Understanding the link between earthquakes and volcanic eruptions using machine learning techniques

Arshad Ahmad Khan Mohammed
Assistant Professor (C.S.E),
GITAM University,
Hyderabad, India.
amohamma2@gitam.edu

N Madhumita
Student, B.tech (C.S.E),
GITAM University,
Hyderabad, India.
221710304036@gitam.in

Vantipalli Pravarsha
Student, B.tech (C.S.E),
GITAM University,
Hyderabad, India.
221710304064@gitam.in

Jujaray Naveen
Student, B.tech (C.S.E),
GITAM University,
Hyderabad, India.
221710304022@gitam.in

Pitchika Raghavendra Rao
Student, B.tech (C.S.E),
GITAM University,
Hyderabad, India.
221710304044@gitam.in

ABSTRACT

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, but there are three main factors that trigger an eruption, the first factor is the buoyancy of the magma; the second factor is the pressure due to the gases dissolved in the magma; the third factor: Injection of a new batch of magma into an already filled magma chamber. It also is believed that sometimes tectonic earthquakes can also cause volcanic eruptions.

The existing model is using two different volcanoes , Nevado del Ruiz(Colombia) and Telica(Nicaragua). Nevado del Ruiz is formed by the subduction of the Nazca oceanic plate beneath the South American continental plate (which is infact an ocean-continent convergence)and Telica in Nicaragua is formed by the subduction of Cocos Plate beneath the Caribbean Plate (which is infact ocean-ocean convergence). Various supervised machine learning classification methods like Support Vector Machine, Logistic Regression, Random Forest and Gaussian Process Classifiers have been used for predicting the eruption of the two strato volcanoes as mentioned above

In the proposed model we predict the probability of a volcanic eruption being triggered by an earthquake. An earthquake which has a magnitude greater than 6 can trigger a volcanic eruption. Instead of restricting to a specific volcano this model focus on the regions formed by ocean-ocean convergence. Eg. Japan . Here the Pacific plate is moving westwards and is being subducted beneath the Okhotsk Plate (Northern part of Japan) . Various supervised machine learning classification methods like Support Vector Machine, Logistic Regression, Random Forest and Gaussian Process Classifiers have been used for predicting the eruption.

KEYWORD'S:- SVM (Support Vector Machine), Logistic Regression, Gaussian Processes Classifier, Random Forest Algorithm, TKinter Tool.

1. INTRODUCTION

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 Phillipines. 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.

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 FORMATION OF JAPANESE ISLAND

- 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.
- As the sediment-laden ocean floor crust (oceanic plate) subducts into the softer asthenosphere, the rocks in the subduction zone metamorphose (change in the composition or structure of a rock) under high pressure and temperature.
- After reaching a depth of about 100 km, the plates melt. Magma (metamorphosed sediments and the melted part of the subducting plate) has a lower density and is at high pressure.
- It rises upwards due to the buoyant force offered by the surrounding denser medium.
- The magma flows out to the surface. A continuous upward movement of magma creates constant volcanic eruptions on the ocean floor.
- Layers of rocks are formed as a result of constant volcanism above the subduction zone. As this process continues over millions of years, a volcanic landform is formed, which in some cases rises above the ocean.
- Such volcanic landforms all along the boundary form a chain of volcanic islands known as Island Arcs

(Indonesian Island Arc or Indonesian Archipelago, Philippine Island Arc, Japanese Island Arc etc.).

- Orogenesis (mountain formation) initiates the process of forming continental crust by replacing oceanic crust (this occurs much later). Every few years, for example, new islands appear around Japan. Japan will be a single land after a million years.

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

Fig: Only earthquakes above six can cause a volcanic

2.METHODS

At its most basic, machine learning uses programmed algorithms that receive and analyse input data to predict output values within an acceptable range. As new data is fed to these algorithms, they learn and optimise their operations to improve performance, developing 'intelligence' over time. We used supervised machine learning techniques , wherin the machine is given a set of labelled data ,which is called the train data. After that a new set of data is provided which is the test data. The machine must be able to produce the correct results for the new data based on the previous labelled data. We have focused on three features for feature extraction: the latitide, longitude and magnitude. Using the latitude and longitude we can locate the epicentre of the earthquake. Magnitude represents the intensity of the earthquake. Using the magnitude we can classify whether an eruption can occur or not. We have used four supervised machine learning classification methods: Support Vector Machine, Logistic Regression, Random Forest and Gaussian process classifiers.

2.1 *Support Vector Machine(SVMs)*

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 SVM in a probabilistic classification setting).

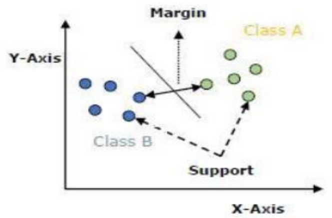


Fig: SVM

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, 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 categorisation -SVMs support text and hypertext categorisation in both inductive and transductive models. They classify documents into different categories using training data. It categorises 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.

2.2 *Logistic Regression(LR)*

Logistic regression is a classification algorithm that uses supervised learning to predict the likelihood of a target variable. Because the nature of the target or dependent variable is dichotomous, there are only two possible classes. Simply put, the dependent variable is binary, with data coded as either 1 (for success/yes) or 0 (for failure/no). A logistic regression model predicts P(Y=1) as a function of X mathematically. It is one of the most basic ML algorithms. It can be used to solve various classification problems such as spam detection, diabetes prediction, cancer detection, etc.

Logistic regression is classified into the following types:

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.

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.

Assumptions for Logistic Regression: The target variables in binary logistic regression must always be binary, and the desired outcome is represented by factor level 1.The model should not have any multi-collinearity, which means that the independent variables must be independent of one another. In order for our

model to be meaningful, we must include meaningful variables. For logistic regression, we should use a large sample size.
Applications of Logistic regression: We are using healthcare to identify disease risk factors and plan preventive measures.We use a weather forecasting app to forecast snowfall and weather conditions. We use voting apps to determine whether voters will vote for a specific candidate.To forecast whether a loan applicant will be approved or denied.

2.3 Random Forest(RF)

Random Forest is a well-known machine learning algorithm from the supervised learning technique. It can be applied to both classification and regression problems in machine learning. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and improve the model's performance."Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset," as the name implies. Instead of relying on a single decision tree, the random forest takes the predict ions from each tree and makes decisions based on the values. Random forest constructs multiple decision trees and 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.

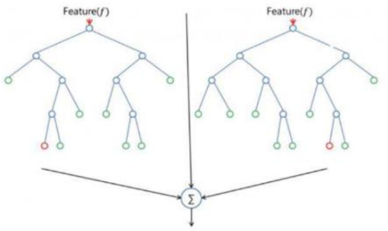


Fig: Random forest with two trees

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 Applications: Random Forest algorithm is used in banking to find loyal customers, which our customers who can take out many loans and pay their interest to the bank on time, and fraud customers, who are customers who have bad records, such as failing to pay back a loan on time or engaging in dangerous behaviour. Random Forest algorithm can be used in medicine to identify the correct combination of medicine components and identify diseases by analysing the patient's medical records. Random Forest algorithm can be used in the stock market to identify a stock's behavior and the expected loss or profit. In order to be used in e-commerce.

2.4 Gaussian Process classification (GPs)

Gaussian Processes generalises the Gaussian probability distribution that can be used to underpin sophisticated non-parametric machine learning algorithms for classification and regression. They are a type of kernel model, similar to SVMs, and unlike SVMs, they can predict highly calibrated class membership probabilities, though the selection and configuration of the kernel used at the heart of the method can be difficult. Its 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. A Gaussian process machine-learning algorithm uses lazy learning and a measure of point similarity (the kernel function) to predict the value of an unseen point from training data. The prediction is a one-dimensional Gaussian distribution with uncertainty information, not just an estimate for that point. Multivariate Gaussian processes are used for multi-output predictions. The multivariate Gaussian distribution is the marginal distribution at each point.

Because it is based on the Gaussian distribution, the concept of Gaussian processes is named after Carl Friedrich Gauss (normal distribution). Gaussian processes can be thought of as an infinite-dimensional extension of multivariate normal distributions. Gaussian processes are helpful in statistical modelling because they inherit properties from the normal distribution. For example, the distributions of various derived quantities can be obtained explicitly if a random process is modelled as a Gaussian process. The average value of the process over a range of times and the error in estimating the average using sample values at a small set of times are
examples of such quantities.

Applications of Gaussian Process Classifier:

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

3. MODEL

Our project aims to predict the possibility of a volcanic eruption if an earthquake occurs. We want to link earthquakes and volcanic eruptions. Our classifications have a higher correspondence of eruptive activity with visual activity than with database records of eruption start and end. We analyze the relative importance of the different features of seismic activity used in our models (e.g. peak event amplitude, daily event counts) and find little consistency between the two volcanic systems in terms of the most important features Identifying the critical thresholds which govern when large-scale changes in volcanic behaviour occur is acknowledged as one of the fundamental research questions associated with understanding the beginning, evolution and termination of volcanic activity. Development of models for these critical thresholds is necessary to understand the processes which drive large-scale change in volcanic behaviour, but this in turn requires knowing when these changes occur in the timeline of an eruption.

which determine whether activity is eruptive or non-eruptive. These initial results look promising and our approach may offer a robust tool to help determine when an eruption has ended in the absence of visual confirmation. Machine learning methods can successfully classify overall patterns of eruptive and non-eruptive behaviour in seismic time series. This study is the first to apply machine learning techniques to singlestation seismic data to classify overall volcanic state as eruptive or non-eruptive. We define a decisiveness index to evaluate classification of eruptive state based on the consistency of classification, which is comparable across datasets. Our models have a high agreement in terms of eruptive classification with visual indicators of eruption, such as ash emissions.

Features: We are focusing on the following features:
Latitude, Longitude and Magnitude

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.

3.1 SYSTEM DESIGN:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. An activity diagram visually presents a series of actions or flow of control in a system similar to a flowchart or a data flow diagram. Activity diagrams are often used in business process modeling. They can also describe the steps in a use case diagram. Activities modeled can be sequential and concurrent.

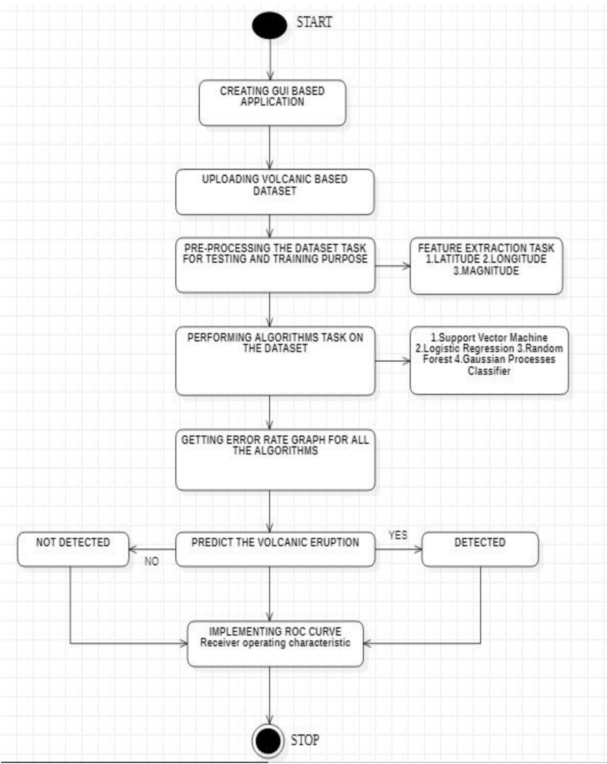


Fig 3.1: Model Design

4.IMPLEMENTATION

4.1 FUNCTIONAL REQUIREMENTS:

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation. The appropriation of requirements and implementation constraints gives the general overview of the project regarding the areas of strength and deficit and how to tackle them. Python idle 3.7 version (or) Anaconda 3.7 (or) Jupiter (or) Google colab

Hardware Requirements: Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM. In contrast, applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

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

4.2 PYTHON PACKAGES:

NumPy: NumPy includes the tools for purpose of 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 analysing, cleaning, exploring, and manipulating data. Pandas enables us to draw conclusions and analyse large amounts of data using statistical theories. We can use pandas to clean up messy data sets and make them relevant and readable.

The pandas. read_csv() function enables us to load the dataset from the system. You can find the dataset here. As the dataset contains categorical variables as well, we have thus created dummies of the categorical features for an ease in modelling using pandas.

Tkinter: Tkinter is the most important and widely used framework for creating graphical user interfaces. These include buttons, radio buttons, checkboxes, etc. It connects Python to the TK GUI toolkit, which runs on almost every modern operating system.

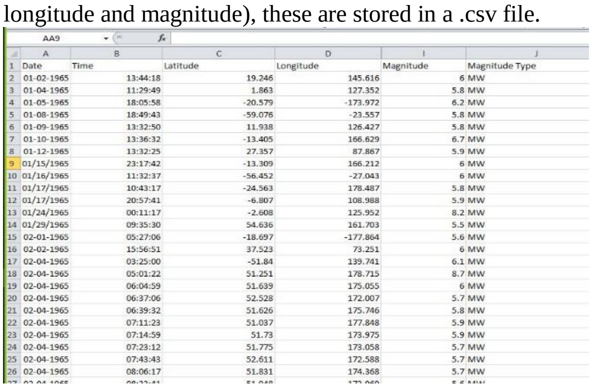
4.3 DATASET

This dataset contains the seismic magnitude of many volcanoes. If its value is greater than ‘6.0’, then it will be considered that a volcano is about to erupt.

We had used 80% of dataset records to be trained Machine Learning algorithms and 20% of the records will be tested to calculate its classification accuracy.

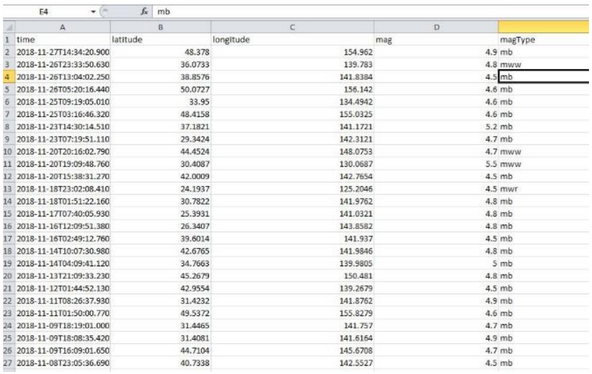
We have used two datasets: The first dataset consists of earthquakes across the globe from 1965-2016. This dataset consists of 23413 rows.

The second dataset is of earthquakes in Japan from 2001-2018. This dataset has 18751 rows. Based on the features(latitude, longitude and magnitude), these are stored in a .csv file.



Date	Time	Latitude	Longitude	Magnitude	Magnitude Type
01-02-1965	13:44:18	19.246	145.616	6	MW
01-04-1965	11:29:49	1.863	127.352	5.8	MW
01-05-1965	18:05:58	-20.579	-173.972	6.2	MW
01-08-1965	18:49:43	-59.076	-23.557	5.8	MW
01-09-1965	13:32:50	11.938	126.427	5.8	MW
01-10-1965	13:36:32	-13.405	166.629	6.7	MW
01-12-1965	13:32:25	27.357	87.867	5.9	MW
01/15/1965	23:17:42	-13.309	166.212	6	MW
01/16/1965	11:32:37	-56.452	-27.043	6	MW
01/17/1965	10:43:17	-24.563	178.487	5.8	MW
01/17/1965	20:57:41	-6.807	108.988	5.9	MW
01/24/1965	00:11:17	-2.608	125.952	8.2	MW
01/29/1965	09:35:30	54.636	161.703	5.5	MW
02-01-1965	05:27:06	-18.697	-177.864	5.6	MW
02-02-1965	15:56:51	37.523	73.251	6	MW
02-04-1965	03:25:00	-51.84	139.741	6.1	MW
02-04-1965	05:01:22	51.251	178.715	8.7	MW
02-04-1965	06:04:59	51.639	175.055	6	MW
02-04-1965	06:37:06	52.528	172.007	5.7	MW
02-04-1965	06:39:32	51.626	175.746	5.8	MW
02-04-1965	07:11:23	51.037	177.848	5.9	MW
02-04-1965	07:14:59	51.73	173.975	5.9	MW
02-04-1965	07:23:12	51.775	173.058	5.7	MW
02-04-1965	07:43:43	52.611	172.588	5.7	MW
02-04-1965	08:06:17	51.831	174.368	5.7	MW

Fig: DataSet-1 earthquakes across the globe from (1965-2016)



	A	B	C	D	
1	time	latitude	longitude	mag	magType
2	2018-11-27T14:34:20.900	48.378	154.962	4.9	mb
3	2018-11-26T23:33:50.630	36.0733	139.783	4.8	mww
4	2018-11-26T13:04:02.250	38.8576	141.8384	4.5	mb
5	2018-11-26T05:20:16.440	50.0727	156.142	4.6	mb
6	2018-11-25T09:19:05.010	33.95	134.4942	4.6	mb
7	2018-11-25T03:16:46.320	48.4158	155.0325	4.6	mb
8	2018-11-23T14:30:14.510	37.1821	141.1721	5.2	mb
9	2018-11-23T07:19:51.110	29.3424	142.3121	4.7	mb
10	2018-11-20T20:16:02.790	44.4524	148.0753	4.7	mww
11	2018-11-20T19:09:48.760	30.4087	130.0687	5.5	mww
12	2018-11-20T15:38:31.270	42.0009	142.7654	4.5	mb
13	2018-11-18T23:02:08.410	24.1937	125.2046	4.5	mwr
14	2018-11-18T01:51:22.160	30.7822	141.9762	4.8	mb
15	2018-11-17T07:40:05.930	25.3931	141.0321	4.8	mb
16	2018-11-16T12:09:51.380	26.3407	143.8582	4.8	mb
17	2018-11-16T02:49:12.760	39.6014	141.937	4.5	mb
18	2018-11-14T10:07:30.980	42.6765	141.9846	4.8	mb
19	2018-11-14T04:09:41.120	34.7663	139.9805	5	mb
20	2018-11-13T21:09:33.230	45.2679	150.481	4.8	mb
21	2018-11-12T01:44:52.130	42.9554	139.2679	4.5	mb
22	2018-11-11T08:26:37.930	31.4232	141.8762	4.9	mb
23	2018-11-11T01:50:00.770	49.5372	155.8279	4.6	mb
24	2018-11-09T18:19:01.000	31.4465	141.757	4.7	mb
25	2018-11-09T18:08:35.420	31.4081	141.6164	4.9	mb
26	2018-11-09T16:09:01.650	44.7104	145.6708	4.7	mb
27	2018-11-08T23:05:36.690	40.7338	142.5527	4.5	mb

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

In the screen below, 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 "Running each and every Algorithm' button to train the SVM model with the previous dataset.

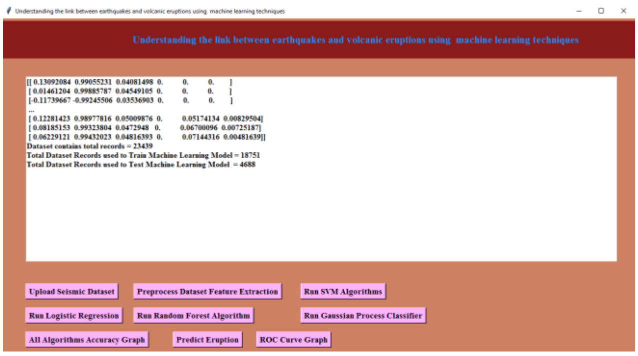


Fig: Pre-processing the Dataset (GUI Application)

After processing each and every algorithm, we get below mentioned accuracy values:

SVM Accuracy on Eruption Dataset: 67.81143344709898

Logistic Regression Accuracy on Eruption Dataset: 67.74744027303754

Random Forest Accuracy on Eruption Dataset: 99.74402730375427

Gaussian Process Classifier Accuracy on Eruption Dataset: 67.83276450511946

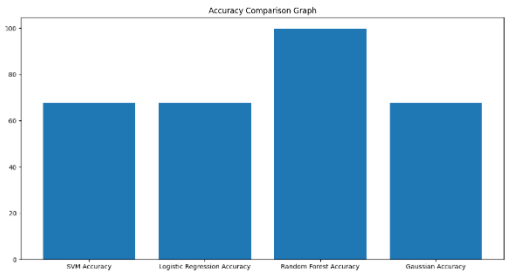
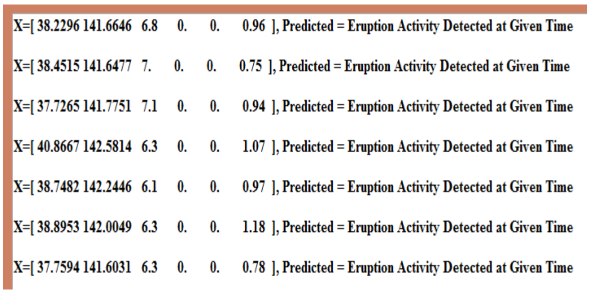


Fig: Algorithm's Accuracy and it's Graph



X=[38.2296 141.6646 6.8 0. 0. 0.96], Predicted = Eruption Activity Detected at Given Time

X=[38.4515 141.6477 7. 0. 0. 0.75], Predicted = Eruption Activity Detected at Given Time

X=[37.7265 141.7751 7.1 0. 0. 0.94], Predicted = Eruption Activity Detected at Given Time

X=[40.8667 142.5814 6.3 0. 0. 1.07], Predicted = Eruption Activity Detected at Given Time

X=[38.7482 142.2446 6.1 0. 0. 0.97], Predicted = Eruption Activity Detected at Given Time

X=[38.8953 142.0049 6.3 0. 0. 1.18], Predicted = Eruption Activity Detected at Given Time

X=[37.7594 141.6031 6.3 0. 0. 0.78], Predicted = Eruption Activity Detected at Given Time

Fig: Predict Eruption Function

Here (Fig.), 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 >=6.0, it classes that record time as 'eruption activity detected'.

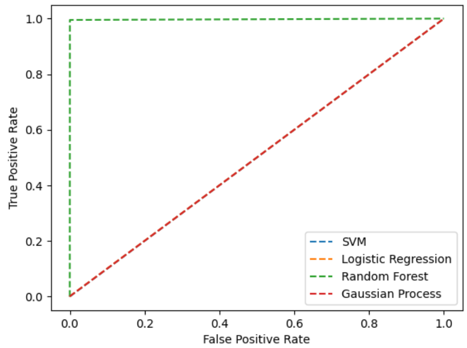


Fig: ROC Curve Graph

A useful tool when predicting the probability of a binary outcome is the Receiver Operating Characteristic curve, or ROC curve. It is a plot of the false positive rate (x-axis) versus the true positive rate (y-axis) for a number of different candidate threshold values between 0.0 and 1.0.

Probability Derivation:

$$P[X = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Where **X** is the number of volcano eruptions within unit time **t** , waiting time between volcano eruptions is a random variable with an exponential distribution with rate:

λ=0.10 eruptions/year and **t** is time(1 year)

There are ove a 100 active volcanoes in Japan. We assumed that out of 100 if 10 volcanoes erupt(10/100=0.10) and if there is 1 volcano which has been triggered by an earthquake , we would like to find out the probability of such eruptions triggered by an earthquake over a period of 60 years.

Eg: **λt** =0.10* 60=6

P[X<=9]= P[X=6]+P[X=7]+P[X=8]+P[X=9]

$P[X=6] = (6^6\ e^{-6})/6! = 0.16062$

$P[X=7] = (6^7\ e^{-7})/7! = 0.0506$

$P[X=8] = (6^8\ e^{-8})/8! = 0.01397$

$P[X=9] = (6^9 e^{-9})/9! = 0.003427$
$P[X<=9] = 0.228617$
We consider that value of x should be between 6 to 9 because the magnitude values less than 6 will be considered as Non-Eruptive and Greater than 6 will be considered as Eruptive volcano and the range is taken upto 9 because the magnitude value greater than 9 has not been occurred till now either in present situation nor in history. Finally, After Probability Derivation we get approx. value of **0.23.**

5. RESULT

Based on our datasets and depending on the frequency of the volcanic eruption in Japan we can say that the probabilty that an earthquake can trigger a volcanic eruption over a period of 60 years is 0.23 or 23%.

6.CONCLUSION AND FUTURE SCOPE

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 (63-68).

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.

REFERENCES

1) Brodsky, E. E.,B. Sturtevant, andH. Kanamori, Earthquakes, volcanoes, and rectified diffusion, J. Geophys. Res., 103, 23,827–23,838, 1998.
2) Dziewonski, A. M.,T. A. Chou, andJ. H. Woodhouse, Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, J. Geophys. Res., 86, 2825–2852, 1981.
3) Heiken, G., Will Vesuvius erupt? Three million people need to know, Science, 286, 1685–1687, 1999.
4) Simkin, T., and L. Siebert, Volcanoes of the World, Geosciences, Tucson, Ariz., 1994.
5) Yokoyama, I., Volcanic eruptions triggered by tectonic earthquakes, Geophys. Bull. Hokkaido Univ., 25, 129–139, 1971.E-websites
6) Kelly, P. M., and C. B. Sears, Climatic impact of explosive volcanic eruptions, Nature, 311, 740–743, 1984.
7) Kerr, R. A., Can great quakes extend their reach? Science, 280, 1194– 1195, 1998
8) Marzocchi, W., R. Scandone, and F. Mulargia, The tectonic setting of Mount Vesuvius and the correlation between its eruptions and the earthquakes of the southern Apennines, J. Volcanol. Geotherm. Res., 58, 27–41, 1993
9) Pacheco, J. F., and L. R. Sykes, Seismic moment catalog of large shallow earthquakes, 1900 to 1989, Bull. Seismol. Soc. Am., 82, 1306–1349, 1992.
10) Perez,O. J., Revised world seismicity catalog (1950–1997) for strong (Ms ≥ 6) shallow (h ≤ 70 km) earthquakes, Bull. Seismol. Soc. Am., 89, 335–341, 1999.