MAJOR PROJECT REPORT

19/04/21 **[Batch D08]**

Project title**: Understanding the Timing of Eruption End using a Machine Learning Approach to Classification of Seismic Time Series**

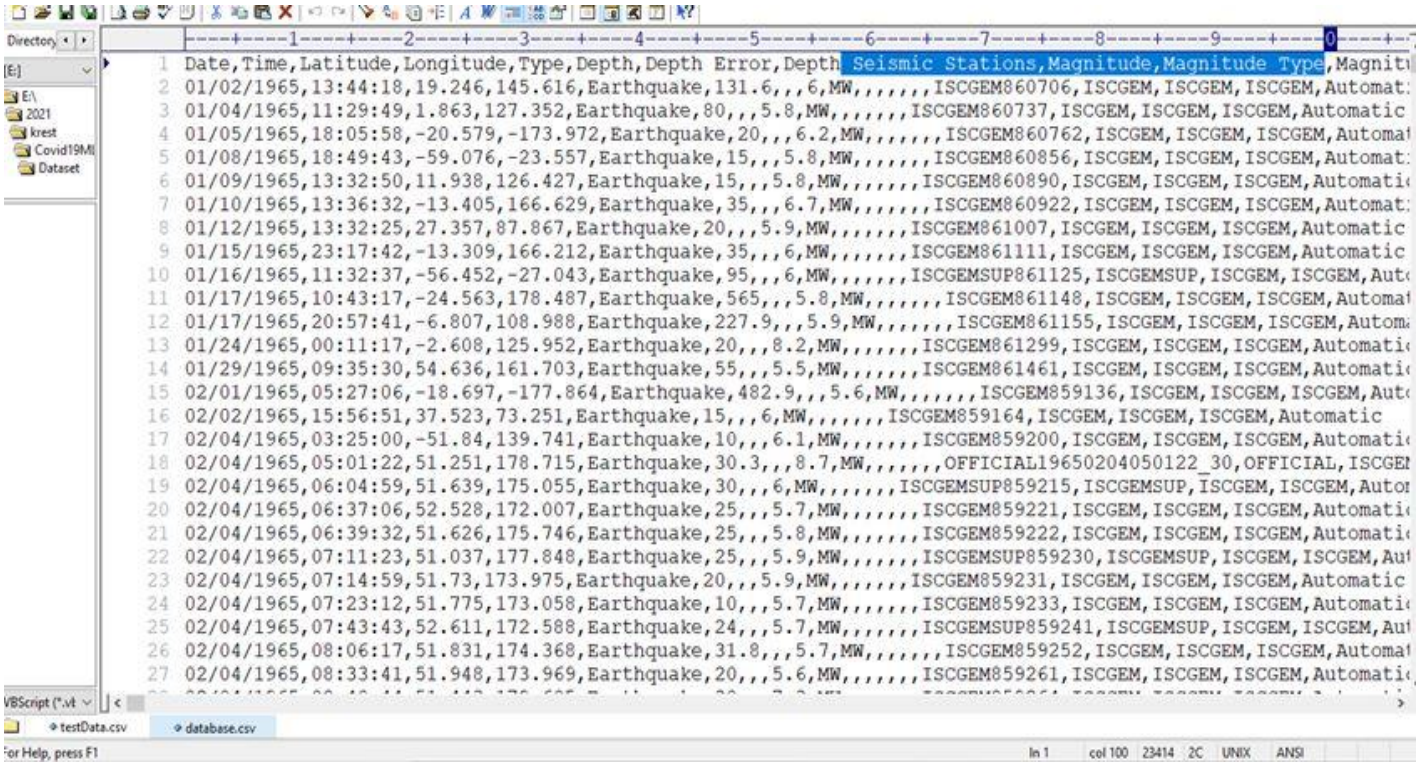
Overview:

In this research paper 2 volcanoes Telica and Nevado del Ruiz were chosen. We wanted to choose the 2 different volcanoes. The 2 volcanoes which we decided to work on were Barren Island(Andaman Islands) and mount vesuvius(Italy).

But unfortunately,despite the rigorous searches we could not get the data sets related to that. Though we got the dataset for mount vesuvius, the data set could only be accessed by the members of the organization. At last we found a dataset from kaggle.

This dataset contains seismic magnitude of many volcanoes and if its value is greater than 6 then it will be considered that a volcano is about to erupt.

We had used 80% dataset records to train Machine Learning algorithms and 20% records to calculate its classification accuracy.



REVIEW 2 FEEDBACK:

1. The review panel asked us to create our own data set. Since volcanic seismic data can only be obtained by sophisticated instruments it is not possible to create our own data set.

2. They asked us to extract the features using deep learning techniques. The problem is we do know much about deep learning. Generally deep learning techniques(like CNN etc.) are suitable when we have data in the form of images etc. Writing an algorithm would be very difficult in our case. In the base papers also they preferred ML over deep Learning for feature extraction, hence we worked on ML.

the training period.

For this preliminary work, we use machine learning models where features are calculated and chosen as inputs to the model, as opposed to other methods such as deep learning wherein features are calculated and chosen within the model. Choosing features as model input is preferred so that we can use features derived from the seismic data that are similar to those used in current monitoring practices. These features, such as event rate or peak signal frequencies, have had widespread success in a monitoring

FEATURES:

We are focusing on the following features:

1. Latitude
2. Longitude
3. Depth
4. Magnitude

IMPLEMENTATION:

Software requirements:

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

Hardware requirements:

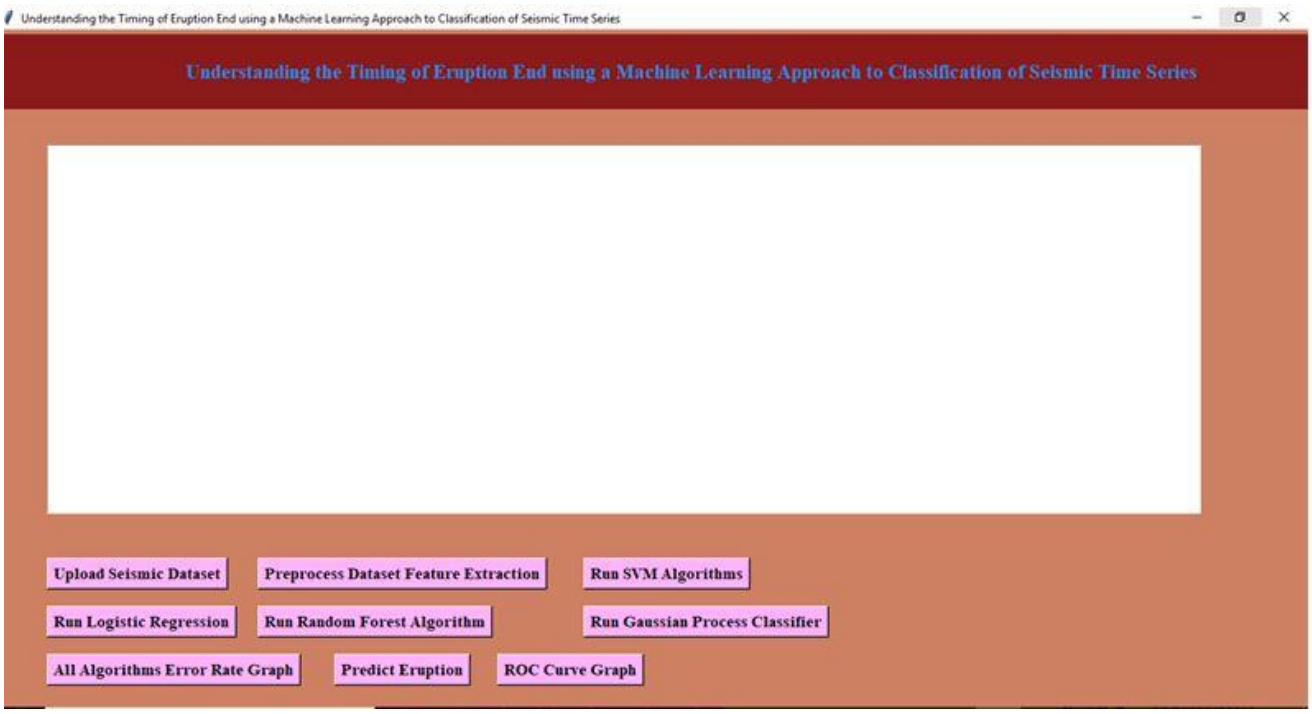
- **Operating system : windows, linux**
- **Processor : minimum intel i3**

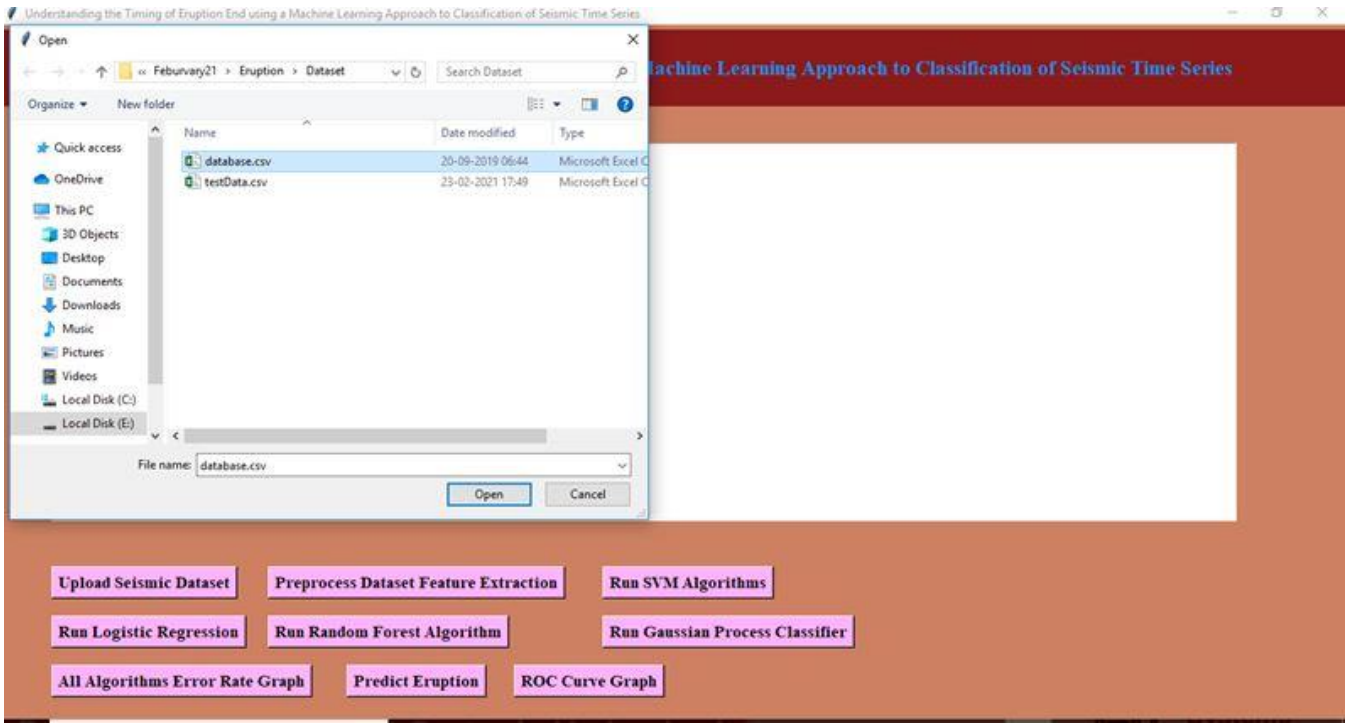
- **Ram : minimum 4 GB**
- **Hard disk : minimum 250 GB**

Four supervised machine learning methods were used to measure the accuracy:

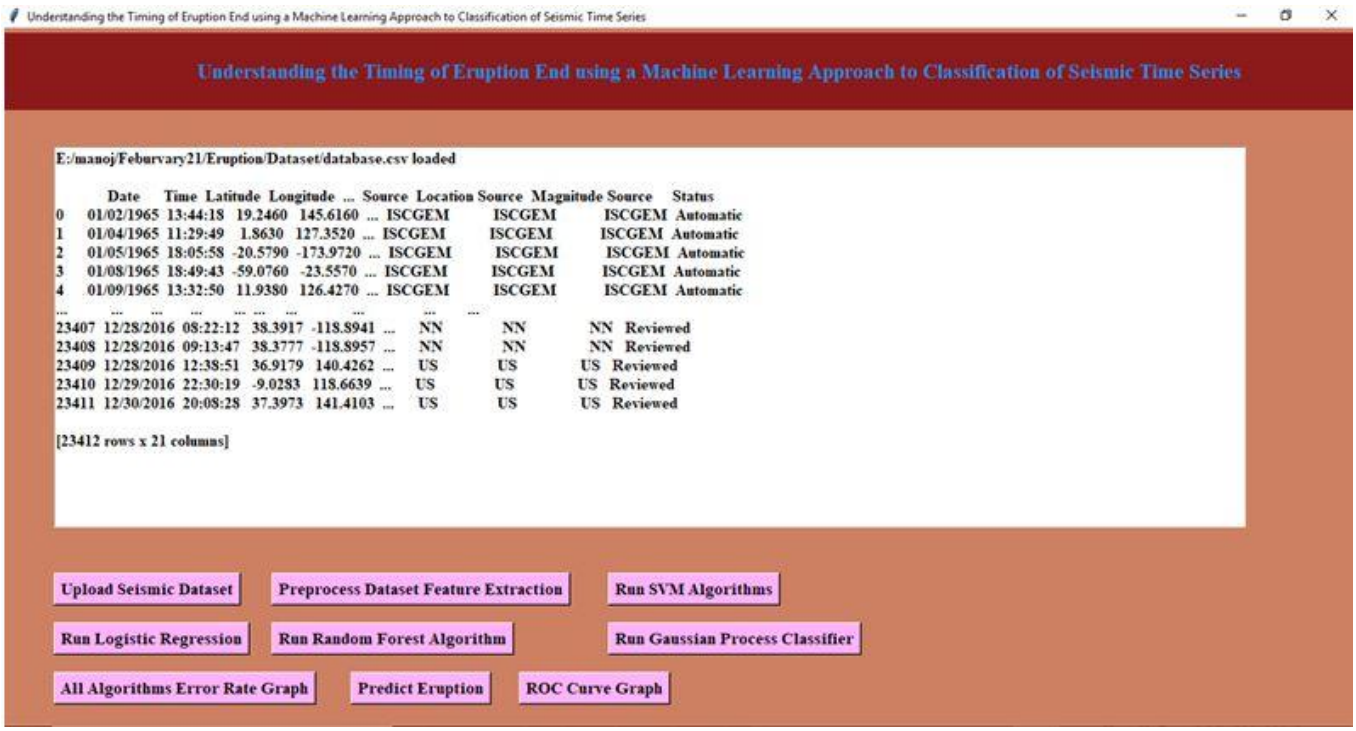
1. SVM
2. Logistic Regression
3. Random Forest
4. Gaussian Process Classifier

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



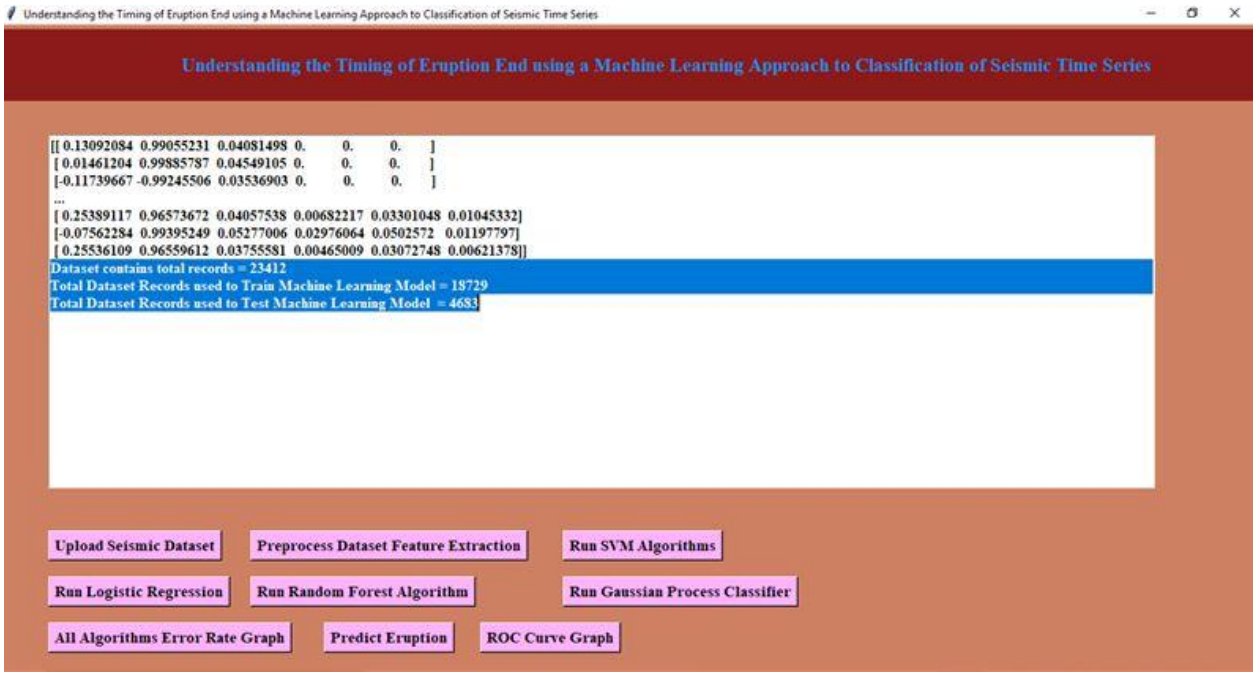


We uploaded the ‘database.csv’ file and then clicked on the ‘Open’ button to load the dataset .



In the above screen dataset is loaded and is displaying certain records. We can see there are 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 normalize format .



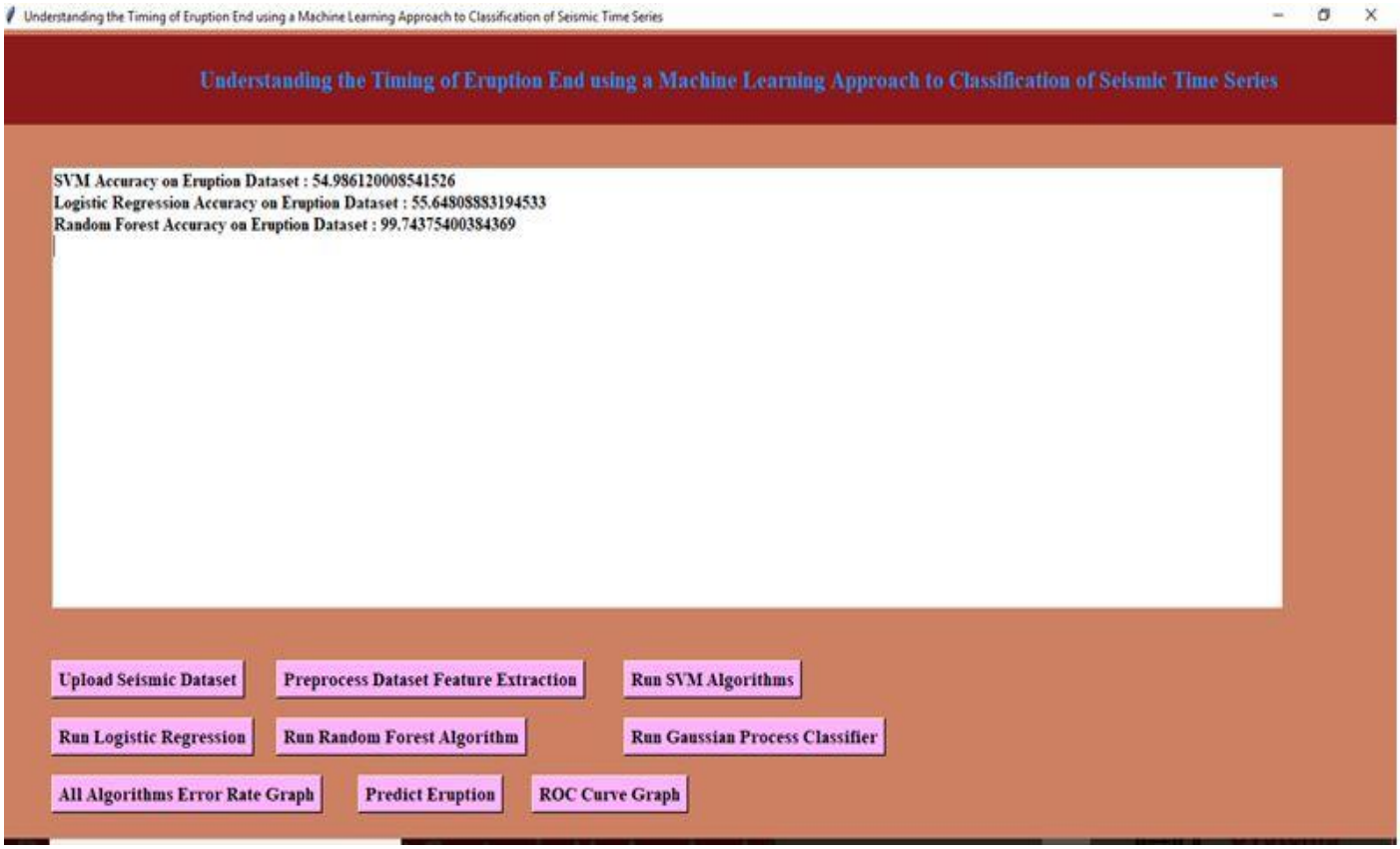
In the above screen all records were converted to numeric values. We can see that the application contains a total 23412 records and application uses 18729 records to train machine learning algorithms and 4683 records to test them.

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.

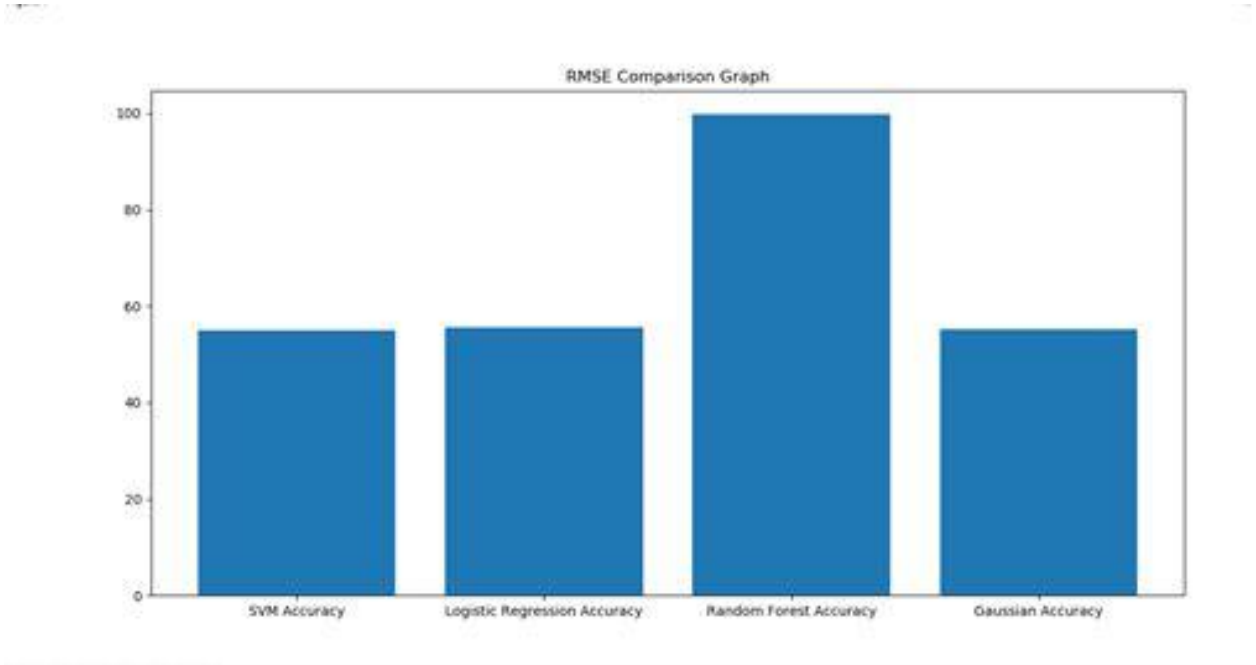
In the above screen, we trained the SVM model and its accuracy is 54%. Next we click on the 'Run Logistic Regression' button to get its accuracy.

In the above screen, we can see that logistic regression got an accuracy of 55%. We move on to ‘Run Random Forest Algorithm’ button to get its accuracy



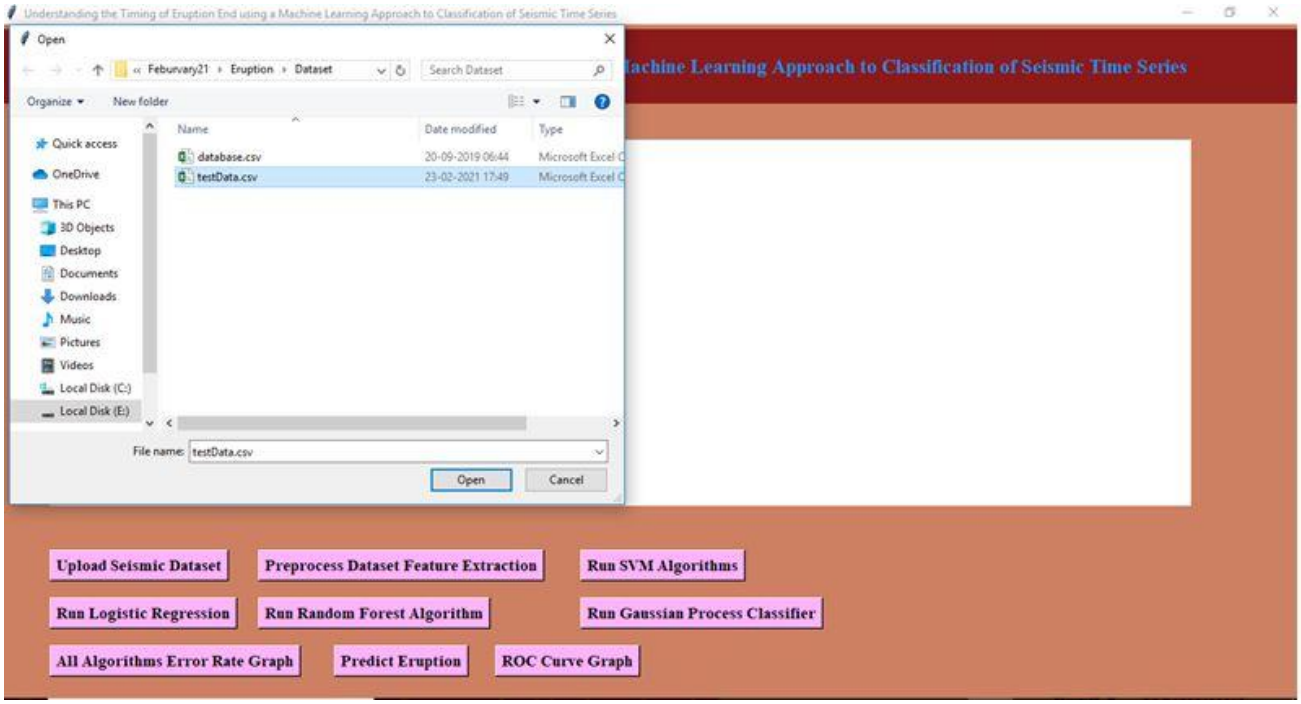
In the above screen, we can see that the random forest algorithm got 99.74% classification accuracy . We now click on the ‘Run Gaussian Process Classifier’ button to get its accuracy.

In the above screen ,we can see that the Gaussian process Classifier has an accuracy rate of 55%.

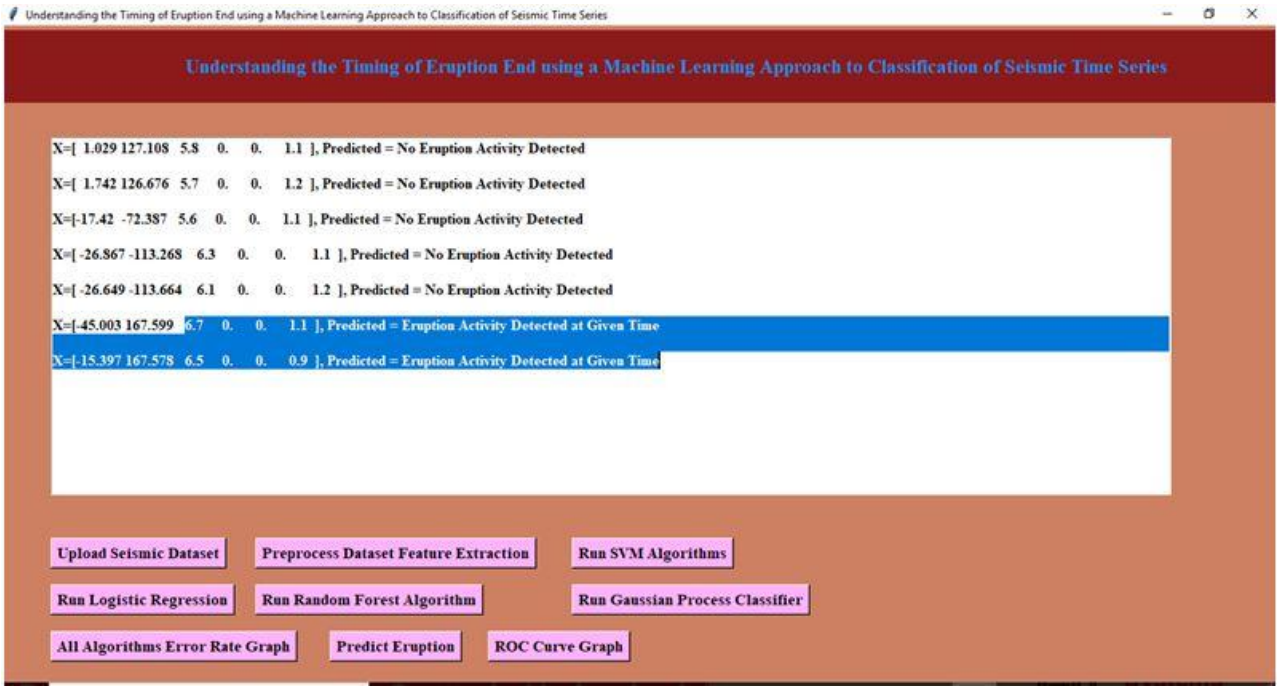


In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can conclude that Random Forest gives better result.

We click on the 'Predict Eruption' button and then upload the test file. The application detects eruption activity from that uploaded time data.



In the above screen we upload the 'testData.csv' file . We then get the below result.Stating whether an eruption will occur or not.



Here, in square brackets we can see the test data and 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 classies that record time as 'eruption activity detected'.

RESULTS

Random forest algorithm performed well for our data set. It gave an accuracy of 99%.Where as SVM, Gaussian Process Classifier and Logistic regression gave us an accuracy of (53-55)%. Whereas in the base paper SVM and Gaussian Process classifier performed well.