

Earthquake prognosis using machine learning

Sachin Sawant^{1*}, Purva Golegaonkar^{2*}, Prayas Gondane³, Rushikesh Gole⁴, Srushti Gole⁵, Aniruddha Gondkar⁶, Aditya Gorave⁷ and Rupali Deshpande⁸

¹⁻⁷Department of Engineering Sciences and Humanities, Vishwakarma Institute of Technology, Pune, Maharashtra, India

Abstract. One of the deadliest and riskiest natural disasters is an earthquake. They often occur without a warning or any further alert. Therefore there was a need for its prognosis as it is extremely important for mankind as well as the environment. In this project, the successful application of machine learning techniques have been used for different elements of research which would be possible to use to make a more accurate short-term prognosis of upcoming earthquakes. Random Forest Classifier is the algorithm used for the research.

Keywords. Earthquake, Random Forest, Machine Learning, Latitude, Longitude.

1. Introduction

As the world is growing in the field of science and technology, there are still a few things that human beings cannot stop. Natural disasters are one of them. The consequences are large number of deaths, property loss, and damage to the environment as well as mankind. Humans cannot avoid them, but early prediction and appropriate precautions can reduce the casualties and the loss of life as well as property. As Earthquake is one of the main such disasters which is very disastrous, Its after effects are very deadly as it results in disasters like landslides, tsunamis, volcano eruptions, etc..

2. Literature Review

At present, a specific technique which could have been used for predicting an earthquake does not exist, which made it much more destructive. For example the 2023 Turkey Earthquake tendency (There were 3 consecutive earthquakes all with magnitudes 5+ which were followed by land slides that resulted in nearly 25,000+ death cases).

In today's era Data Analytics is one of the tool used to predict future outcomes from the current and past available observations. E.g. Data Analytics is used in the medical field to improve the patient's health. Sports, defence, marketing, education are some of the other fields where it is used. In sports it is used in management of sports big data [1]. Data Analytics in the medical field tells us about the use of data analytics to predict the demand of oxygen which is very important for diseases like Covid -19[2]. Its use in defence and finance is

* Corresponding Author: sachin.sawant@vit.edu

* purva.golegaonkar22@vit.edu

helpful [3][4]. A. Moghar presented a paper in which machine learning was used to predict the stocks[5]. Use of machine learning in agriculture to decide which crop should be grown is significant [6].

Earthquake Prediction using Machine Learning has been done with six machine learning algorithms. To choose the best model, various mechanisms are used separately, and accuracy levels in the training and testing datasets are compared[7,8]. Earthquake Damage Prediction using Machine Learning by D. T. Nandwani et.al [9] reported how to reduce the damage caused by earthquakes. Earthquake trend prediction using long short-term memory run predicts the future trend of occurring earthquakes [10,11]. Use of SVR and ANN was done in the earthquake prediction model [12-15]. It has been reported to use Dynamic Fourier Analysis to distinguish between seismic signals from natural earthquakes and mining explosions [16]. Use of GIS technology was done to predict the landslide [17]. ANFIS has been used to predict the earthquake [18]. Use of mobile gravity data to predict the earthquake has been reported[19]. It has been done to predict earthquake-induced landslide susceptibility using real-time earthquake-induced landslide data [20]. Very Recently, research for post earthquake damaged reinforced concrete has been done [21-23]. Recently S.Gentili et.al presented a paper in which innovative machine learning approach called NESTORE is used to predict the probability of earthquake [24].

In the present work, Random Forest Classifier is used to predict the magnitude of an earthquake. Three parameters were considered to predict the earthquake latitude, longitude and depth from the epicentre.

3. Methodology

An algorithm for classification and regression in machine learning called random forest is based on supervised learning. The Random Forest classifier used here is based on the Wisdom of Clouds idea, which claims that many diverse models acting as a committee might perform admirably in each set of separate constituent models. It is a well-known algorithm for machine learning that falls under the umbrella of ensemble learning techniques. Using a randomly selected subset of characteristics and samples from the training data, the method creates several decision trees. The final forecast is based on the consensus of the separate decision trees' predictions, each of which is trained on a distinct subset of the training data. It is a strong method that handles high-dimensional datasets and a large number of features with ease and is less prone to overfitting than single decision trees.

The dataset used in the project is taken from United States Geological Survey [25]. The dataset includes time, latitude, longitude, depth, magnitude, magnitudetype, nst, gap, dmin, rms, place, horizontal Error, dept Error, magError, magNst, location Source, magSource. Parameters considered for the research are latitude, longitude, depth, magnitude. Figure 1 depicts the block diagram of project.

3.1 Block Diagram

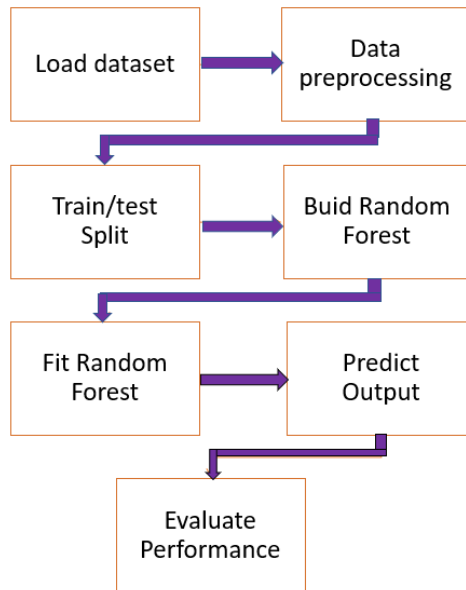


Fig. 1. Flowchart showing flow of the model

4. Result and Discussions

In the given map below in Fig. 2, the earthquakes that have occurred in the past throughout the world are noted. The area which is marked in red tells us about the earthquakes occurred in the past. Similarly in Fig. 4 and Fig. 5 specifically the earthquakes that have occurred in the different continents before which are included in the dataset. A lot of seismic activity has taken place earlier in Europe. Similarly, a lot of earthquakes have been occurred in South America. With the help of these snippets a clear picture about the earthquakes that have occurred before is noted. All of these maps were generated through basemap(matplotlib extension).

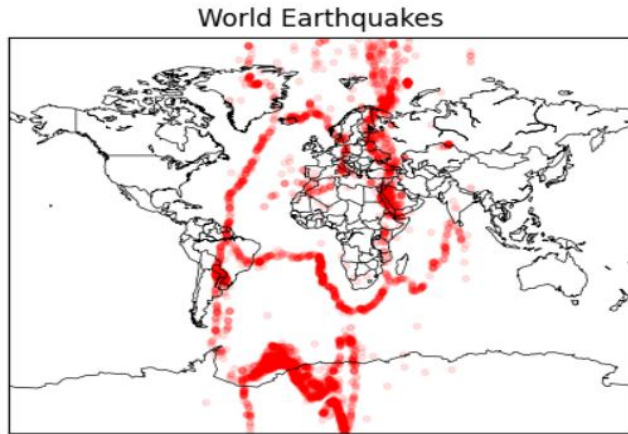


Fig. 2. World Earthquakes

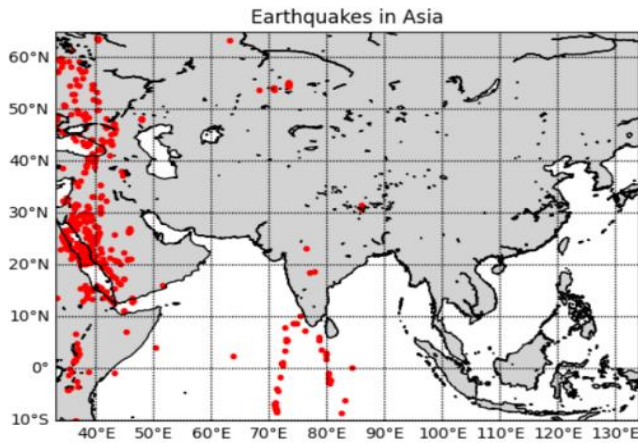


Fig. 3. Past Earthquakes in Asia

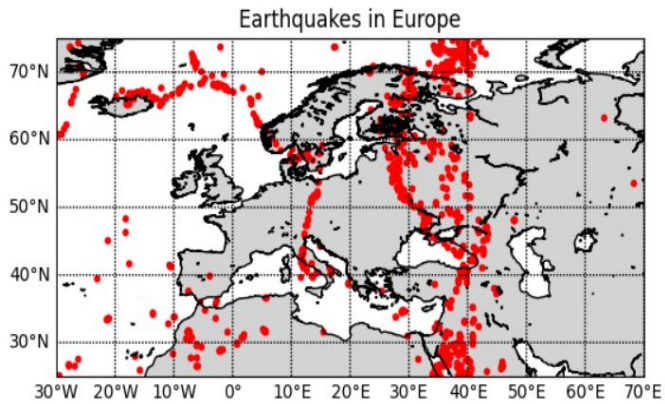


Fig. 4. Past Earthquakes in Europe

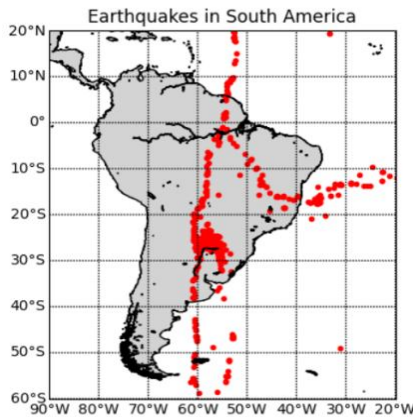


Fig. 5. Past Earthquakes in South America

So, talking about the model, by inputting the values of latitude, longitude and depth, the model predicts the magnitude of the earthquake that will be occurring at that place. Study of the underground formation of fractures, faults would help to decide the depth of the specific place. So the accuracy of the model came to be 62.90%. Besides this F1 score of the model was noted as 0.60 which tells us that the model is working moderately well. Also the values for precision and recall are 0.59 and 0.62 respectively. The precision is 0.59, which meant that out of all the positive predictions made by the model, only 59% were correct. The recall is 0.62, which means that the model was able to correctly identify 62% of the actual positive instances.%. Fig. 8 shows the performance metrics of the model

```
print("ACCURACY :",metrics.accuracy_score(y_test ,y_predict))
```

```
ACCURACY : 0.629060838747785
```

Fig. 6. Accuracy of the model

```
from sklearn.metrics import f1_score, precision_score, recall_score, roc_auc_score
f1= f1_score(y_test, y_predict, average='weighted')
precision = precision_score(y_test, y_predict, average='weighted')
recall = recall_score(y_test, y_predict, average='weighted')

# print the evaluation metrics
print("F1 score: {:.2f}".format(f1))
print("Precision: {:.2f}".format(precision))
print("Recall: {:.2f}".format(recall))
```

```
F1 score: 0.60
Precision: 0.59
Recall: 0.62
```

Fig. 7. Evaluation Metrics

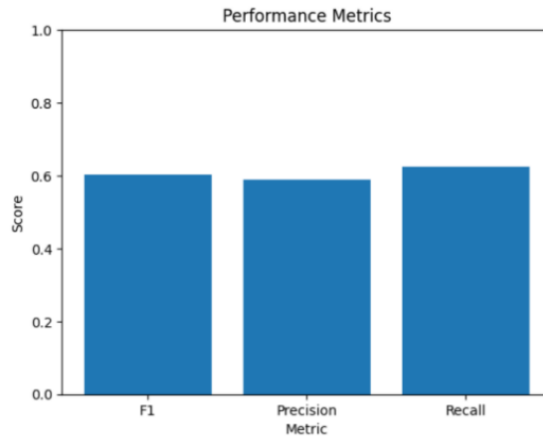


Fig. 8. Performance metrics

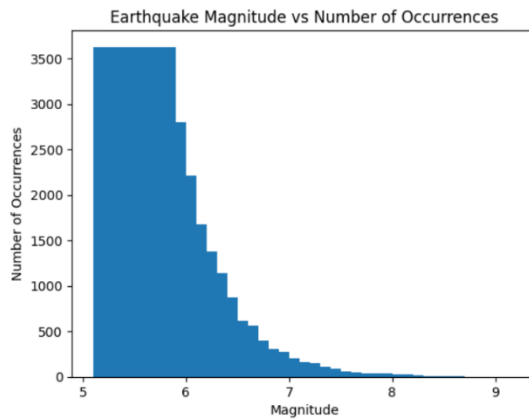


Fig. 9. Magnitude vs No. of occurrences of earthquake of world dataset

As a case study implementation of this model on Turkey and Himalayan Range by considering previous data of Turkey, Himalayan range which was taken from the USGS website. After implementing the model on Turkey we got its accuracy as 64.62. After entering the coordinates and depth of Yesilyurt in Turkey the predicted result came to be in good agreement with the original which is 5 Richter Scale. And the earthquake that occurred there recently on 23/2/2023 was of magnitude 5.4 Richter Scale. In the Fig.11 given below, previous earthquakes that have occurred in Turkey are shown.

```
[5]  
: print("ACCURACY :", metrics.accuracy_score(y_test ,y_predict))  
ACCURACY : 0.6462480857580398
```

Fig. 10. Accuracy of the Turkey model.

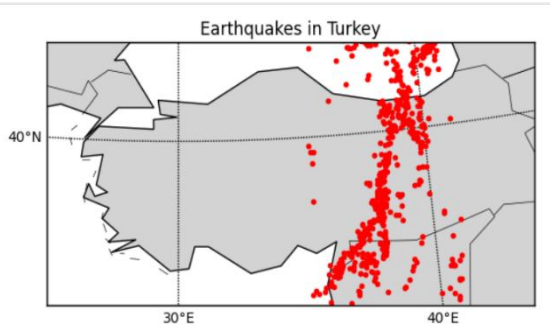


Fig. 11. Past Turkey Earthquakes

Talking about the F1 score, precision, it as shown in the following graphs.

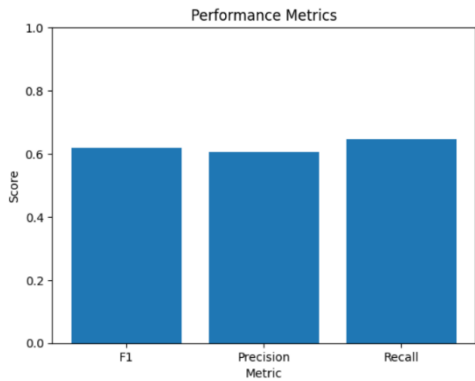


Fig. 12. Performance metrics

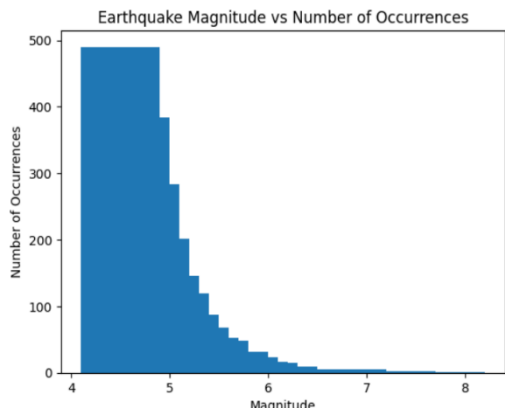


Fig. 13. Magnitude vs No. of occurrences of earthquake of Turkeydataset

The above figures shows the graph depicting the earthquake’s magnitude vs number of occurrences of the earthquakes that have happened before. Majority of the earthquakes occurred between the range of magnitude 4 to 5 Richter Scale.

While implementing the same on Himalayan range,by considering the dataset of previous years its accuracy was noted as 67.82% which is shown in Fig.14. After entering the coordinates and depth of latest earthquake which happened in the Himalayan range in Bhadarwah on 13/06/2023, the magnitude of earthquake was reported as 4 Richter scale

which is again in good agreement with the original earthquake which was of 5 Richter scale. Fig.15 shows the past earthquakes happened in the Himalayan Range. Fig. 16 and Fig. 17 shows the graph depicting F1 score, precision metric recall and magnitude vs No. of occurrences of Himalayan dataset respectively.

```
: print("ACCURACY :", metrics.accuracy_score(y_test ,y_predict))
```

ACCURACY : 0.6782608695652174

Fig. 14. Accuracy of the Himalayan model.

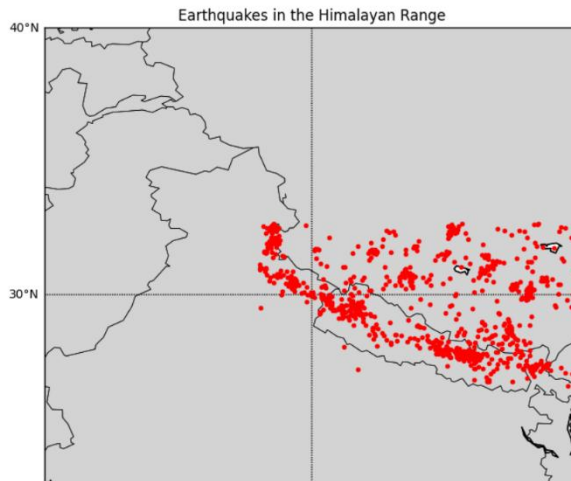


Fig. 15. Past earthquakes in Himalayan Range

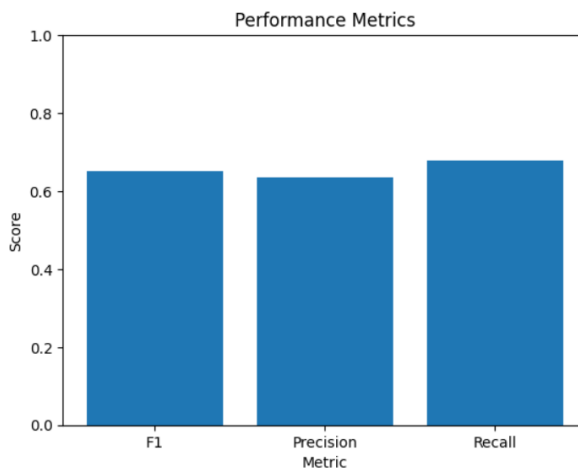


Fig. 16. Performance Metrics

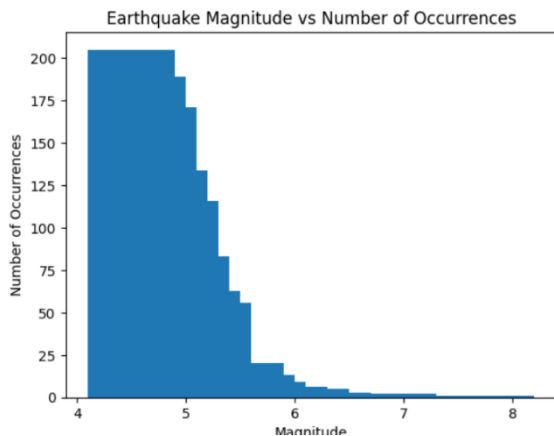


Fig. 17. Earthquake Magnitude vs Number of Occurrences of Himalayan Range dataset

Hence it is understood how important it is to predict the earthquake in today's generation as even today, thousands of people are dying due to this horrible disaster. Further by including more geological factors like fault density, construction dataset, increase in the accuracy of the model can be done.

5. Conclusion

This research paper presented a model for earthquake magnitude prediction using seismic data. The dataset includes information of the latitude, longitude, depth, and magnitude of earthquakes that have occurred worldwide. The model takes in these inputs and other relevant features to predict the earthquake's magnitude. The model's efficiency was tested through a case study of earthquake data from Turkey and Himalayan Range, and it was found to predict the magnitude of earthquakes in the region with a moderate degree of precision and recall. This study highlights the potential of using models to predict the earthquake's magnitude based on the seismic data. The model's predictions can be useful in real-time earthquake monitoring systems to help mitigate the damage caused by earthquakes and save lives.

6. Acknowledgement

We want to thank the honourable Prof. (Dr.) R. M. Jalnekar, Director of the Vishwakarma Institute of Technology (VIT), Pune, and Prof. (Dr.) C. M. Mahajan, HOD, for their encouragement and strong moral support. Additionally, we would like to express our gratitude to Dr. Sachin S. Sawant, our project's mentor, for his invaluable advice.

References

1. Zhongbo Bai, Xiaomei Bai, "Sports Big Data: Management, Analysis, Applications, and Challenges", *Complexity*, Article ID 6676297, 11 pages, **vol. 2021** (2021).
2. S. S. Sawant, C. B. Dongre, A. D. Dolas, Oxygen Demand Forecasting with Data Analytics, in P.Mahalle , R. Talware , G. Patil, S. Sakahre, P. Futane, Y. Dandawate (Eds) book " Artificial Intelligence in Information and Communication Technologies, Healthcare and Education, CRC Press, 1st Edition, Boca Ronton Florida, USA, (2022)

3. Bohdan L. Kaluzny, Data analytics in military human performance: Getting in the game: Summary of a keynote address, *Journal of Science and Medicine in Sport*, **Vol (24)** Issue 10, (2021)
4. Huang, J., Chai, J. & Cho, S. Deep learning in finance and banking: A literature review and classification. *Front. Bus. Res. China* **14**, 13 (2020).
5. A. Moghar, M. Hamiche, Stock Market Prediction Using LSTM Recurrent Neural Network, *Procedia Computer Science*, **vol 170**, (2020)
6. S. S. Sawant, D. Bobby, A. Dusane, G. Durge (in press), Crop Yield Prediction and Leaf Disease Detection Using Machine Learning, *Smart Innovations and Technological Advancements in Civil and Mechanical Engineering*, CRC Press, July (2023)
7. S. A. Kumar, A. Kumar, A. Dhanraj, A. Thakur, "Earthquake Prediction using Machine Learning", *IRJET*, **8** (2021).
8. P. Bangar, D. Gupta, S. Gaikwad, B. Marekar, J. Patil, "Earthquake Prediction using Machine Learning Algorithm", *IJRTE*, **vol 8** (2020).
9. D. T. Nandwani and V. Buradkar, "Earthquake Damage Prediction using Machine Learning" vol(10) 7 July-(2022).
10. T. K. Bhandar, V. N. Satish, S. Sridhar, R. Sivakumar, and S. Ghosh. "Earthquake trend prediction using long short-term memory rnn", *IJECE*, **vol 9** (2019).
11. H. S. Kuyuk, and O. Susumu, "Real time classification of earthquake using deep learning" *Procedia Computer Science* **vol 140** (2018), 298-305.
12. K. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks", *PLOS ONE*, **vol 13** (2018).
13. Jena, R., Pradhan, B. Integrated ANN-cross-validation and AHP-TOPSIS model to improve earthquake risk assessment. *Int. J. Disast. Risk Reduction* **50**, 101723 (2020).
14. K. M. Asim, A. Idris, F. Martínez-Álvarez and T. Iqbal, "Short Term Earthquake Prediction in Hindukush Region Using Tree Based Ensemble Learning," *International Conference on Frontiers of Information Technology (FIT)*, Islamabad, (2012) pp. 365-370
15. A. E. Ruano, G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, P. M. Ferreira "A Support Vector Machine Seismic Detector for Early-Warning Applications", *IFAC Proceedings* **vol 51**, (2013) pp. 400-405.
16. M. C. Mariani, H. Gonzalez, Huizar, M. A. M. Bhuiyan, O. K. Twenwoah. "Using Dynamic Fourier Analysis to Discriminate Between Seismic Signals from Natural Earthquakes and Mining Explosions", *AIMS Geosciences*, **vol 3 (3)**, 438-449 (2019).
17. S Carrara A, Guzzetti F, Cardinali M, Reichenbach P "Use of GIS technology in the prediction and monitoring of landslide hazard" (1999).
18. M. Bhatia, T. A. Ahanger, A. Manocha, Artificial intelligence based real-time earthquake prediction *Engineering Applications of Artificial Intelligence*, **vol 120**, (2023).
19. Y. Zhu, F. Liu, X. You, W. Liang, Y. Zhao, Lian Liu "Earthquake Prediction from China's mobile gravity data, *Geodesy and geodynamics*, **vol (6)**, Issue 2, (2015).
20. A. Pyakurel, B. K. Dahal, D. Gautam, Does machine learning adequately predict earthquake induced landslides?, *Soil Dynamics and Earthquake Engineering*, **vol 171**, (2023).
21. W. Wang, L. Li, Z. Qu, Machine learning-based collapse prediction for post-earthquake damaged RC columns under subsequent earthquakes, *Soil Dynamics and Earthquake Engineering*, **vol 172** (2023).

22. C. Cosgun, Machine learning for the prediction of evaluation of existing reinforced concrete structures performance against earthquakes, *Structures*, **vol 50**, (2023).
23. K. C. Sajan, A. Bhusal, D. Gautam, R. Rupakhety, Earthquake damage and rehabilitation intervention prediction using machine learning, *Engineering Failure Analysis*, **vol 144**, (2023).
24. S. Gentili, R. Di Giovambattista, Forecasting strong subsequent earthquakes in California clusters by machine learning, *Physics of the Earth and Planetary Interiors*, **vol 327**, (2022).
25. <https://earthquake.usgs.gov/earthquakes/search/> dataset link