IJAIT (International Journal
of Applied Information Technology)

http://journals.telkomuniversity.ac.id/ijait/

Early Estimation of Earthquake Magnitude Using Machine Learning

Astri Novianty [a, *], Anggunmeka Luhur Prasasti [a], Randy Erfa Saputra [a]

[a] *Computer Engineering Department, Telkom University, Indonesia*
astrinov@telkomuniversity.ac.id, anggunmeka@telkomuniversity.ac.id, resaputra@telkomuniversity.ac.id

ARTICLE INFO

Received May 10th, 2023
Revised June 5th, 2023
Accepted June 13th, 2023
Available online December 14th, 2023

Keywords
Early estimation, magnitude, earthquake, machine learning, regression

ABSTRACT

Seismic parameters provide important information that describes the characteristics of an earthquake. The magnitude parameter is one of the essential seismic parameters in making the right decision regarding earthquake disaster mitigation. Determining the magnitude of an earthquake must be done early because this information represents the size of the earthquake and the potential damage it causes. If the determination of the earthquake's magnitude is delayed, emergencies such as the evacuation of residents and post-disaster recovery may be disrupted. This study attempts to estimate the earthquake magnitude parameters based on Primary (P) wave signals using several machine learning algorithms for regression, such as Neural Network Regression (NNR), Random Forest Regression (RFR), and Support Vector Machine Regression (SVMR). The experimental results show that the RFR can produce the best estimation with an R-squared (R^2) value of 0.946 and a root mean square error (RMSE) of 0.087.

* Corresponding author at:
Computer Engineering Department, Telkom University
Jl. Telekomunikasi No. 1, Terusan Buah Batu, Bandung, 40257
Indonesia
E-mail address: astrinov@telkomuniversity.ac.id

ORCID ID:
- First Author: 0000-0002-8109-5544
- Second Author: 0000-0001-6197-4157

https://doi.org/10.25124/ijait.v7i02.5994

1. Introduction

The characteristics of an earthquake can be identified through various seismic parameters calculated after the earthquake occurs. These parameters include the location of the epicenter of the earthquake, the origin time, the magnitude, the depth of the earthquake, and so on. Earthquake magnitude calculation is used to estimate the energy released by earthquakes. It also provides essential information for scientists in studying seismology and earthquake behavior. Information on the earthquake’s magnitude can provide initial clues about the earthquake's strength. Earthquakes with higher magnitudes have greater energy and are more likely to cause further disasters such as tsunamis. Therefore, calculating the estimated magnitude of an earthquake needs to be done early and quickly so that decisions related to disaster mitigation can be made earlier.

In general, calculating various seismic parameters can be done using seismic signals or Global Navigation Satellite System (GNSS) data. Today the seismic signal is the primary raw data used in calculating seismic parameters, including the magnitude parameter.

Seismic wave signals are categorized into several types of waves based on the arrival time at the seismometer as the receiver, namely Primary (P) waves, Secondary (S) waves, and surface waves [1][2]. P-waves are the first seismic wave signals that arrive at the receiver after an earthquake occurs. S-waves are seismic wave signals that come after the P signal. Surface waves are seismic wave signals propagating on the earth's surface that arrive at the receiver after the P and S signals. Most of the seismic parameters are calculated using S signals. However, this study seeks to utilize the P signal for the early estimation of the magnitude so that the information obtained can be more beneficial for disaster mitigation.

Many studies have been carried out to estimate the magnitude quickly either based on seismometer signals [3]-[7] or based on GNSS data [8][9][10]. Apart from [5] and [8], the studies use signal processing and specific formulas to estimate the magnitude value, requiring a relatively high computational cost. Machine learning is expected to be an alternative solution for the heavy computation problem.

The study [5] in 2018 used a machine learning approach, namely the Support Vector Machine Regression (SVMR) method. The study focuses on earthquake events with a magnitude less than 5.0 using 12 features of the P signal. Earthquakes greater than 6.5 are not covered yet in the study. Another study in 2021 [8] used Genetic Algorithm (GA) to estimate the magnitude of earthquakes based on GNSS time series data. The method proposed in the study still requires many predetermined assumptions to estimate accurately. So it still needs further improvement.

In [5], SVMR gives a good accuracy of the magnitude estimation. However, many studies show various performances of SVMR by comparing the performance of SVMR with Neural Network Regression (NNR) [11][12], SVMR with Random Forest Regression (RFR) and NNR [13], and others [14]. In [11], SVMR performs better than NNR even though both performances are relatively poor. NNR performs well in [15] and is even better than SVMR in [12]. In another study [13], RFR outperformed SVMR and NNR in performance. In [14], RFR is better than SVMR and in [16], RFR is better than NNR. RFR performance is also excellent in [17].

This study aims to estimate the earthquake’s magnitude from the seismic signals using machine learning-based regression techniques. Due to all performance variations shown in previous studies, as mentioned in the last paragraph, this study tested three classic machine learning algorithms widely used, namely NNR, SVMR, and RFR, to determine their performance in estimating the earthquake’s magnitude. The performance of each method is measured to obtain the best estimation result.

2. Material and Methods

There are at least two essential requirements that must be fulfilled by the proposed system. First, the estimation must be calculated early and immediately after the earthquake. Second, the estimation results must be close to the actual magnitude value. The seismic agency usually officially announces the actual magnitude value several hours after the earthquake.

2.1. Proposed Scheme of Magnitude's Early Estimation

The estimation scheme proposed in this study can be seen in Figure 1. The input data for this estimation system is the P-wave of the seismic signal. The P signals are used as input because they are the earliest part of the earthquake waves received by the seismometer. The signals are then processed in the feature extraction to generate relevant features to the magnitude estimation. The features obtained then become the input for machine learning to estimate the magnitude value using the regression technique.

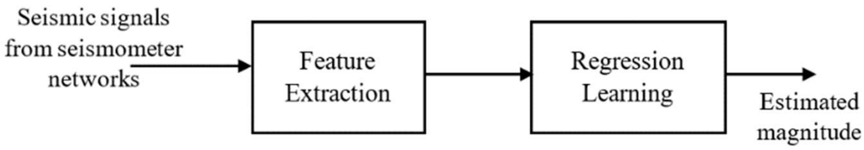

Figure 1 The Scheme of The Proposed Early Estimation System

2.2. Feature Extraction

The features used for magnitude estimation are extracted from the P-wave signal. The signal is used to calculate the magnitude estimation as early as possible, considering that the P-wave signal is the earliest signal to arrive at the seismometer device after an earthquake occurs. The estimation proposed in this study uses three P-wave signals as input acquired from three different seismometer stations.

In this study, the feature extraction generates two features from each P-wave signal, i.e., rupture duration (T_0) and P-wave dominant period (T_d). Hence, the feature vector generated from the feature extraction is stated in Equation 1 as follows.

$$f = [T_{0_1}, T_{d_1}, T_{0_2}, T_{d_2}, T_{0_3}, T_{d_3}] \qquad \text{Equation 1}$$

with T_{0_i} and $T_{\mathrm{d_i}}$ are the rupture duration and P-wave dominant period, respectively, generated from the i-th seismometer. The calculation process including the formula of each feature is described in the following subsections.

2.2.1. Rupture Duration (T_0)

It is the amount of time that an area of fault needs to rupture when an earthquake occurs. The following procedure is performed to obtain this feature [18][19]. The seismic signal is filtered first using a high-pass filter (HPF) to obtain the P-wave. The filtering results are then squared to obtain velocity-squared time-series data from the signal. After that, a smoothing technique using a triangle function is used to obtain the envelope function of the signal. From the envelope signal, the time parameters (Tp), which indicate the signal duration when the signal amplitude reaches p % of the peak amplitude, with p = 90, 80, 50, and 20, are calculated. After the signal durations are obtained, the rupture duration can be calculated using Equation 2 as follows [18][19].

$$T_0 = (1 - w) \cdot T_{90} + w \cdot T_{20} \qquad \text{Equation 2}$$

with w is the weight calculated using the formula in Equation 3 [10].

$$w = \frac{\left[\frac{T_{80}+T_{50}}{2} - 20\right]}{40} \qquad \text{Equation 3}$$

and T_{90}, T_{80}, T_{50}, T_{20} are the signal durations Tp calculated starting from the P-wave arrival time with p = 90, 80, 50, and 20, respectively.

2.2.2. P-wave Dominant Period (T_d)

It is the period representing the most energetic part of the wave in the signal spectrum. The steps to calculate this feature are as follows [20]. The seismogram, the velocity signal, is integrated to obtain the displacement signal. The ratio r between the velocity signal integral and the displacement signal integral in a specific time interval, generally set to 3 seconds [20], is calculated with the formula stated in Equation 4 as follows [20].

$$r = \frac{\int_0^{\tau_0} \dot{u}^2(t)dt}{\int_0^{\tau_0} u^2(t)dt} \qquad \text{Equation 4}$$

with τ_0 representing the observation time in seconds, $\dot{u}(t)$ is the velocity signal in t and $u(t)$ is the displacement signal in t, and t represents time in seconds. The Equation 5 then calculates the T_d parameter as follows.

$$\tau_c = \frac{2\pi}{\sqrt{r}} \qquad \text{Equation 5}$$

where r is the previously calculated ratio.

2.3. Machine Learning used for the Magnitude's Early Estimation

This study uses machine learning-based regression techniques to estimate the magnitude values. The regression methods tested in this study were NNR, RFR, and SVMR. The performance parameters measured are R-square (R^2) and Root Mean Square Error (RMSE).

The dataset used in this study consists of seismic features processed from earthquake signals with the magnitude above 7.0 that occurred from 1990 to 2019. There are 298 earthquakes included in the dataset-making process. All seismic signals used for the dataset were taken from the official public repositories of the GEOFON [21] and Incorporated Research Institutions for Seismology (IRIS) [22]. For the experiment purpose, the dataset is divided into three parts of data, namely training data, validation data, and testing data, with a ratio of 60%:20%:20% for each data part, respectively.

For each regression method tested, hyperparameters tuning is performed to obtain the best parameter configuration that produces the best performance for each method. The hyperparameter tuning is performed on the validation data, which is 20% of the data, using the Grid Search method. The tuning and testing results are then validated using the Nested K-Fold Cross Validation. The nested cross validation consists of two cross validation schemes, i.e., the inner loop and the outer loop. Each scheme uses the K-fold concept with the K values are set to be $K = 4$ for the hyperparameter tuning and $K = 5$ for the final testing validation.

3. Results and Discussions

A list of hyperparameters of each regression method which are tuned for the experiment carried out in this study are described as follows.

1. NNR. The tuned hyperparameters are the number of hidden layers (*hl_size*) including the size parameter of each layer, and the activation function (*act_func*).
2. RFR. There are three hyperparameters that are tuned in the experiment, i.e., the number of estimators (*num_estimators*), the maximum depth of the tree (*max_depth*), and the minimum samples allowed to form new leaves in the generated tree (*min_samples*).
3. SVMR. The hyperparameter tuning is performed on the kernel type (*type_kernel*), the C values (*val_C*), and the gamma values (*gamma_val*) of the regressor model.

The range of values that is set in the tuning process for each hyperparameter can be seen in Table 1. According to the hyperparameter's range of values shown in Table 1, the number of configurations tested in the experiment for each algorithm are 32, 135, and 36 configurations for NNR, RFR, and SVMR, respectively.

Table 1 The Range of Hyperparameter Values

Regressor	Hyperparameter	Range of Values
NNR	*hl_size* *act_func*	[(6,), (8,), (10,), (6,10), (8,10), (10,20), (10,50), (10,100)] ['identity', 'logistic', 'tanh', 'relu']
RFR	*num_estimators* *max_depth* *min_samples*	[10, 20, 50, 100, 200] [1, 2, 3, 4, 5, 10, 15, 20, 25] [2, 5, 10]
SVMR	*type_kernel* *val_C* *gamma_val*	['linear', 'poly', 'rbf'] [1, 5, 10, 30] [0.01, 0.1, 0.2]

The best performance of the experiment results of each regression method tested is shown in Table 2. As mentioned earlier, this study measured two performance parameters in the experiment, i.e., R^2 and RMSE.

Table 2 The Best Results and Hyperparameter Settings

Regressor	R^2	RMSE	Hyperparameter's Best Setting
NNR	0.151	0.342	*hl_size* = (10, 50) *act_func = 'relu'*
RFR	0.946	0.087	*num_estimators* = 200 *max_depth = 25* *min_samples = 2*
SVMR	0.227	0.337	*type_kernel* = 'rbf' *val_C = 30* *gamma_val = 0.2*

The R^2 is the coefficient of determination that provides information about the goodness of fit of the regression model, measuring the proportion of variance in the dependent variable explained by the independent variables. For the case in this study, the dependent variable is the earthquake magnitude, while the independent variables are the predictors represented by the seismic features. A higher R^2 indicates a better fit of the model, meaning that the model accounts for more of the variation in the data.

Table 2 shows that the RFR model has a much higher R^2, which is 0.98, than the NNR and SVMR models, which are 0.151 and 0.22, respectively. It suggests that the RFR model fits the data much better and explains more of the variation in the magnitude estimation. This could be due to the stronger ability of the RFR algorithm to capture complex interactions and relationships among the predictor variables.

On the other hand, the RMSE provides information about the accuracy of the magnitude estimation, measuring the average difference between the predicted and actual values. A lower RMSE indicates better accuracy, meaning that the model has smaller errors in its estimation.

It is shown in Table 2 that the RFR model has the lowest RMSE, which is 0.087. It indicates that it has the highest accuracy of the three models in predicting the earthquake magnitude. The NNR and SVMR models have much higher RMSE values, i.e., 0.342 and 0.337, respectively, suggesting higher errors and lower accuracy in their estimation.

R^2 and RMSE results suggest that the RFR model has a much better fit to the data than the NNR and SVMR models. Therefore, for the earthquake magnitude estimation discussed in this study, the RFR model might be a more appropriate choice for the data.

4. Conclusions

The high R^2 value of 0.98 for the earthquake magnitude estimation using RFR suggests that this method has strong predictive power and can be considered a reliable approach for this task. On the other hand, the low R^2 values obtained from NNR (0.151) and SVMR (0.22) indicate that these methods may not be suitable for earthquake magnitude estimation or require further optimization to improve their performance.

Furthermore, the RMSE results also support the conclusion that the RFR has better accuracy in predicting earthquake magnitudes, as it has the lowest RMSE value of 0.087. The higher RMSE values obtained from NNR (0.342) and SVMR (0.337) indicate that these methods have a higher error rate in predicting earthquake magnitudes.

Overall, the conclusion is that the RFR is the most suitable method for earthquake magnitude early estimation among the three tested methods, at least for the data used in this study.

Bibliography

[1] J. Havskov and L. Ottemöller, "Processing earthquake data", Book in Preparation, Preliminary Version, 2010.

[2] T. Steinmetz, U. Raape, S. Teßmann, C. Strobl, M. Friedemann, T. Kukofka, … and S. Dech, "Tsunami early warning and decision support", Natural Hazards and Earth System Science, 10(9), 1839–1850, 2010.

[3] R. Heidari, "Quick Estimation of the Magnitude and Epicentral Distance Using the P Wave for Earthquakes in Iran", Bulletin of the Seismological Society of America 106(1), 2015.

[4] S. Atefi, R. Heidari, N. Mirzaei and H. R. Siahkoohi, "A frequency-based parameter for rapid estimation of magnitude", Journal of Theoretical and Applied Physics, 2017.

[5] L. H. Ochoa, L. F. Nino and C. A. Vargas, "Fast magnitude determination using a single seismological station record implementing machine learning techniques", Geodesy and Geodynamics, 2018.

[6] A. Katsumata, M. Tanaka and T. Nishimiya, "Rapid estimation of tsunami earthquake magnitudes at local distance", Earth, Planets and Space volume 73, 2021.

[7] J. Zhu, S. Li, S Li, Y. Wei and J. Song, "Rapid earthquake magnitude estimation combining a neural network and transfer learning in China: Application to the 2022 Lushan M6.1 earthquake", Frontiers in Physics, Vol. 11, 2023.

[8] D. Goldberg, Y. Bock and D. Melgar, "Rapid Earthquake Magnitude Estimation for Early Warning Applications", Proceeding of 19th EGU General Assembly, EGU2017, p.11571, 23-28 April, 2017

[9] A. Novianty,; I. Meilano, C. Machbub, S. Widiyantoro and S. Susilo, "Rapid Estimation of Earthquake Magnitude and Source Parameters Using Genetic Algorithms", Applied Science, 11, 11852, 2021.

[10] I. Meilano, S. Susilo and D. Sarsito, "Rapid Magnitude Estimation of the August 5, 2018, Lombok Earthquake using High-Rate GNSS Data", International Journal of GEOMATE, Oct., Vol.23, Issue 98, pp.57-65, 2022.

[11] A. Pfaffinger, "Support Vector Regression vs. Neural Network Regression in the Field of Predicting Drug-Target Binding Activities", thesis, Johannes Kepler University Linz, 2015.

[12] K. Shahare, A. Mitra, D. Naware, R. Keshri, H. M. Suryawanshi, "Performance analysis and comparison of various techniques for short-term load forecasting", Energy Reports, Volume 9, Supplement 1, March 2023.

[13] G. Dudek, "A Comprehensive Study of Random Forest for Short-Term Load Forecasting", Energies Journal, Vol. 15, 2022.

[14] J. Huo, T. Shi, J. Chang, "Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources", 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2016.

[15] M. Ohsaki, N. Kishimoto, H. Sasaki, R. Ikeura, S. Katagiri, K. Ohnishi, Y. Sebastian, P. Then, "NNR-GL: A Measure to Detect Co-Nonlinearity Based on Neural Network Regression Regularized by Group Lasso", IEEE Access, Vol. 9, 2021.

[16] G. F. Fan, L. Z. Zhang, M. Yu, W. C. Hong, S. Q. Dong, "Applications of random forest in multivariable response surface for short-term load forecasting", International Journal of Electrical Power & Energy Systems, Vol. 139, 2022.

[17] Ganesh N., P. Jain, A. Choudhury, P. Dutta, K. Kalita, P. Barsocchi, "Random Forest Regression-Based Machine Learning Model for Accurate Estimation of Fluid Flow in Curved Pipes", Processes Journal, Vol. 9, 2021.

[18] A. Lomax, "Rapid estimation of rupture extent for large earthquakes: Application to the 2004, M9 Sumatra-Andaman mega-thrust", Geophysical Research Letters, 32(10), 1–4, 2005.

[19] A. Lomax, and A. Michelini, "Mwpd: A duration-amplitude procedure for rapid determination of earthquake magnitude and tsunamigenic potential from P waveforms", Geophysical Journal International, 176(1), 200–214, 2009.

[20] Y. M. Wu, and H. Kanamori, "Experiment on an onsite early warning method for the Taiwan early warning system", Bulletin of the Seismological Society of America, 95(1), 347–353, 2005.

[21] GFZ German Research Centre for Geosciences, GEOFON and EIDA Data Archives, 2013. [Retrieved July 10, 2018, from http://eida.gfz-potsdam.de/webdc3/]

[22] Incorporated Research Institutions for Seismology (IRIS), Wilber 3: Select Event, 2010. [Retrieved July 10, 2018, from http://ds.iris.edu/wilber3/find_event]