Earthquake prediction model using python

Abstract:
An earthquake is a type of natural disaster that is well-known for the devastation it causes to both naturally existing and artificial structures, including buildings, bungalows, and residential areas, to name a few. Seismometers, which pick up vibrations caused by seismic waves moving through the earth's crust, are used to measure earthquakes. The damage caused by an earthquake was categorised in this work into damage ratings, which have values ranging from one to five. The damage grade of a certain structure, which is linked to a Unique Identification String, was predicted using a previously gathered data set and a number of criteria. An analysis of current machine learning classifier techniques was used to make the forecast. Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier, and K-Nearest Neighbors were the machine learning techniques employed in this study. The best algorithm was taken into consideration after a review of a number of attributes. The method used to predict the property underwent a thorough investigation, and the data analysis that followed revealed information that could help future earthquakes' effects be lessened. Regression, K Nearest Neighbors, and predictive analysis.

Keywords: Machine learning, Support Vector Machine (SVM), Random Forest Classifier, Logistic

I. INTRODUCTION

A catastrophic event such as an earthquake is harmful to human interests and has negative effects on the environment. Incalculable harm to buildings and other assets has always been done by earthquakes, which have also claimed millions of lives around the world. Numerous national, international, and transnational organizations implement various disaster warning and preventive strategies to lessen the effects of such an incident. Organization managers have a number of challenges when it comes to allocating the organization's resources because time and quantity are constraints. To estimate the extent of damage done to buildings after an earthquake, it is possible to use machine learning. This is accomplished by categorizing these buildings according to a degree of damage severity based on a number of elements, including their age, foundation, number of floors, kind of material used, and others. Then, ward-by-ward in a district, the number of families and the likely casualties are considered. This enables the proportionate distribution of relief forces by ward and their prioritizing according to the severity of the damage. Such models can contribute to the fastest possible lifesaving and prove to be a successful and affordable option.[1-3] It can be further enhanced by include the distribution of goods like food, clothing, medical care, and money in accordance with the number of fatalities among people and the degree of structural damage.

METHODOLOGY

A. Importing Libraries

The Python code to import libraries. We have used four libraries

- Python has a library called Numpy that is used for scientific computing. This library is utilized throughout the project and is imported as np.
- Pandas are used for data analysis and manipulation. An open source, BSD-licensed library called pandas offers simple data structures and tools for data analysis. It is imported as Pd.
- Matplotlib is a python library. The command-style utilities in pyplot enable matplotlib to behave similarly to MATLAB. It is imported as plt.
- Seaborne is a matplotlib-based Python data visualization package for aesthetically pleasing and educational statistical visuals. It is imported as sns.

B. Importing data

Figure 2 displays the Python code for importing data from the appropriate directory or file and allocating it to It imports the data that is kept in CSV format.

C. Checking for NaN

Checking for NaN is a critical step in the pre-processing of data. We were only able to identify a few NaNs in this test.

D. Manipulating NaN values

It is essential to remove the NaN values. This can be done by

- Removing the entire column containing many NaN values
- Forward fillna method
- Backward fillna method
- Mean method

E. Plotting a Heat map

A heat map is used to assess the correlation between the fields of the collected data. When developing various AI prediction models, the magnitude of the values along with the sign (which may be negative or positive) is crucial.

F. Train/Test split

Creating train and test sets from the data is our next step towards developing a Machine Learning model. The Python code to divide the data set into train

MODEL BUILDING

- The yield of an ML algorithm is a 'model'.
- To begin with, the target variable and feature variable are comprehended and fetched.
- Second, the data-set is partitioned into training and testing data-set and third, the regressor/classifier model is constructed and fitted to training data-set.
- In python, scikit-learn is a simple, basic, efficient open source library that executes a range of machine learning algorithms featuring various classification, regression and clustering algorithms using a unified interface.
- Step by step building is as follows:
- Building A Random Forest Regression Model :
- Random forests are an ensemble learning method that can be fabricated for both regression as well as classification chore.
- It takes on the task of constructing multiple of decision trees during training and outputs the class that is mean prediction (regression) of each individual tree or the mode of the classes (classification).
- This huge number of trees represents a forest. Decision trees are rule based models; on a given training data-set with targets and features, the decision tree algorithm will come up with rules to carry out classification

DATA PRE-PROCESSING

- Data Pre-processing is a technique that converts given and regression.
- Features will be nodes and their presence and absence will represent likeliness.

- This helps in constructing a path of rules to work with.
- The root and splitting node is based on information gain or gini index.
- In Random Forest, the root and splitting nodes are calculated in a random manner

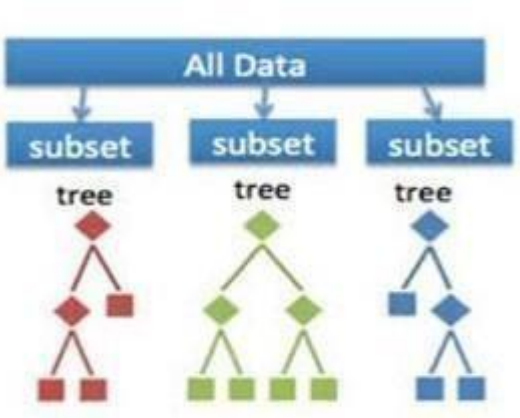


Fig. Random Forest

Therefore random forest is a model comprising of various streets with the capability of making decision based on rule and the procedure of choosing root nodes and parent node sis random.

Building A Support Vector Machine Regression Model:

- SVM segregates different data classes using a decision line named hyperplane.
- When predicting a numerical value, SVR attempts to find a function f(x) in the form of decision boundary at a certain deviation from €, which is a threshold value for all prediction to be within, from obtained targets value Yi, the original hyperplane, such that data points are within the boundary line.
- This decision boundary is the Margin of tolerance - a boundary that allows errors under given range.

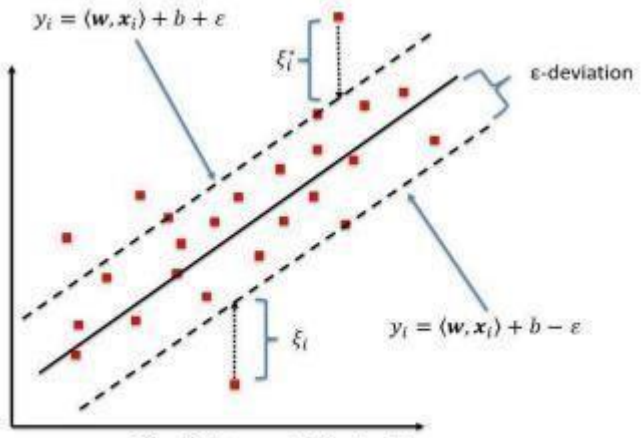


Fig. 2. Support Vector Regressor

Building A Stacking Regress or Model :

Stacking regression is an ensemble learning method. Several regression models collaborate, as a result, meta-regress or is build & itself finds its best fit by making use of output of

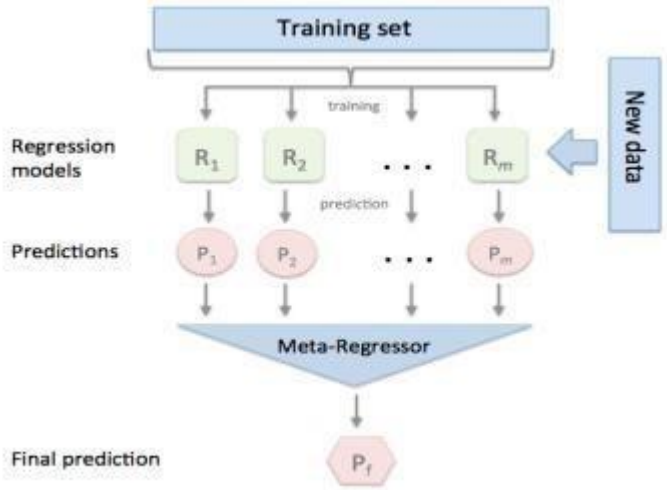


Fig. 3. Stacking

RESULT

- The random forest-support vector machine model in combination work well for large dataset.
- The accuracy obtained for stacking model is the highest- 83% as compared to the accuracy of bagging and boosting. Response time is same for all the methodologies. Training time taken is slightly higher for stacking.
- Results are as follows :

Table- I: Result Table

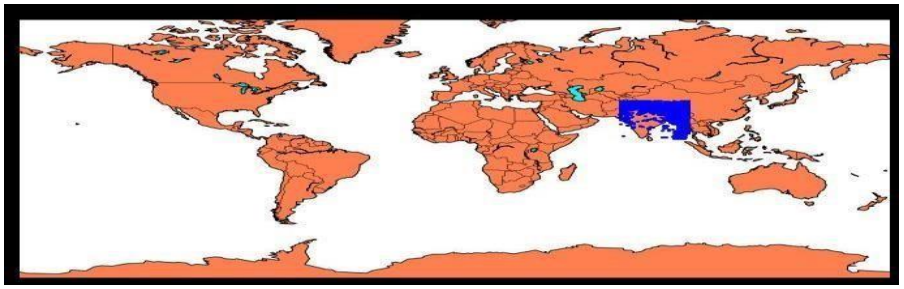
Parametes/ Algorithms	ACCURACY	TRAINING TIME	RESPONSE TIME
Bagging	74%	3m5sec	5 sec
Boosting	76%	3m19sec	5sec
Stacking	83%	11m37sec	5sec

PREDICTIONS

Algorithm:

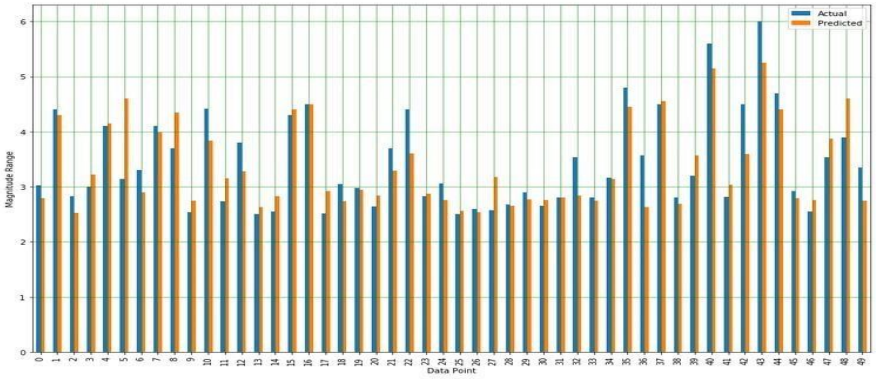
1. Input data-set and load libraries.
2. Data Pre-processing.
3. Model Building.
4. Making Predictions. **Data Visualization:**

1. Affected Areas



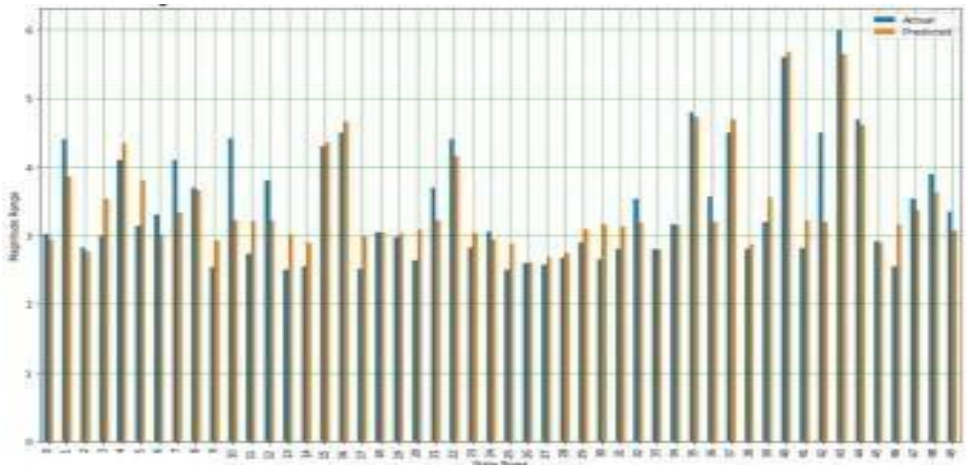
Prediction Using Stacking

Accuracy: 83%



Prediction Using Boosting

Accuracy: 76%



Bar plot for Bagging

CONCLUSION

- Thus we can conclude that integration of seismic activity with machine learning technology yields efficient and significant result and can be used to predict earthquakes widely, given the past history of the same is well maintained.
- Our attempt can be termed successful.
- The collaboration of the two can further be advanced to guard earthquakes more acutely. Large datasets prove to be very significant.
- Prediction models can be deployed in an area- centric manner, thus increasing the chances of accurate prediction exponentially but at the cost of studying algorithms used to build Stacking model, as it will perform well only if the algorithms chosen to build metaregressor are accurate themselves.
- The use of the methodology can be expanded in predicting various natural disasters as well.