(main, text="Preprocess Dataset Feature Extraction",
command=preprocess, bg='#ffb3fe')
processButton.place(x=270,y=550)
processButton.config(font=font1)

svmButton1 = Button(main, text="Run SVM Algorithms", command=runSVM, bg='#ffb3fe')
svmButton1.place(x=610,y=550)
svmButton1.config(font=font1)

lrButton = Button(main, text="Run Logistic Regression", command=runLR, bg='#ffb3fe')
lrButton.place(x=50,y=600)
lrButton.config(font=font1)

rfButton = Button(main, text="Run Random Forest Algorithm",
command=runRandomForest, bg='#ffb3fe')
rfButton.place(x=270,y=600)
rfButton.config(font=font1)

gpButton = Button(main, text="Run Gaussian Process Classifier", command=runGaussian,
bg='#ffb3fe')
predictButton.place(x=350,y=650)
predictButton.config(font=font1)

rocButton = Button(main, text="ROC Curve Graph", command=rocGraph, bg='#ffb3fe')
rocButton.place(x=520,y=650)
rocButton.config(font=font1)
```

```
main.config(bg='LightSalmon3')
main.mainloop()
```

map.py

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.basemap import Basemap
import warnings
warnings.filterwarnings('ignore')
dataset = pd.read_csv("database.csv")
dataset = dataset[['Latitude','Longitude','Magnitude']]
m = Basemap(projection="mill")
longitudes = dataset["Longitude"].tolist()
latitudes = dataset["Latitude"].tolist()
x,y = m(longitudes,latitudes)
fig = plt.figure(figsize=(12,10))
plt.title("All affected areas")
m.scatter(x,y, s = 4, c = "blue")
m.drawcoastlines()
m.fillcontinents(color='coral',lake_color='aqua')
m.drawmapboundary()
m.drawcountries()
plt.show()
minimum = dataset["Magnitude"].min()
maximum = dataset["Magnitude"].max()
average = dataset["Magnitude"].mean()
print("Minimum:", minimum)
print("Maximum:",maximum)
print("Mean",average)
(n,bins, patches) = plt.hist(dataset["Magnitude"], range=(0,10), bins=10)
plt.xlabel("Earthquake Magnitudes")
```

```
plt.ylabel("Number of Occurences")
plt.title("Overview of earthquake magnitudes")
print("Magnitude" +"   "+ "Number of Occurence")
for i in range(5, len(n)):
 print(str(i)+ "-"+str(i+1)+"          "+str(n[i]))
```

```
]: import map
```

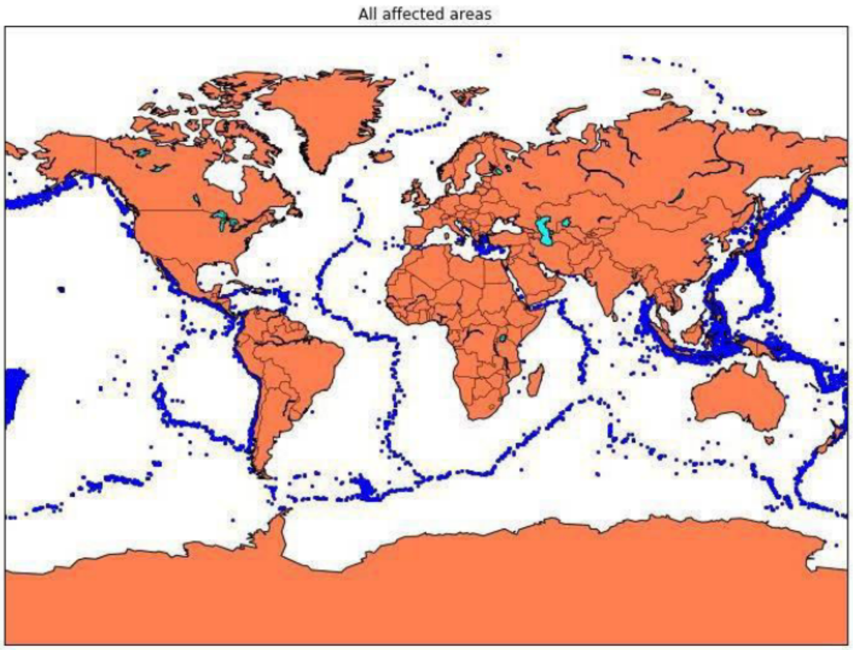


Fig:6.4 BaseMap for dataset 1(earthquakes across the globe from 1965-2016)

The graph below Fig.7.5 shows the number of earthquakes that occurred between different ranges that occurred across the globe.

```
('Minimum:', 5.5)
('Maximum:', 9.1)
('Mean', 5.882530753459764)
Magnitude   Number of Occurence
5-6         16058.0
6-7         6616.0
7-8         698.0
8-9         38.0
9-10         2.0
```

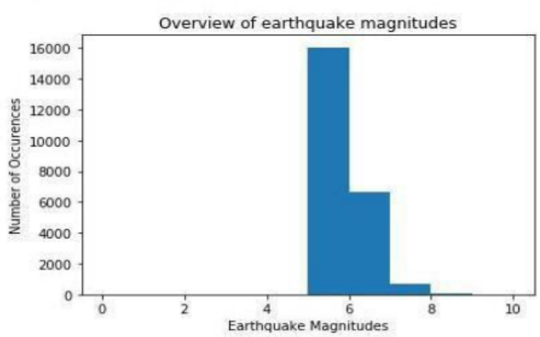
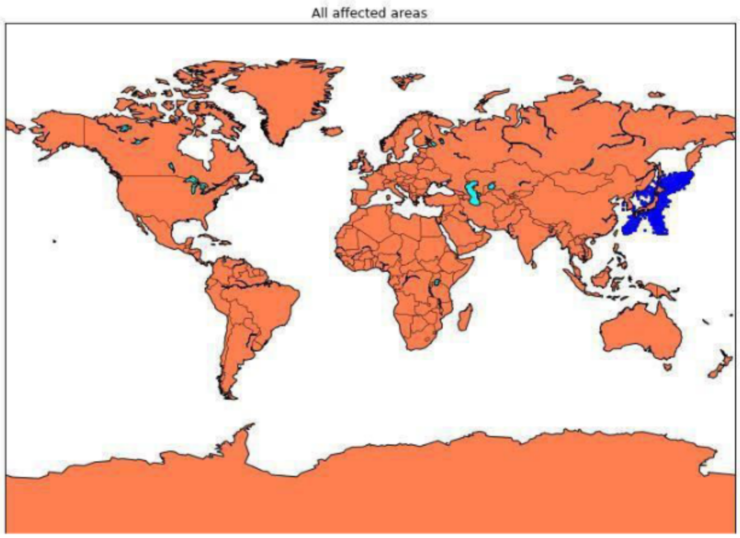


Fig.6.5 Number of earthquakes that occurred across the globe from 1965-2016.

Fig.6.6 Number of earthquakes that occurred in Japan from (2002-2018)

```
import japan
```



CHAPTER 6: RESULT

6.1 LIMITATIONS

Limitations of our model: A model cannot be universal. Our model is applicable in the regions of ocean-ocean convergence. In ocean-ocean convergence, two oceanic plates converge or collide. The denser plate subducts beneath the convergence zone into the asthenosphere, forming a trench at the surface. The zone of subduction is the region beneath the convergence zone. Our model may or may not be applicable for ocean-continent convergence or continent-continent convergence because for a volcano to form, the presence of magma is essential. This entirely depends on the depth of subduction. For E.g. Earthquakes of magnitude greater than six have occurred in the Himalayas, but there has been no volcanic eruption. The Himalayas were formed by the collision of the Indo-Australian plate (continental plate) and the Eurasian plate (continental plate). However, the subduction of the Indian plate was not so deep that the subducted plate melted to form magma. As a result, there is no volcanic eruption in the Himalayas. Hence instead of continent-continent convergence, we restricted our model to ocean-ocean convergence.

RESULT

After analysing the datasets ie. The global dataset, containg the magnitude as well as the location of earthquakes from 1965 to 2017 and the Japanese dataset from 1901 to 2021 we roughly estimated that the probability of a volcanic eruption after a major earthquake is 6 to 11% according to our datasets.

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

CONCLUSION:

This project is useful in predicting volcanic eruptions from earthquakes, in the regions which are formed by ocean-ocean convergence like Japan, Philippines etc. Depending on our dataset, the Random forest algorithm performed well. It gave an accuracy of 99%.Whereas SVM, Gaussian Process Classifier and Logistic regression gave us an accuracy of (53-55).

FUTURE SCOPE:

The project has been implemented by calculating the accuracies of the Random Forest algorithm, SVM, Gaussian Process Classifier, and Logistic regression. Though our model is limited to the regions formed by ocean-ocean convergence, the model can be extended to the regions formed by continent-continent convergence and ocean-continent convergence by conducting additional research on the correlation between earthquakes and volcanoes.

.

123

ORIGINALITY REPORT

7%	4%	5%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Gitam University Student Paper	2%
2	curve.coventry.ac.uk Internet Source	1%
3	cms.geolsoc.org.uk Internet Source	1%
4	Submitted to University of Liverpool Student Paper	1%
5	ir.canterbury.ac.nz Internet Source	1%
6	Ken'Ichiro Yamashina, Kazuaki Nakamura. "Correlations between tectonic earthquakes and volcanic activity of izu-oshima volcano, japan", Journal of Volcanology and Geothermal Research, 1978 Publication	1%
7	Watt, S.F.L.. "The influence of great earthquakes on volcanic eruption rate along the Chilean subduction zone", Earth and Planetary Science Letters, 20090130	<1%

Publication

8 Servando De la Cruz-Reyna, Marta Tárraga, Ramón Ortiz, Alicia Martínez-Bringas. "Tectonic earthquakes triggering volcanic seismicity and eruptions. Case studies at Tungurahua and Popocatépetl volcanoes", Journal of Volcanology and Geothermal Research, 2010 <1%

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		