Earthquake Prediction using Machine Learning

Hrithik Sarda and S Srinibha | Guide: Dr. Renuga Devi S. | SENSE

Introduction

An earthquake is caused by a sudden slip on a fault. Forecasting earthquakes is one of the most important problems in Earth Science because of their devastating consequences which includes the loss of countless lives and billions of dollars of infrastructure. Mainly, Earthquake forecasting studies focus on three key points: **When, Where and How** big will the earthquake be. The prediction of earthquakes as a whole has proved to be a challenge which is essentially impossible but here we've attempted to use ML algorithms on 2 datasets to predict the Time to Failure in seconds and Magnitude of an Earthquake respectively.

Objective of the Project

- Visualize and locate the affected areas due to earthquake near the Himalayan Plateau using python and data collected from USGS website.
- Acoustic properties of seismic signals underneath the surface of an Earthquake prone area were used to predict the Time to Failure in seconds.
- Geographical properties of the area and data from past earthquake incidents was used to predict the Magnitude of the upcoming earthquakes.

Methodology

Dataset

Parameter	Dataset – I	Dataset – II
Data	Acoustic Data vs Time to Failure	Earthquakes occurred near the Himalayan Plateau.
Area of Study	Los Almos National Laboratory	Himalayan Plateau
Years of Study	2017	26/02/1950- 26/12/2020
Courtesy	Online Website www.kaggle.com	Online Website https://earthquake.usgs.gov/
Forecast Frequency	----	Daily

Process Flowchart

	Dataset 1 (Laboratory data)	Dataset 2 (Real-time data)
Data Pre-processing	• Missing values are filled with mean •Datatype conversions to save computing power • Standard Feature Scaling	• Missing values are filled with mean • Removing time and region columns
Feature Extraction	•Mean •RMS •Standard deviation •Peak •Mean of absolute value • Skewness • Kurtosis •Crest Factor •Log-log ratio	• Latitude • Longitude • Depth • Magnitude
ML Algorithms	• Random forest regressor • Support Vector Machine • Kernel Ridge Regression • Catboost Regressor	• Random forest regressor • Support Vector Machine • Decision Tree
Result Visualisation	• Plotting predicted value vs actual value for all the algorithms. • Comparing the Machine learning models by plotting their output waveforms with the original waveforms.	• Visualising Earthquakes occurred between 26-02-1950 – 26-12-2020. • Comparing the Machine learning models by scattering predicted outputs vs actual values.

Random Forest Regressor
It is an ensemble technique which means it creates multiple models and then combines them to produce improved results.
Here, we are using RF algorithm on both the datasets.
In case of Dataset 1, the CV score we obtained was 0.3524.
In case of Dataset 2, the mean absolute error (MAE) obtained was 2.5059

Decision Tree
It uses a tree-like model of decisions to predict outcomes.
We used this only on Dataset 2, the value of R-squared obtained was 0.1612.

Support Vector Regressor
SVR tries to fit the best line within a threshold value
Here we are using SVR algorithms on both the datasets.
In case of Dataset 1, the CV score we obtained was -2.1722.
In case of Dataset 2, the score we obtained was -15.658.

Kernel Ridge Regression
It combines ridge regression with kernel tricks.
We used this only in Dataset 1, the value of CV score obtained was -2.2046

Results & Discussions

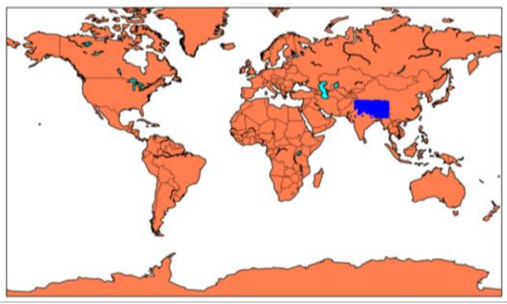

Figure 1

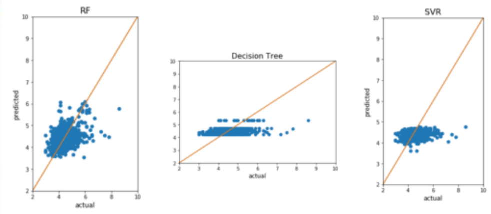

Figure 2

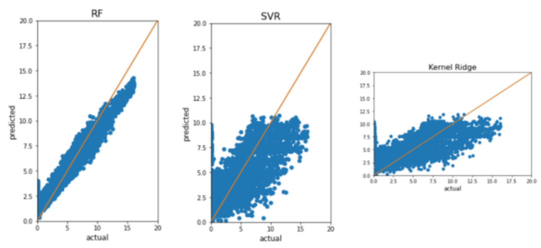

Figure 3

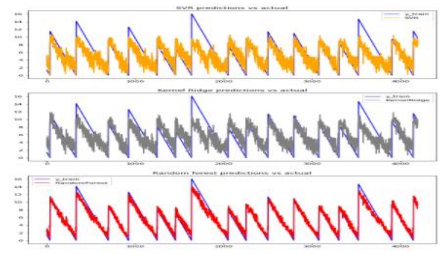

Figure 4

Figure 1: Earthquake affected areas in Himalayan Plateau using Dataset – 2
Figure 2: Result Visualization of ML Algorithms used for Dataset – 2. (RF, SVR & Decision Tree)
Figure 3: Result Visualization of ML Algorithms used for Dataset – 1. (RF, SVR & Kernel Ridge)
Figure 4: SVR, Kernel Ridge and RF prediction vs. actual for Dataset – 1.

- Prediction of Time to Failure, Magnitude and Depth is carried out by the mentioned ML algorithms (RF, SVM, Catboost, Kernel Ridge & Decision Tree.)
- Main metrics for Model evaluation in Regression are R squared, RMSE & MAE.

Performance Analysis of the given ML algorithms on both the Datasets.

ML Algo	MAE		RMSE		R- Squared		Best CV Score	
Dataset	1	2	1	2	1	2	1	2
SVM	2.09421	0.4276	2.76348	0.5612	0.43386	0.0374	-2.1722	-0.4099
RF Algorithm	0.80280	0.4162	1.03247	0.5561	0.92097	0.1374	0.3524	0.1445
Kernel Ridge	2.10980	---	2.71360	---	0.45412	---	-2.2046	---
Catboost	3.21763	---	3.82865	---	0.0866	---	1.3776	---
DT	---	0.4276	---	0.5612	---	0.1215	---	-0.2808

- Dataset 1: Results show that acoustic data from seismic waves successfully predicted the time to failure within acceptable limits. As observed from the table, Random Forest Algorithm was able to predict time to failure with higher accuracy than the other three algorithms mainly because the size of the dataset and complexity of the features involved after Feature Extraction
$$RF > Kernel\ Ridge > SVM > Catboost$$
- Dataset 2: Results show that Latitude and Longitude from the Real-time data successfully predicted the Magnitude of the Earthquake with acceptable limits. As observed from the table, Random forest algorithm, was able to predict Magnitude with higher accuracy than the other two algorithms.
$$RF > SVM > Decision\ Tree$$

References

1. Ajit Kundu, Y. S. Bhaduria, S. Basu, S. Mukhopadhyay, "Artificial neural network-based estimation of moment magnitude with relevance to Earthquake Early Warning", International Conference on Wireless Communications, 2017.
2. Yih-Min Wu, Li Zhao, "Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning", Geophysical Research Letters, August, 2006.
3. Yih-Min Wu, Hiroo Kanamori, Richard M. Allen, Egill Hauksson, "Determination of earthquake early warning parameters, τ c and Pd, for southern California", The Authors Journal Compilation, 2007.
4. Kemal Gunaydin, Ayten Gunaydin, "Peak Ground Acceleration Prediction by Artificial Neural Networks for Northwestern Turkey", 2008.
5. Jasperson H.A, Bolton D.C, Johnson P.A, Marone C, Dehoop M, "Toward laboratory earthquake prediction using unsupervised classification of acoustic emissions", American Geophysical Union, December, 2018.
6. F Corbi, L Sandri, J Bedford, F Funiciello, S Brizzi, M Rossenau, S Lallemand, "Machine Learning Can Predict the Timing and Size of Analog Earthquakes", American Geophysical Union, February, 2019.

Contact Details

hrithiksarda.4@gmail.com shreenibhaseri@gmail.com