Journal of Soft Computing Exploration

Analysis of earthquake forecasting using random forest

Kholiq Budiman[1], Yahya Nur Ifriza[2]
[1,2]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:
Received Aug 30, 2021
Revised Sep 6, 2021
Accepted Sep 7, 2021

Keywords:

Analysis
Earthquake
Forecasting
Random Forest
predictions

ABSTRACT

The subject of forecasting earthquakes is an intriguing one to investigate. As a natural calamity, earthquakes continue to be devastating, not just to the economy but also to the lives of individuals. This gave rise to the concept of creating an early warning system against seismic catastrophes to minimize deaths. Researchers have been making earthquake forecasts and seismic hazard ratings of a location for a few years now. In this work, we attempt to forecast earthquakes before they occur using p-arrival data, which includes information on disaster arrival time and amplitude height from the arrival station. Several studies on earthquake prediction have been carried out so far and have developed and used the Random Forest method and one of the Machine Learning. According to, the process of predicting earthquakes has been studied for a long time, but there is still uncertainty due to the diversity and complexity of the earthquake phenomenon itself. According to, conducting a random forest prediction model to identify the structural safety status of buildings damaged by the earthquake is probabilistic. An earthquake's latitude, longitude, magnitude, and depth may be predicted using the random forest algorithm. A random forest with multioutput technique is employed, with variables being each station's recorded value and geographic position. This study's predictions were accurate to within 63 percent.

Corresponding Author:

Kholiq Budiman,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: kholiq.budiman@mail.unnes.ac.id

1. INTRODUCTION

Indonesia is a seismically active territory, having 3,486 earthquakes with a Richter magnitude greater than 6.0 between 1976 and 2006. The Meteorology, Climatology, and Geophysics Agency's (19-year) research has resulted in 27 damaging earthquakes and 13 earthquakes that caused tsunamis [1][2]. Every year, Indonesia is hit by an average of 2 earthquakes and one tsunami [3]. There are several methods to classify earthquakes depending on their causes. Seismologically, tectonic earthquakes are those that occur when elastic energy held in tectonic plates releases. In other words, volcanic earthquakes occur when volcanic activity causes earthquakes to occur [4]. Last but not least, earthquakes can be attributed to human activity as a contributing factor. To give you a few examples, consider building high-dams and injecting enormous amounts of water into rock aquifer or oil wells [5].

Several studies on earthquake prediction have been carried out so far and have developed and used the Random Forest method and one of the Machine Learning. According to [1], the process of predicting earthquakes has been studied for a long time, but there is still uncertainty due to the diversity and complexity of the earthquake phenomenon itself. According to [2], conducting a random forest prediction model to identify the structural safety status of buildings damaged by the earthquake is probabilistic. According to [6], in this study, an evaluation of the performance of the machine learning algorithm "Random Forests" was carried out to classify seismic signals recorded at the Piton de la Fournaise volcano, La Reunion Island (France). Meanwhile, [7], with data on water pipe damage due to the earthquake from February and June 2011 in Christchurch, New Zealand, compares four methods [8][9][10].

The ability to make appropriate choices after a large earthquake is critical for allowing short-term decisions, such as those involving building evacuation or repair. Furthermore, by integrating fast building damage assessment with short-term forecasts of seismicity and related hazards, it is possible to enhance seismic crisis management and safety measures [11]. For short-term forecasting on a city-wide scale or in the case of important buildings, performance levels linked to imminent occupancy or damage grades should be taken into account. In particular, fast methods based on the Markov Chain have been utilized to evaluate seismic risk assessment; however, the evaluation must be extended throughout the full length of the aftershock series in order to account for possible damage accumulations resulting from the aftershock sequence. After then, operational methods are needed to integrate aftershock occurrence predictions, damage accumulation models, and building health characterizations in order to achieve the desired levels of damage. It is common practice to make earthquake risk predictions during an aftershock series, which is based on operational earthquake forecasting (OEF). Hermann et al. [12] utilized the OEF to predict the time-varying seismic risk during an earthquake series scenario mimicking the 1356 Basel earthquake, which they found to be accurate. They calculated the number of lives lost and the hazards associated with them in the near term in order to support choices such as evacuation or the suspension of essential activities in reinforced concrete (RC) structures. Aftershock sequences in Italy were analyzed retrospectively by Chioccarelli and Iervolino [4], who utilized OEF to conduct their study of loss estimates [13][14][15][16].

In seismic crises, decision-makers must take into account the increased susceptibility of structures as a result of mainshock damage, as well as the time-variant vulnerability of buildings as a result of the possibility of aftershock damage accumulation. Aftershock damage accumulation has been quantified using probabilistic models in order to determine the relative contribution of aftershock damage accumulation to the damage produced by the mainshock [17][18]. Evaluation of seismic performance and development of damage accumulation models, in conjunction with a probabilistic assessment of aftershock occurrence, have been carried out using Markov Chain-based methods, as has been done for aftershock occurrence. The majority of research find that aftershocks provide a significant contribution to prediction of consequences and losses. The short-term variability of a building's vulnerabilities, as measured in relation to the damage accrued over the course of the whole seismic sequence, is a critical factor to consider when making short-term decisions [18][11].

When buildings' stiffness progressively deteriorate before collapse, monitoring the elongation of their fundamental period may aid in the assessment of seismic damage to such structures [19]. When considering apparent structural stiffness and structural health, the fundamental period (or frequency) is considered to be a surrogate [20][17][21]. It has been possible, for example, to investigate the impact of seismic damage accumulation on a macro-seismic intensity estimate by using the residual stiffness of masonry structures derived from period measurements [22][23]. We conducted laboratory experiments on unreinforced masonry specimens in order to measure the fundamental frequency shift as a function of structural drift and the degree of damage. Researchers have discovered empirical connections between the frequency shift and the damage index for RC structures using experimental and computational methods [24][25][26]. In spite of this, building tagging processes often give a red tag to structures that have suffered significant damage without collapsing, and to structures that are deemed dangerous, unrepairable, or unworthy of repair. In this regard, the methods described below are concerned with the probability of damage states ranging from minor to severe, depending on the length of time that has passed since the collapse was seen [27][28][29].

Furthermore, [30][31][32] used Decision Tree Bagging and Random Forest in the field of earthquake precursors to detect GPS-TEC (Total Electron Content) seismoionospheric anomalies around the time and location of the Chile earthquake 27 February 2010. In [33], built an earthquake alert system in 9 countries to minimize the earthquake's impact. Furthermore, [34], to predict the earthquake's strength in California using the big data algorithm. And [35][36] applied the Artificial Neural Network Cross-Validation and AHP-

TOPSIS methods to calculate the risk of an earthquake in Aceh, Indonesia. Based on previous research, it can be seen that Random Forest can be used to predict earthquake disasters.

2. METHOD

Predicting the depth, strength/magnitude, and earthquake location (longitude and latitude) of an earthquake using datasets taken from three monitoring stations. By using Random Forest to get accurate results and using google collab as software that helps process data. Steps to make predictions using google collab:

First, enter the dataset into Google Drive, then connect Google Drive with Google Collab to use the dataset. Second, by using a random forest regressor and doing a split test. Third, determine the variables x and y, where x is the data from the earthquake monitoring station, and y is the desired prediction result. Fourth, separate train and test data with 289 train data sizes and 30 tests. Fifth, by using a random forest regressor, predictions of the value will be made using the x test variable. Sixth, compare the results between data, multi-random forest, and random forest in the form of plots. Seventh, the prediction results were obtained from the earthquake's depth, strength, and location with a sure accuracy. Eighth, actions can be taken to see the results more clearly.

In addition to using google collab, here are the four steps in the Random Forest algorithm [37]:

1. Choose a random bootstrap sample of size n (randomly taken n samples from the training set with replacement)
2. Build a decision tree from the bootstrap sample. On each node:
   a. Randomly select d features without replacement
   b. Split the nodes using the feature that provides the best split according to the objective function.
3. Repeat steps 1) and 2) k
4. Combine predictions based on each tree to assign a class label based on the most votes.

Furthermore, by using the Random Forest algorithm with the Scikit-learn library, the following equation is obtained [38]

1) The number of samples used is as many as the original data lines

$$n = n_{baris\ data\ asli} \tag{1}$$

2) The following equation determines the value of d

$$d_{fitur} = \sqrt{jumlah\ fitur} \tag{2}$$

3) Decision Tree is an algorithm Cart Tree (Classification and Regression Trees)
   a) Split data

$$split\ \theta = (j, t_m) \tag{3}$$

$$Q_{left}(\theta) = (x, y) | x_j \leq t_m \tag{4}$$

$$Q_{right}(\theta) = (x, y) | x_j > t_m \tag{5}$$

Information:

$split\ \theta$: split data
j : feature
t_m : threshold value specified at node m
$Q_{left}(\theta)$: data partition on the left child where the x value in feature j is less than or equal to the threshold value at node m
$Q_{right}(\theta)$: data partition on the right child where the x value in feature j is more than the threshold value at node m
x : training data
y : class label

   b) Impurity value of child node m

$$G_{child}(Q, \theta) = \frac{n_{left}}{N_m} H\left(Q_{left}(\theta)\right) + \frac{n_{right}}{N_m} H\left(Q_{right}(\theta)\right) \tag{6}$$

$$H(X_m) = \sum_k P_{mk}(1 - P_{mk}) \tag{7}$$

$$P_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = 1) \tag{8}$$

Information:

$G_{child}(Q,\theta)$: information gain child or impurity node m
n_{left} : the number of data partitions on the left child
n_{right} : the number of data partitions on the right child
N_m : the amount of data on node m
$H(X_m)$: function to find the impurity value
k : label class value
P_{mk} : the proportion of class label k on node m
I : a lot of data that the value of the label classy is the same as the class k
i : row region on node m $[0,1,2,\dots,N_m-1]$

Split data is done by selecting feature parameters and threshold node m, producing the smallest Information Gain Child value.

$$\theta^* = argmin_\theta G_{child}(Q,\theta) \quad (9)$$

Information:

θ^* : the smallest information gain value of child node m

c) Information Gain

$$IG(Q,\theta) = \frac{N_m}{N}(H(Q(\theta)) - G_{child}(Q,\theta)) \quad (10)$$

Information:

$IG(Q,\theta)$: information gain node m
N : the amount of data
$H(Q(\theta))$: impurity parent pada node m
$G_{child}(Q,\theta)$: information gain child pada node m

4) The Feature Importance function can be calculated from the average impurity reduction of all decision trees in a Random Forest without assuming whether the data used are linearly separated or not [12]. Find the value of feature importance can be seen in the following equation [39]:
   a) Menghitung nilai importance pada setiap decision tree:

$$FI_i = \frac{\sum_j IG_j}{\sum_k IG_k} \quad (11)$$

Information:

FI_i : importance value for feature I in the decision tree
IG_j : information gain value of the essential feature at node j
j : node in a decision tree
i : feature index to i
k : all nodes in the decision tree

   b) Calculating the value of feature importance in a random forest:

$$RFFI_i = \frac{\sum_j FI_{ij}}{T} \quad (12)$$

Information:

$RFFI_i$: importance value for feature I in the decision tree
FI_{ij} : a most crucial feature I in decision tree j
j : decision tree pada random forest
T : the number of decision trees in a random forest

5) Confusion Matrix is a matrix that describes the performance of an algorithm [12]. From the Confusion Matrix, the values of false-negative (TFN), false positive (TFP), and true negative (TTN) can be obtained with the equation as [18]

$$TFN_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} x_{ij} \quad (13)$$

$$TFP_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} x_{ji} \quad (14)$$

$$TTN_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} x_{jk} \quad (15)$$

$$TTP_{all} = \sum_{j=1}^{n} x_{ji} \quad (16)$$

From the above equation, the accuracy value is determined to determine the performance of the algorithm. Accuracy is the overall average performance [40], following the form of the equation

$$accuracy = \frac{TTP_{all}}{Total\ number\ of\ testing\ entries} \quad (17)$$

3. RESULTS AND DISCUSSIONS

1) Using Google Collab

From table 1, it can be seen that the prediction and data values are almost the same, although they are still inaccurate because the accuracy value is still 63%. Furthermore, it can be seen more clearly in the table to compare the importance of the data and predictions of random forests.

Table 1. the prediction and data values

	Long	Lat	Depth(km)	Mag	pred_long	pred_lat	pred_depth	pred_mag
0	113.87	-10.77	81.47	5.0	113.770315	-10.760219	49.227277	4.724745
1	113.86	-10.68	78.84	3.9	113.828766	-10.790116	61.577034	4.066505
2	114.01	-9.43	106.01	3.9	113.842565	-9.645616	48.447260	4.093076
3	112.74	-9.35	7.44	3.6	112.787658	-9.071596	64.325799	3.360742
4	112.49	-8.97	31.07	3.2	112.543365	-8.730904	92.877491	3.427932
5	114.01	-9.39	13.13	3.6	114.010511	-9.397245	36.133379	3.684806
6	111.72	-8.58	106.58	3.8	111.744921	-8.468248	110.943944	3.675805
7	113.44	-8.61	107.24	4.1	113.402667	-9.258137	51.144672	4.374296
8	114.18	-9.81	14.65	4.1	114.163239	-9.438448	42.947346	4.058913
9	113.41	-9.15	61.01	3.6	113.439193	-9.189722	49.957257	3.566677
10	112.84	-9.42	110.06	3.3	112.850624	-9.222438	95.281329	3.323480
11	113.84	-10.71	56.44	4.7	113.844871	-10.747330	54.516259	4.030911
12	113.89	-10.69	23.12	4.6	113.844369	-10.554225	49.385308	4.143208
13	113.22	-9.28	71.52	4.0	113.208710	-9.210613	55.020431	4.197230
14	113.93	-10.59	78.56	3.6	113.840103	-9.907759	74.764856	3.526163

The above method obtains the comparison results between the data, multi-random forest, and random forest. The comparison of the values is illustrated in latitude and longitude coordinates. With squares as data, circles as multi-random forest values , and triangles as random forest values, figure 1.

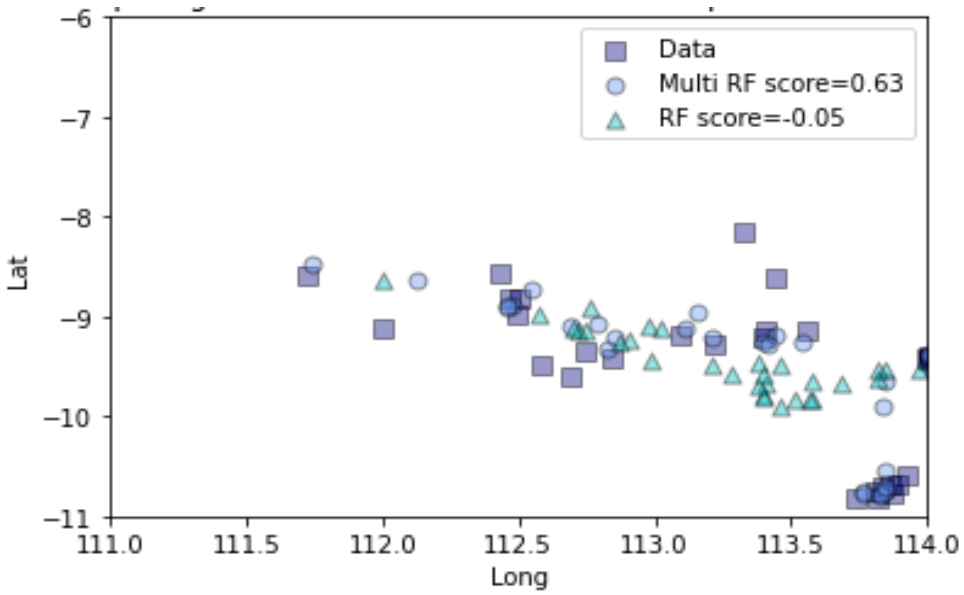

Figure 1. Comparing random forests and the multi-output meta estimator

It can be seen more clearly that the prediction results are close to the data, so it can be said that the prediction results are comparable to accurate. It's just that some predictions are wrong, as in the magnitude at point 11; according to the data, the result is 4.7 SR while the forecast shows 4.03. To overcome this problem, further iterations are needed to obtain the most fantastic accuracy of results.

2) Manual calculation

Using (1) to determine how many samples will be used in this study and obtained n = 386. Next, look for the value of d using (2) and get the value d=√16=4. This means that four features will be used. Then split the data using (3), where j is the magnitude and tm is used 4.0. Next, do a split with the conditions (4) and (5) so that you get two groups, namely left ($x_j \leq 4$) and right ($x_j > 4$), as follows in table 2:

Table 2. split with the conditions

Class	N Left	N Right
0	16	43
1	16	1
2	7	2
4	3	1
5	19	13
7	5	0
8	0	2
9	17	2
10	0	3
12	14	10
13	1	6
16	30	18
17	4	0
18	25	16
19	1	1
20	4	3
21	4	4

22	24	8
24	1	0
25	3	0
26	2	0
27	1	3
28	2	1
29	15	3
30	23	6
31	2	1
Total	239	147

Next, the Information Gain Child value of the magnitude feature will be searched using (6). First, look for the value of left P_mk and right P_mk using (8) and then look for the value of H(Q_left (θ)) and H(Q_right (θ)) using (7). By doing the calculations, the values obtained are H(Q_left (θ))=0.9227 and H(Q_right (θ))=0.869. This followed by finding the value of Gchild(Q,θ) with the following formula:

$$G_{child}(Q,\theta) = \frac{n_{left}}{N} H\left(Q_{left}(\theta)\right) + \frac{n_{right}}{N} H\left(Q_{right}(\theta)\right)$$
$$= \frac{239}{386}(0{,}9227) + \frac{147}{386}(0{,}869) = 0{,}9022$$

Based on these calculations, it is known that the child's information gain for the magnitude feature with a threshold of 4 is 0.9022. Next, look for the attribute with the minimum child information gain value. For the example of manual calculations, this value will be selected to find the information gain value. Find the value of information by using (10), where the importance of H(Q(θ)) and P_mk is the total number in each class. After calculating the value of H(Q(θ))=0.9164, which will be used to find the value of IG(Q,θ) with the following formula:

$$IG(Q,\theta) = \frac{N_m}{N}\left(H\left(Q(\theta)\right) - G_{child}(Q,\theta)\right)$$
$$= \frac{386}{386}(0{,}9164 - 0{,}9022) = 0{,}0142$$

Next, the essential features will be searched using the IG(Q,θ) value of the magnitude feature only, so that by using (11), the value of FI_i=1 will be obtained. Furthermore, the value of the essential components of the random forest was searched by using (12) the value of RFFI_i = 0.25. The importance value of feature magnitude is 0.25.

Then look for the accuracy value using the original data and the prediction results with the value TTP_all=243. TTP_all is the number of actual data entries equal to the predicted data. Then by using (17), the following accuracy is obtained.

$$accuracy = \frac{TTP_{all}}{total\ entries\ tested} = \frac{243}{386} = 62{,}95\%$$

4. CONCLUSION

From the two methods, it can be seen that the accuracy of the two calculations is relatively similar, which is close to 63%. However, the accuracy of manual calculations can be improved by finding the minimum child information gain value from other features such as deep, latitude, and longitude. As for the measures using Google Collab, it can be iterated again to get the highest accuracy.

REFERENCES

[1] T. H. Jordan *et al.*, "Operational earthquake forecasting: State of knowledge and guidelines for utilization," *Ann. Geophys.*, vol. 54, no. 4, pp. 319–391, 2011, doi: 10.4401/ag-5350.

[2] A. Barkat *et al.*, "Time series analysis of soil radon in Northern Pakistan: Implications for earthquake forecasting," *Appl. Geochemistry*, vol. 97, pp. 197–208, 2018, doi: 10.1016/j.apgeochem.2018.08.016.

[3] T. Hardy, B. Nurdiyanto, D. Ngadmanto, P. Susilanto, and B. Sunardi, "Kajian Kerawanan Gempabumi Berbasis Sig Dalam Upaya Mitigasi Bencana Studi Kasus Kabupaten Dan Kota Sukabumi," 2012, [Online]. Available: https://publikasiilmiah.ums.ac.id/handle/11617/1422%0Ahttps://publikasiilmiah.ums.ac.id/bitstream/handle/11617/1422/5-SNPJ-SIG-2012-Bambang Sunardi.pdf?sequence=1&isAllowed=y.

[4] G. Asencio-Cortés, A. Morales-Esteban, X. Shang, and F. Martínez-Álvarez, "Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure," *Comput. Geosci.*, vol. 115, pp. 198–210, 2018, doi: 10.1016/j.cageo.2017.10.011.

[5] Y. Ogata, "A prospect of earthquake prediction research," *Stat. Sci.*, vol. 28, no. 4, pp. 521–541, 2013, doi: 10.1214/13-STS439.

[6] G. Cremen and C. Galasso, "Earthquake early warning: Recent advances and perspectives," *Earth-Science Rev.*, vol. 205, no. February, p. 103184, 2020, doi: 10.1016/j.earscirev.2020.103184.

[7] E. E. H. Doyle *et al.*, "Interpretations of aftershock advice and probabilities after the 2013 Cook Strait earthquakes, Aotearoa New Zealand," *Int. J. Disaster Risk Reduct.*, vol. 49, p. 101653, 2020, doi: 10.1016/j.ijdrr.2020.101653.

[8] M. C. Mariani, M. A. M. Bhuiyan, O. K. Tweneboah, and H. Gonzalez-Huizar, "Long memory effects and forecasting of earthquake and volcano seismic data," *Phys. A Stat. Mech. its Appl.*, vol. 559, p. 125049, 2020, doi: 10.1016/j.physa.2020.125049.

[9] K. Trevlopoulos, P. Guéguen, A. Helmstetter, and F. Cotton, "Earthquake risk in reinforced concrete buildings during aftershock sequences based on period elongation and operational earthquake forecasting," *Struct. Saf.*, vol. 84, no. February 2019, p. 101922, 2020, doi: 10.1016/j.strusafe.2020.101922.

[10] H. Yepes, J. M. Nocquet, B. Bernard, P. B. Palacios, S. Vaca, and S. Aguaiza, "Comments on the paper 'Two independent real-time precursors of the 7.8 M earthquake in Ecuador based on radioactive and geodetic processes – Powerful tools for an early warning system' by Toulkeridis et al. (2019)," *J. Geodyn.*, vol. 133, p. 101648, 2020, doi: 10.1016/j.jog.2019.101648.

[11] D. Bindi, I. Iervolino, and S. Parolai, "On-site structure-specific real-time risk assessment: perspectives from the REAKT project," *Bull. Earthq. Eng.*, vol. 14, no. 9, pp. 2471–2493, 2016, doi: 10.1007/s10518-016-9889-4.

[12] M. Herrmann, J. D. Zechar, and S. Wiemer, "Communicating time-varying seismic risk during an earthquake sequence," *Seismol. Res. Lett.*, vol. 87, no. 2A, pp. 301–312, 2016, doi: 10.1785/0220150168.

[13] K. Budiman and I. Akhlis, "Changing user needs and motivation to visit a website through ad experience: A case study of a university website," *J. Phys. Conf. Ser.*, vol. 1918, no. 4, 2021, doi: 10.1088/1742-6596/1918/4/042008.

[14] K. Budiman, S. Subhan, and D. A. Efrilianda, "Business Process re-engineering to support the sustainability of the construction industry and sales commodities in large scale transaction during Covid 19 with integrating ERP and Quotation System," *Sci. J. Informatics*, vol. 8, no. 1, pp. 84–91, 2021, doi: 10.15294/sji.v8i1.27969.

[15] Sugianto, Z. Abidin, A. T. Putra, and K. Budiman, "Knowledge management system in a higher education institution: Development of an expertise search system," *J. Phys. Conf. Ser.*, vol. 1918, no. 4, 2021, doi: 10.1088/1742-6596/1918/4/042021.

[16] K. Budiman, A. T. Putra, Alamsyah, E. Sugiharti, M. A. Muslim, and R. Arifudin, "Implementation of ERP system functionalities for data acquisition based on API at the study program of Universities," *J. Phys. Conf. Ser.*, vol. 1918, no. 4, 2021, doi: 10.1088/1742-6596/1918/4/042151.

[17] J. W. Baker, "Measuring bias in structural response caused by ground motion scaling," *Pacific Conf. Earthq. Eng.*, no. 056, pp. 1–6, 2007, doi: 10.1002/eqe.

[18] Y. Zhang, H. V. Burton, H. Sun, and M. Shokrabadi, "A machine learning framework for assessing post-earthquake structural safety," *Struct. Saf.*, vol. 72, pp. 1–16, 2018, doi:

10.1016/j.strusafe.2017.12.001.
[19] Y. Reuland, P. Lestuzzi, and I. F. C. Smith, "A model-based data-interpretation framework for post-earthquake building assessment with scarce measurement data," *Soil Dyn. Earthq. Eng.*, vol. 116, no. August 2018, pp. 253–263, 2019, doi: 10.1016/j.soildyn.2018.10.008.
[20] S. Grimaz and P. Malisan, "How could cumulative damage affect the macroseismic assessment?," *Bull. Earthq. Eng.*, vol. 15, no. 6, pp. 2465–2481, 2017, doi: 10.1007/s10518-016-0016-3.
[21] M. Brun, P. Labbe, D. Bertrand, and A. Courtois, "Pseudo-dynamic tests on low-rise shear walls and simplified model based on the structural frequency drift," *Eng. Struct.*, vol. 33, no. 3, pp. 796–812, 2011, doi: 10.1016/j.engstruct.2010.12.003.
[22] P. Probst and A. L. Boulesteix, "To tune or not to tune the number of trees in random forest," *J. Mach. Learn. Res.*, vol. 18, pp. 1–8, 2018.
[23] P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 3, pp. 1–15, 2019, doi: 10.1002/widm.1301.
[24] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005, doi: 10.1080/01431160412331269698.
[25] Y. Qi, "Ensemble Machine Learning," *Ensemble Mach. Learn.*, pp. 307–323, 2012, doi: 10.1007/978-1-4419-9326-7.
[26] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003, doi: 10.1021/ci034160g.
[27] Y. N. Ifriza, C. E. Edi, and J. E. Suseno, "expert system irrigation management of agricultural reservoir system using analytical hierarchy process (AHP) and forward chaining method," pp. 74–83, 2017.
[28] M. Sam and Y. N. Ifriza, "A combination of TDM and KSAM to determine initial feasible solution of transportation problems," *J. Soft Comput. Explor.*, vol. 2, no. 1, pp. 17–24, 2021, doi: 10.52465/joscex.v2i1.16.
[29] Y. N. Ifriza and M. Sam, "Irrigation management of agricultural reservoir with correlation feature selection based binary particle swarm optimization," *J. Soft Comput. Explor.*, vol. 2, no. 1, pp. 40–45, 2021, doi: 10.52465/joscex.v2i1.23.
[30] S. Gentili and R. Di Giovambattista, "Forecasting strong aftershocks in earthquake clusters from northeastern Italy and western Slovenia," *Phys. Earth Planet. Inter.*, vol. 303, no. September 2019, p. 106483, 2020, doi: 10.1016/j.pepi.2020.106483.
[31] A. Zarola and A. Sil, "Forecasting of future earthquakes in the northeast region of India considering energy released concept," *Comput. Geosci.*, vol. 113, pp. 1–13, 2018, doi: 10.1016/j.cageo.2018.01.003.
[32] Y. Zhu, F. Liu, G. Zhang, and Y. Xu, "Development and prospect of mobile gravity monitoring and earthquake forecasting in recent ten years in China," *Geod. Geodyn.*, vol. 10, no. 6, pp. 485–491, 2019, doi: 10.1016/j.geog.2019.05.006.
[33] Q. Hong, S. Crampin, and Y. Gao, "Changes in shear-wave splitting at the 2014 Bárðarbunga seismic crisis and dyke intrusion in Iceland compared with earthquakes and other eruptions," *Phys. Earth Planet. Inter.*, vol. 300, no. February, p. 106446, 2020, doi: 10.1016/j.pepi.2020.106446.
[34] M. Akhoondzadeh, "Decision Tree, Bagging and Random Forest methods detect TEC seismo-ionospheric anomalies around the time of the Chile, (M w = 8.8) earthquake of 27 February 2010," *Adv. Sp. Res.*, vol. 57, no. 12, pp. 2464–2469, 2016, doi: 10.1016/j.asr.2016.03.035.
[35] M. Battarra, B. Balcik, and H. Xu, "Disaster preparedness using risk-assessment methods from earthquake engineering," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 423–435, 2018, doi: 10.1016/j.ejor.2018.02.014.
[36] R. Jena and B. Pradhan, "Integrated ANN-cross-validation and AHP-TOPSIS model to improve earthquake risk assessment," *Int. J. Disaster Risk Reduct.*, vol. 50, p. 101723, 2020, doi: 10.1016/j.ijdrr.2020.101723.
[37] K. Nikolopoulos, F. Petropoulos, V. S. Rodrigues, S. Pettit, and A. Beresford, "A disaster response model driven by spatial–temporal forecasts," *Int. J. Forecast.*, no. xxxx, 2020, doi: 10.1016/j.ijforecast.2020.01.002.
[38] C. Hibert, F. Provost, J. P. Malet, A. Maggi, A. Stumpf, and V. Ferrazzini, "Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random

Forest algorithm," *J. Volcanol. Geotherm. Res.*, vol. 340, pp. 130–142, 2017, doi: 10.1016/j.jvolgeores.2017.04.015.

[39] A. Bagriacik, R. A. Davidson, M. W. Hughes, B. A. Bradley, and M. Cubrinovski, "Comparison of statistical and machine learning approaches to modeling earthquake damage to water pipelines," *Soil Dyn. Earthq. Eng.*, vol. 112, no. June 2017, pp. 76–88, 2018, doi: 10.1016/j.soildyn.2018.05.010.

[40] P. Jiang, X. Liu, and M. Zheng, "Emergency blood demand forecasting after earthquakes," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 773–777, 2019, doi: 10.1016/j.ifacol.2019.11.209.