Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 6 (2023) 1439 - 1447 ISSN Media Electronic: 2580-0760

Forecasting the Magnitude Category Based on The Flores Sea Earthquake

Adi Jufriansah[1], Azmi Khusnani[2], Sabarudin Saputra[3], Dedi Suwandi Wahab[4]
[1]Physics Education, Faculty of Teacher Training and Education, Universitas Muhammadiyah Maumere, Maumere, Indonesia
[2]Informatics, Faculty of Engineering, Business and Health, Universitas Muhammadiyah Maumere, Maumere, Indonesia
[1]saompu@gmail.com, [2]husnaniazmi@gmail.com, [3]sicoccinela@gmail.com, [4]dediwahab04@gmail.com

Abstract

Earthquakes are a phenomenon that is still a mystery in terms of predicting events, one of which is the magnitude. As technology develops, there are many algorithms that can be used as approaches in earthquake forecasting. In the context of magnitude forecasting, the application of GaussianNB, Random Forest and SVM has the potential to reveal these patterns and relationships in the data. With the six main phases of this research, namely data acquisition, data preprocessing, feature selection, model training, forecasting result evaluation, and performance analysis. From these results we obtain, firstly that the GaussianNB model has a relatively simple and fast method in training its model. However, the weakness lies in the assumption of a Gaussian distribution which may not always suit the complex and diverse characteristics of earthquake data. Based on GaussianNB model, the model accurately predicts magnitude category 1 for 421 observations and magnitude category 2 for 33 observations. Meanwhile, the magnitude 3 and magnitude 4 categories did not produce accurate predictions from the model. Second, Random Forest, this method can increase accuracy and overcome the overfitting problem that occurs when forecasting magnitudes. In contrast to GaussianNB, it tends to result in models with greater complexity and require more time to compute. In our findings, we obtained an MSE value of 0.12 with an R^2 score of -0.10, this indicates conditions that are less effective in explaining differences in test data. The third option is SVM, which has both benefits and drawbacks that must be taken into account. The capacity of SVM to separate data that has both linear and non-linear separation is one of its key advantages; nevertheless, the main drawback is that it is sensitive to hyperparameter adjustments. It is clear from the results of the algorithm comparison that SVM has more potential for earthquake forecasting, especially the linear SVM and polynomial SVM model. The accuracy of the standard SVM is 0.587, which indicates relatively low performance. Linear SVM obtained a very high accuracy of 0.998. Meanwhile, Polynomial SVM achieves perfect accuracy of 1.0. while RBF SVM has the same accuracy as standard SVM, namely 0.587.

Keywords: gaussiannb; random forest; support vector machine; earthquake; forecasting

1. Introduction

Due to the potential for bodily harm and fatalities, earthquakes have long been a plague for people [1]. Consequently, in an effort to safeguard civilization from these possible risks, understanding and forecasting earthquakes is crucial [2]. The Flores maritime, a maritime region affected by the Flores Back Arc Trust and susceptible to seismic activity, is one of the regions vulnerable to earthquakes [3]–[6]. The Flores Sea Earthquake Data Catalog is a valuable data source that keeps records of earthquakes that occurred in this region [7]. One of the important parameters in earthquake forecasting is magnitude [8]–[10]. Magnitude is a measure of the size of the energy released by an earthquake and can reflect the level of potential danger caused [11]–[13]. In this research, we focus on forecasting earthquake magnitude categories in the Flores Sea using various approaches, including Gaussian Naive Bayes (GaussianNB), Random Forest, and Suppot Vector Machine (SVM). This method belongs to a family of machine learning algorithms that have proven effective in many forecasting cases [14]–[16].

In the context of magnitude forecasting, the application of GaussianNB, Random Forest and SVM has the potential to reveal hidden patterns and relationships in the data [17], [18]. By utilizing attributes such as earthquake location, depth and time, we hope to be able to predict magnitude categories with a higher degree of accuracy. This will be very useful in making decisions regarding risk mitigation and disaster response preparation in the Flores Sea area. In addition, a study of forecasting earthquake magnitude categories in the Flores Sea using this algorithm can help in a deeper understanding of seismic patterns in this region, as well

Accepted: 16-10-2023 | Received in revised: 25-12-2023 | Published: 28-12-2023

as in developing better forecasting models in the future [2].

This research can also be an initial step in applying machine learning technology for earthquake forecasting purposes in Indonesia. With advances in technology and increasing availability of data, machine learning algorithms have great potential to help understand and predict earthquake behavior more efficiently and accurately [19]–[22]. The challenges faced in earthquake forecasting involve various aspects, including the natural complexity and uncertainty inherent in this geological phenomenon. Therefore, it is necessary to carry out careful validation and evaluation of the forecasting results to ensure the quality and reliability of the resulting model [23]–[25]. In the context of disaster mitigation, magnitude category forecasting can be an invaluable tool in identifying areas at high risk of earthquakes of significant intensity. With this information, authorities can take preventive action, such as improving earthquake-resistant infrastructure, preparing evacuation plans, and educating the public regarding emergency measures in the face of earthquakes.

2. Research Methods

Based on the Flores Sea Earthquake Data Catalog, this research uses algorithms, GaussianNB, Random Forest and SVM. The six main phases of this research are data acquisition, data preprocessing, feature selection, model training, forecasting results evaluation, and performance analysis [2], [26]. First, we collected information about earthquakes that have occurred in the region from the Earthquake Data Catalog, which contains important details such as epicenter location, depth, time of occurrence, and magnitude. Complete catalogue data is presented in Table 1.

Table 1. Catalogue Data

No	Latitude	Longitude	Depth	MagType	Magnitude
0	-8,6071	121,6032	183,417	mb	4,5
1	-8,2739	121,8694	202,54	mb	4
2	-7,2034	120,3696	577,909	mb	4,3
3	-7,7762	122,5902	289,691	mb	4,3
4	-8,8323	122,5431	118,504	mb	4,1
…	…	…	…	…	…
1526	-8,349	122,5642	33	mL	4,9
1527	-8,6223	121,9686	13,3	mb	5,7
1528	-8,4969	122,4823	33	mb	6
1529	-8,475	121,9023	27,7	mw	7,4
1530	-8,5	121,9	33	mL	5,1

The data is then continued in the process of training and evaluating forecasting models using GaussianNB, Random Forest and SVM. The reliability of predictions is highly dependent on the availability of appropriate and representative data [27], [28]. The next stage is data cleaning, to avoid invalid values. This method is important to ensure the correctness of the data used for model training and evaluation. The data scale will also be normalized to ensure that the attributes have the same influence on model construction [29]–[31].

The next stage is feature selection. The features used in this research are latitude, longitude, depth, magnitude, and magnitude category encoded data. In an effort to increase the efficacy and accuracy of the model, only the characteristics that are most relevant and have a significant impact on the quantity category will be considered. For feature selection, correlation analysis or other feature selection techniques can be used as an important stage of this research. Next, the model instruction stage is carried out and then concluded. Using clean data and relevant attributes, training data and test data will be separated into two categories. Using the training data, the GaussianNB, Random Forest and SVM models will begin the training process, with characteristics serving as features and quantity categories serving as labels. For a model to accurately predict the future, it must be taught to recognize and learn new data patterns [32]. After training the model, the predicted results are then subjected to an evaluation process. These models will be evaluated using test data they have never encountered before. Evaluation is based on a comparison between the model's predicted quantity categories and the actual labels from the test data. Accuracy, precision, recall, and F1 score will be used to compare the effectiveness of the three approaches for predicting earthquake magnitude categories. Using performance analysis, the advantages of the two forecasting systems are then contrasted and evaluated. The results of this performance analysis will provide an idea of how effective each technique is at classifying earthquakes in the Flores Sea area. By combining evaluation results and performance analysis, this study is expected to contribute to the development of more accurate and effective earthquake forecasting methods. The complete research flow is presented in Figure 1

DOI: https://doi.org/10.29207/resti.v7i6.1495

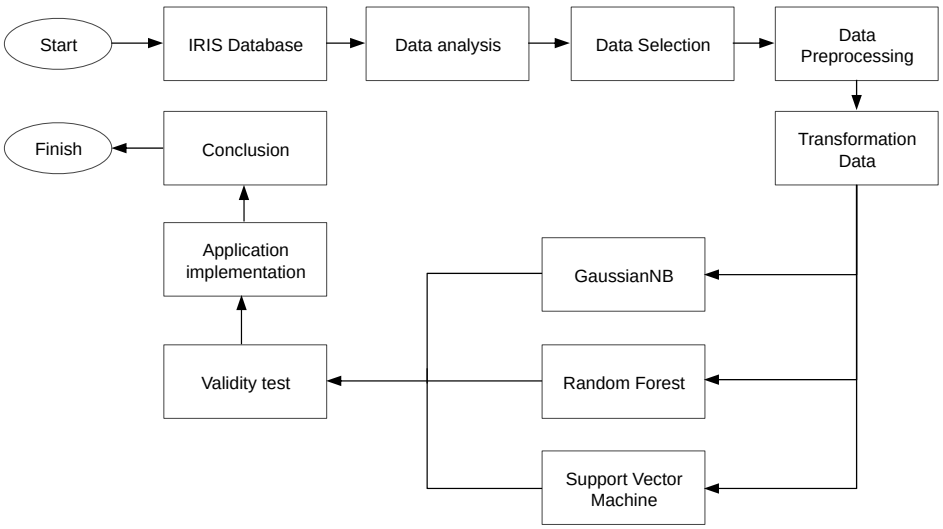


Figure 1. Research flow

3. Results and Discussions

The Gaussian Naïve Bayes (NB) Classifier is a simple probability classification technique that refers to the Bayesian Theorem [33]. The theorem states that the probability of an event occurring is calculated by multiplying the intrinsic probability (estimated using currently available data) by the probability of the same event occurring again in the future (based on historical knowledge). Statistical techniques for drawing inductive conclusions about classification problems [34]. The accuracy results of the Naive Bayes classification are good and consistent. However, if features and data parameters are added, Naive Bayes generates a number of circumstances that will result in poor accuracy scores. Since the data used in this study was numerical, you can use the Probability Density Function (PDF) function to calculate the class probability values.

In Formula 1, Bayes' theorem is shown. while Formula 2 displays the Gaussian NB calculation method.

$$P(h|x) = \frac{p(x|h)p(h)}{p(x)} \quad (1)$$

$$P(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

The GaussianNB method in predicting earthquake magnitude categories in the Flores Sea region has succeeded in showing satisfactory accuracy. This is influenced by a variety of attributes and high-quality data representation. The speed with which GaussianNB can solve multi-class problems makes it a valuable tool for forecasting categories of quantities at various levels [35]. This model uses different probabilities to determine classification, and is easier to understand. Even when the forecasting results show satisfactory performance, there is still room for improvement [36]. This is shown in the results of the confusion matrix presented in Figure 4.

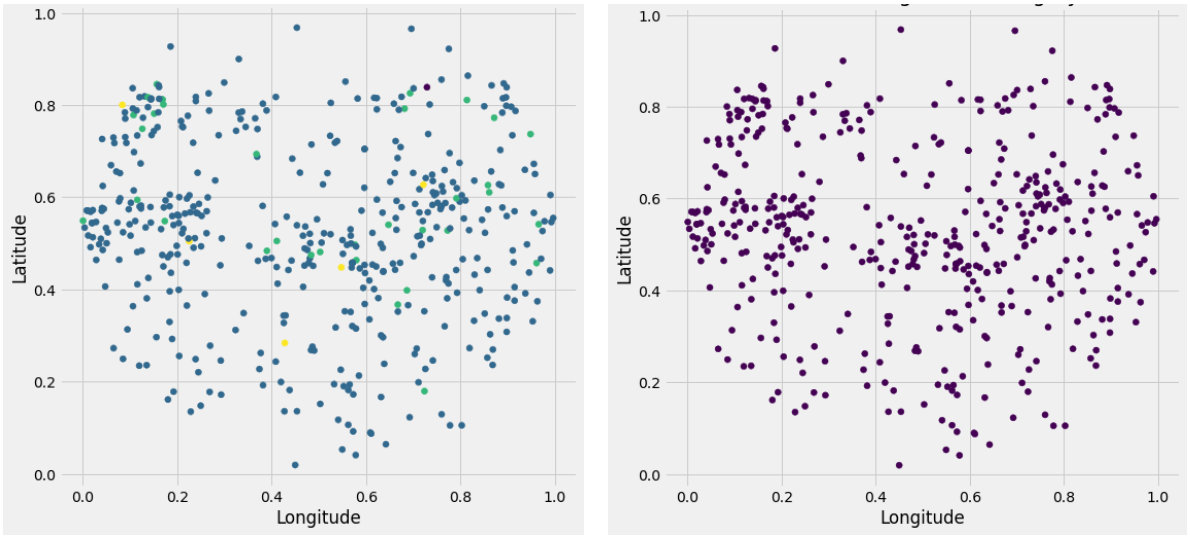
A confusion matrix is a table that shows the extent to which the model correctly and incorrectly identifies each category. The algorithm results show that the confusion matrix provided examines four categories of magnitude, namely 1 (minor), 2 (moderate), 3 (Strong), and 4 (Major). The main diagonal of the confusion matrix represents the number of accurate predictions, while the other diagonal represents the number of inaccurate predictions. Based on these results, the model accurately predicts magnitude category 1 for 421 observations and magnitude category 2 for 33 observations. Meanwhile, the magnitude 3 and magnitude 4 categories did not produce accurate predictions from the model, because none of the categories contained accurate predictions. According to these results (figure 3), the model works well in predicting quantities 1 and 2, but still has difficulty distinguishing quantities 3 and 4. This can be caused by uneven data distribution between categories or variations in attribute characteristics of each category. To improve model performance, additional analysis of variables that influence forecast accuracy for each category is required. Additionally, using data mining techniques or selecting more complex variables can help improve prediction accuracy. The complete classification report is presented as follows.

Classification Report:	precision	recall	f1-score	support
Minor	0.00	0.00	0.00	1
Moderate	0.92	1.00	0.96	421
Strong	0.00	0.00	0.00	33
Major	0.00	0.00	0.00	5
accuracy			0.92	460
macro avg	0.23	0.25	0.24	460
weighted avg	0.84	0.92	0.87	460

DOI: https://doi.org/10.29207/resti.v7i6.1495

The findings of the classification report provide a more comprehensive picture of how well the GaussianNB model performs in predicting various earthquake magnitude categories in the Flores Sea region. The classification report contains evaluation data for each magnitude category, such as precision, recall, and f1-score, in addition to the model's overall accuracy score. Minor, Moderate, Strong, and Major are the four magnitude categories tested in these results. Precision refers to the accuracy of the model in predicting the actual magnitude category. Recall (sensitivity) measures how well the model can recognize and recognize existing categories of quantities. F1-score represents the harmonic mean of recall and precision. With a precision of 0.92 and a recall of 1.00, the categorization report findings demonstrate the model's strong predictive ability for medium-sized categories. In other words, the model is quite adept at identifying and predicting moderate magnitude earthquakes. The high f1 value of this category of 0.96 can serve as proof of the GaussianNB model. This is presented in Figure 4.

In the small, strong, and large magnitude categories, the lack of precision and recall for the three categories indicates that the model is unable to accurately predict these categories. The F1 scores for the Small, Strong, and Major Magnitude categories are also zero, indicating that the model has difficulty distinguishing between earthquakes of these magnitudes. Additionally, the overall accuracy rating of this model is 0.92, indicating that it can effectively predict 92% of the test data used. The accuracy is quite high, but given the subpar performance in certain magnitude categories, these results require further investigation. In fact, to improve model performance, additional research needs to be carried out on variables that influence prediction accuracy in low-performing categories. It is also possible to increase forecasting accuracy by using data mining strategies or by selecting features with a higher level of complexity. The complete forecasting results for the magnitude category are presented in Figure 2.



(a) (b)

Figure 2. (a) Actual Magnitude Category, (b) Predicted Magnitude Category

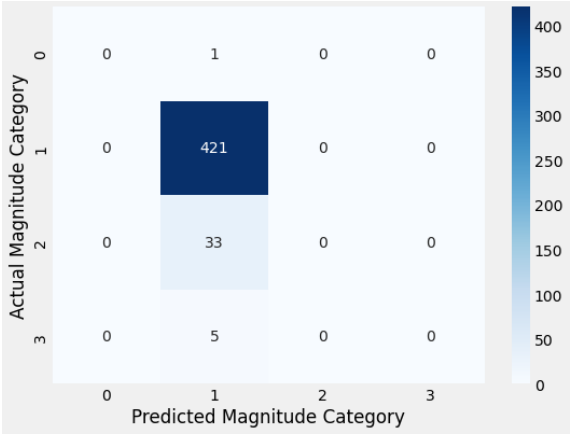


Figure 3. Actual magnitude category vs predicted magnitude category

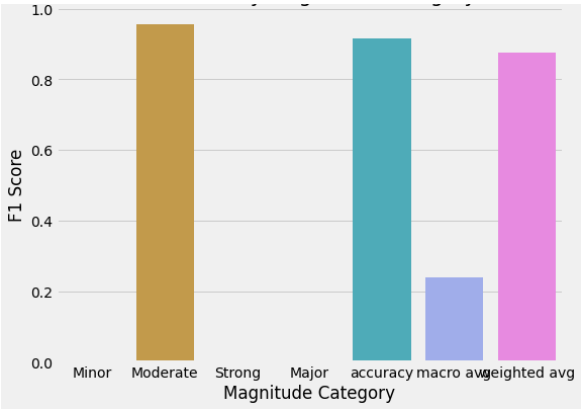


Figure 4. F1 Score by Magnitude Category

DOI: https://doi.org/10.29207/resti.v7i6.1495

Li et. al., [37] claims that the Random Forest (RF) classifier is an ensemble classifier that predicts using a set of CARTs. Using bagging techniques, a portion of the training samples are taken to produce a tree. This implies that some samples may be selected only once, while others may be selected multiple times (figure 1(a)). One-third of the samples referred to as ready-to-use samples are used in the cross-validation procedure to estimate the performance of the resulting Random Forest model, while the remaining two-thirds of the samples are referred to as in-bag samples used to train the tree (Figure 1(a)).

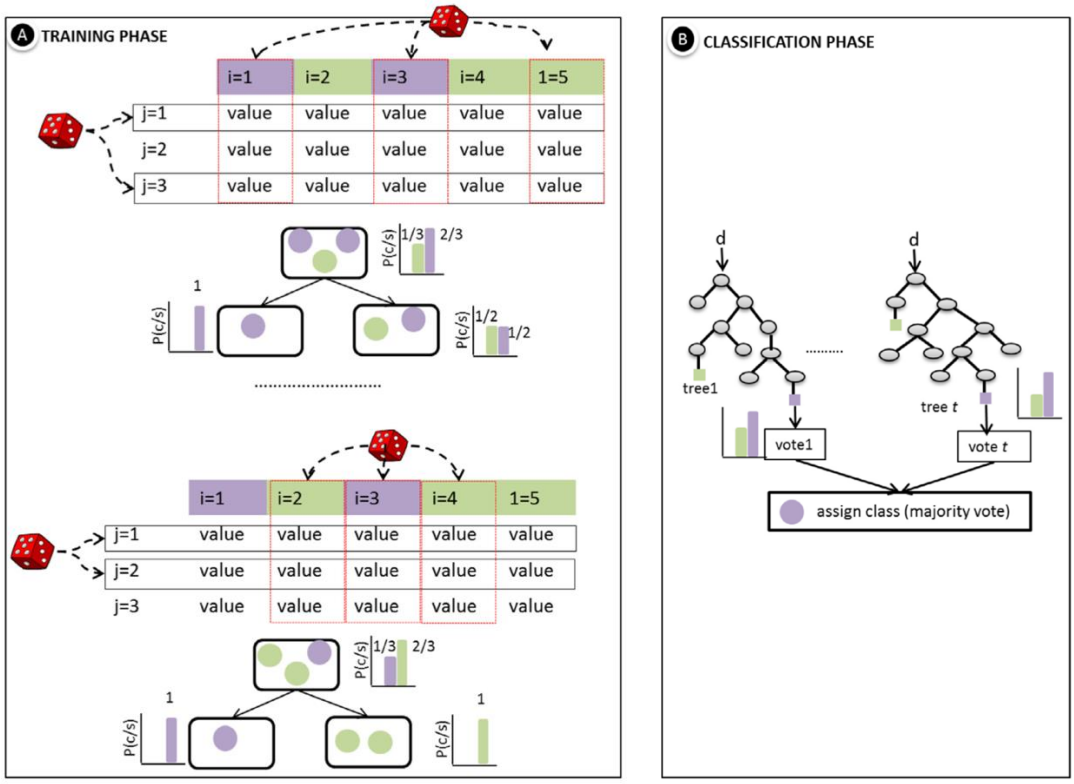
Out-of-bag (OOB) error is the name given to this error estimate. Each tree is created independently, without any pruning, and each node is divided using a randomly selected, user-defined number of features. This approach produces trees with low bias and high variance by expanding the forest to a user-specified number of trees (Ntree) [38]. The class assignment probabilities determined by each resulting tree are then averaged (using the arithmetic mean) to determine the final classification. As a result, each tree built in the ensemble is compared with new input in the form of unlabeled data, and each tree votes for class members. Ultimately, the membership class with the most votes will be selected (Figure 1(b)).

One crucial step in lowering the death toll is estimating when an earthquake will strike a certain location. Using the Random Forest Method is one efficient way to anticipate earthquakes. A modelling technique called Random Forest uses a number of decision trees combined to provide predictions. The benefit of Random Forest lies in its capacity to manage intricate correlations and exchanges across variables, along with its ability to surmount overfitting issues. Random Forest is also an efficient way to handle huge volumes of training data, may provide low error rates, and performs optimally in classification. It is also a useful method for guessing missing data.

The following computational time required to create an Random Forest classification model is presented by Formula 3

$$T\sqrt{MNlog(N)} \quad (3)$$

M is the number of variables used in each split, N is the number of training samples, and T is the number of trees.



(a) (b)

Figure 5. (a) Training phase, (b) Classification phase [38]

Two evaluation measures, namely Mean Squared Error (MSE) and R^2 Score (Coefficient of Determination), can be used to understand the predictions of the Random Forest algorithm [39]. Mean Squared Error, or MSE is a metric that measures how closely a model's predictions match the actual values [14]. The MSE value is determined by the average of the squared differences between the actual value and the predicted value of each test data, with the average of these squared differences functioning as the MSE value [40].

DOI: https://doi.org/10.29207/resti.v7i6.1495

he closer the MSE is to zero, the better, because it indicates that the model predictions are getting closer to the true value [41]. In our findings, we obtained an MSE value of 0.12008229856881615. This result shows that the Random Forest model predictions are quite close to the actual value. Meanwhile, the R^2 score is a metric that evaluates how much variation in test data can be explained by the model [2]. The R^2 value ranges from -1 to 1. A value of -1 indicates that the model has weak predictive ability, while a value of 1 indicates that the model fully explains all the data variance. As shown by the R^2 score using this model of -0.10111866775704192, this indicates conditions that are less effective in explaining differences in test data.

The MSE score calculation shows that the Random Forest model is quite good at predicting earthquake magnitudes in the Flores Sea region, although with some forecasting errors. Meanwhile, a low R^2 score indicates that the model cannot adequately account for data variations. Errors in prediction can be caused by various factors, such as lack of associated features, maximum depth, and node separation criteria that can affect model performance [19]. Model performance can also be influenced by the quality and representativeness of the seismic data used in training and evaluation [42]. The accuracy and precision of the model can be improved with larger and more representative data sets [43]. In addition, a deeper examination of attribute characteristics and their relationships could improve model performance and clarify seismic patterns in the Flores Sea. The results of the Random Forest method show the potential to predict various earthquake magnitude classes in the Flores Sea region. The complete prediction results are presented in Figure 6.

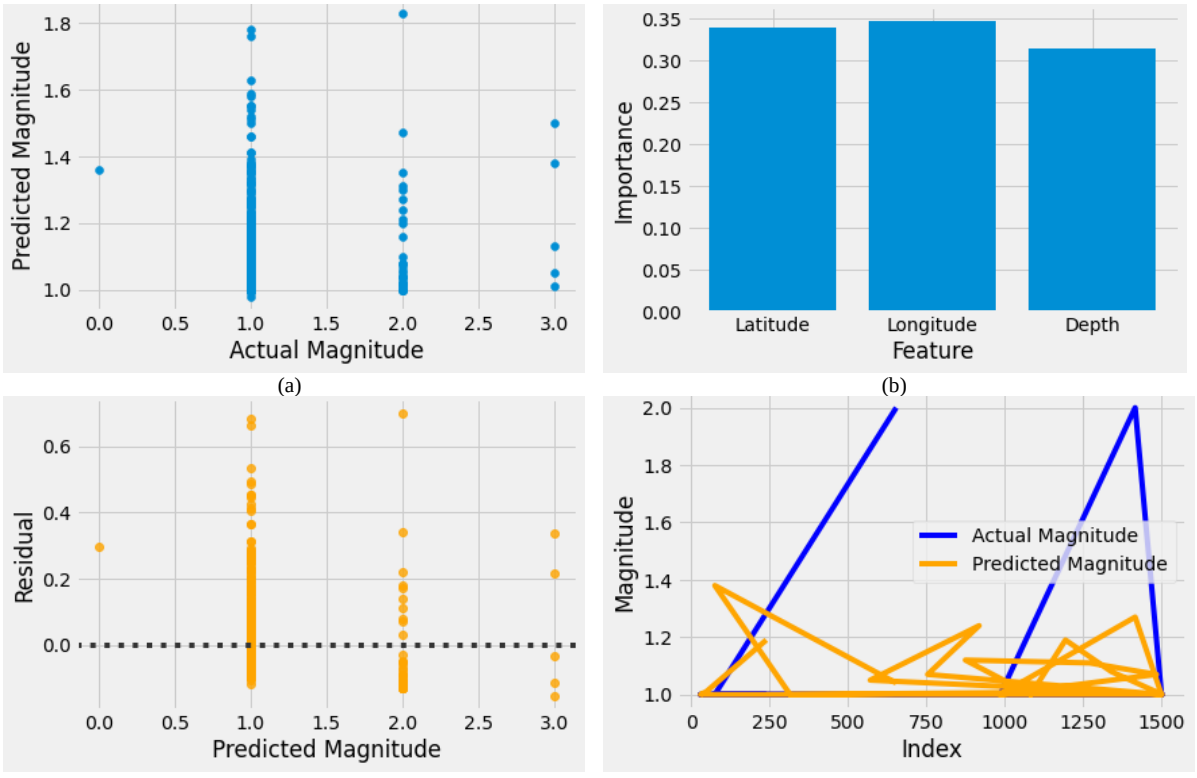


Figure 6. (a) Random Forest Regression Results, (b) Feature Importance Plot, (c) Residual Plot, (d) Actual vs Predicted

Training and test data, each consisting of multiple inputs, are needed for Support Vector Machine (SVM) classification. In the training data, each input has multiple properties in addition to one target value. Based on the attribute values, SVM generates a model that can forecast the test data's target value [44]. The predictor is represented by the symbol x_i, the class by the symbol y_i, which comprises two classes that are presumptively -1 and 1, and the support vector by the symbol w. Formula 4 expresses the hyperplane formulation for linear data.

$$y_i = w.x_i + b \quad (4)$$

Using the Kernel function, which first converts a 2-D vector field into a 3-D vector, the hyperplane for non-linear data is created. Non-linear predictors are simpler to separate in a 3-D vector space than in a 2-D vector field. Formula 5 expresses the hyperplane that is applied to non-linear data using the kernel function.

$$y_i = w.K(x_i, y_j + b \quad (5)$$

30% of the data is used as validation data during the training process, and 70% of the data is used as model

DOI: https://doi.org/10.29207/resti.v7i6.1495

data. For a new input signal, the accuracy results of the classifier with the highest accuracy are used as predictors. Formula 6 can be used to determine the accuracy of the classifier model [15]. True positive, TN true negative, FP false positive, and FN false negative are indicated by the TP parameter.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (5)$$

After Train-Validation Split, for training data, it is compared with test data, which is a significant amount. Two variables form the training set of independent variables and their magnitudes. Table 2 provides a comprehensive comparison of the accuracy of SVM algorithms.

Table 2. Algorithm Accuracy

Algorithm		Accuracy
Support Vektor Machine (SVM)		0.587
Hyperparameter SVM		
1.	SVM Linear	0.998
2.	SVM Polynomial	1.0
3.	SVM RBF	0.587

The data illustrates the accuracy results of various SVM configurations, including SVM with standard settings, Linear SVM, Polynomial SVM, and RBF SVM. The accuracy of the standard SVM is 0.587, which indicates relatively low performance on this particular dataset. However, Linear SVM obtained a very high accuracy of 0.998, demonstrating its excellent ability to separate data linearly. Meanwhile, Polynomial SVM achieved a perfect accuracy of 1.0, which could indicate overfitting if tested on the same dataset as the training data. Finally, the RBF SVM has the same accuracy as the standard SVM, namely 0.587, which may imply that using the RBF kernel on this dataset does not provide a significant advantage. Selection of the appropriate SVM model and hyperparameters is critical, and cross-validation needs to be performed to ensure that the model can generalize well on new data [15], [22], [26], [45]–[49]. In conclusion, Polynomial SVM has the best performance, but needs to be tested further to avoid overfitting, while Linear SVM also shows excellent performance in separating data linearly.

4. Conclusion

Forecasting methods using GaussianNB, Random Forest, and SVM are three interesting approaches and can provide valuable insight in efforts to predict earthquake magnitude categories in the Flores Sea region. Each of these methods is part of a machine learning algorithm that has proven effective in various fields, including forecasting. However, each has its own characteristics and advantages. First, GaussianNB is a method based on Bayes' Theorem with the assumption that attributes have a Gaussian distribution. This method is relatively simple and fast in model training. However, the weakness lies in the assumption of a Gaussian distribution which may not always suit the complex and diverse characteristics of earthquake data. Second, Random Forest is an ensemble algorithm that works by combining several decision trees. This allows this method to increase accuracy and overcome the overfitting problem that often occurs in machine learning models. However, Random Forest has a tendency to increase model complexity and longer computing time compared to GaussianNB. Third is SVM which has a number of advantages and disadvantages that need to be considered. One of the main advantages of SVM is its ability to separate data that has both linear and non-linear separation. This happens because SVM allows the use of various kernels, such as linear, polynomial, and RBF kernels, which allow SVM to address various types of classification problems. Additionally, SVM also has the potential to prevent overfitting if hyperparameters are adjusted correctly and cross-validation is performed well. In particular, the Polynomial SVM configuration can achieve very high accuracy, which is useful in classification cases that require high precision. On the other hand, SVM has several disadvantages that need to be considered. One of its main drawbacks is sensitivity to hyperparameter settings. Choosing the right hyperparameters can be a complex and time-consuming task. Additionally, SVM can be computationally intensive, especially on large datasets, requiring significant computing time to train the model. SVM may also be inefficient if used on very large datasets because it requires large amounts of storage. Additionally, SVMs may not handle highly imbalanced data well without special treatment such as the use of class weights or oversampling/undersampling techniques. So, the use of SVM has the potential to provide good results in a variety of situations, but users must be careful in choosing the right configuration, considering the size of the dataset, and dealing with class imbalance. Cross-validation and careful hyperparameter selection are important to maximize SVM performance in a variety of classification tasks.

Acknowledgment

The researcher would like to thank the Indonesian Ministry of Education, Culture, Research and Technology for providing research funds through the 2023 Beginner Lecturer Research Grant Scheme (PDP).

References

[1] A. Khusnani, A. Jufriansah, and M. Afriyanto, "Utilization of Seismic Data as a Tsunami Vulnerability Review," *Indonesian Review of Physics*, vol. 5, no. 2, pp. 66–72, 2022, doi: https://doi.org/10.12928/irip.v5i2.6706.

[2] A. Jufriansah *et al.*, "Estimation of Flores Sea Aftershock Rupture Data Based on AI," *Indones. Rev. Phys*, vol. 6, no. 1, pp. 46–56, 2023, doi: 10.12928/irip.v6i1.6705.

[3] R. Maneno, B. J. Sentosa, and G. Rachman, "Relocation Of Earthquake Hypocenter In The Flores Region Using

DOI: https://doi.org/10.29207/resti.v7i6.1495

Hypo71," *IPTEK The Journal of Engineering*, vol. 5, no. 2, May 2019, doi: 10.12962/j23378557.v5i2.a5024.

[4] R. Rahardiawan and C. Purwanto, "Struktur Geologi Laut Flores, Nusa Tenggara Timur.," *Jurnal Geologi Kelautan*, vol. 12, no. 3, pp. 153–163, 2014, doi: DOI: 10.32693/jgk.12.3.2014.256.

[5] L. Handayani, "Seismic Hazard Analysis of Maumere, Flores: a Review of the Earthquake Sources," in *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019*, 2020, pp. 1–7. doi: 10.4108/eai.12-10-2019.2296247.

[6] E. D. Banggut and B. J. Santosa, "Estimation of the Source Parameters of the Flores Earthquake and Their Correlation to Aftershocks," *IPTEK The Journal of Engineering*, vol. 5, no. 1, pp. 14–17, 2019, doi: http://dx.doi.org/10.12962/j23378557.v5i1.a5021.

[7] A. Jufriansah, A. Khusnani, Y. Pramudya, and M. Afriyanto, "Comparison of aftershock behavior of the flores sea 12 december 1992 and 14 december 2021," *Journal of Physics: Theories and Applications*, vol. 7, no. 1, pp. 65-74., 2023, doi: https://doi.org/10.20961/jphystheor-appl.v7i1.71609.

[8] A. S. N. Alarifi, N. S. N. Alarifi, and S. Al-Humidan, "Earthquakes magnitude predication using artificial neural network in northern Red Sea area," *J King Saud Univ Sci*, vol. 24, no. 4, pp. 301–313, Oct. 2012, doi: 10.1016/j.jksus.2011.05.002.

[9] S. Cihad, E. Al-Heety, and W. Abdulnaby, "Relocation and Magnitude for Earthquakes in the Outer Arabian Platform of Iraq," *Iraqi Geological Journal*, vol. 56, no. 1B, pp. 198–215, Feb. 2023, doi: 10.46717/igj.56.1B.15ms-2023-2-23.

[10] M. N. Shodiq, D. H. Kusuma, M. G. Rifqi, A. R. Barakbah, and T. Harsono, "Spatial analisys of magnitude distribution for earthquake prediction using neural network based on automatic clustering in Indonesia," in *2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, IEEE, Sep. 2017, pp. 246–251. doi: 10.1109/KCIC.2017.8228594.

[11] W.-Y. Kim, L. Ottemöller, and P. G. Richards, "A Regional *Sn* Magnitude Scale mb(Sn) and Estimates of Moment Magnitude for Earthquakes along the Northern Mid-Atlantic Ridge," *Bulletin of the Seismological Society of America*, vol. 110, no. 6, pp. 3158–3173, Dec. 2020, doi: 10.1785/0120190304.

[12] D. Kusumawati *et al.*, "Sensitivity analysis of moment magnitude (Mw) spectral fitting parameters based on orthogonal design: using aftershocks of the 2019 Mw 6.5 Ambon, Indonesia earthquake," *Bulletin of Earthquake Engineering*, vol. 21, no. 4, pp. 1793–1819, Mar. 2023, doi: 10.1007/s10518-022-01606-6.

[13] A. Jufriansah, Y. Pramudya, A. Khusnani, and S. Saputra, "Analysis of Earthquake Activity in Indonesia by Clustering Method," *Journal of Physics: Theories and Applications*, vol. 5, no. 2, p. 92, Sep. 2021, doi: 10.20961/jphystheor-appl.v5i2.59133.

[14] F. Aden-Antoniów, W. B. Frank, and L. Seydoux, "An Adaptable Random Forest Model for the Declustering of Earthquake Catalogs," *J Geophys Res Solid Earth*, vol. 127, no. 2, Feb. 2022, doi: 10.1029/2021JB023254.

[15] A. E. Ruano, G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, and P. M. Ferreira, "Seismic detection using support vector machines," *Neurocomputing*, vol. 135, pp. 273–283, Jul. 2014, doi: 10.1016/j.neucom.2013.12.020.

[16] H. S. Kuyuk, E. Yildirim, E. Dogan, and G. Horasan, "Application of k-means and Gaussian mixture model for classification of seismic activities in Istanbul," *Nonlinear Process Geophys*, vol. 19, no. 4, pp. 411–419, Aug. 2012, doi: 10.5194/npg-19-411-2012.

[17] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights," *Struct Health Monit*, vol. 21, no. 4, pp. 1906–1955, Jul. 2022, doi: 10.1177/14759217211036880.

[18] F. M. Wani, J. Vemuri, C. Rajaram, and K. S. K. K. Reddy, "Investigating the efficiency of machine learning algorithms in classifying pulse-like ground motions," *J Seismol*, Aug. 2023, doi: 10.1007/s10950-023-10168-2.

[19] R. J. Geller, "Earthquake prediction: a critical review," *Geophys J Int*, vol. 131, no. 3, pp. 425–450, Dec. 1997, doi: 10.1111/j.1365-246X.1997.tb06588.x.

[20] A. P. Menon, A. Varghes, J. P. Joseph, J. Sajan, and N. Francis, "Performance Analysis of different Classifiers for Earthquake prediction: PACE," *International Journal of Innovative Research in Technology (IJIRT)*, vol. 7, no. 2, pp. 142–146, 2020, doi: https://doi.org/10.6084/m9.figshare.12623573.

[21] S. Baranov, C. Narteau, and P. Shebalin, "Modeling and Prediction of Aftershock Activity," *Surv Geophys*, Apr. 2022, doi: 10.1007/s10712-022-09698-0.

[22] C. Jiang, X. Wei, X. Cui, and D. You, "Application of support vector machine to synthetic earthquake prediction," *Earthquake Science*, vol. 22, no. 3, pp. 315–320, Jun. 2009, doi: 10.1007/s11589-009-0315-8.

[23] M. N. Shodiq, D. H. Kusuma, M. G. Rifqi, A. R. Barakbah, and T. Harsono, "Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia," *JOIV : International Journal on Informatics Visualization*, vol. 2, no. 1, p. 37, Feb. 2018, doi: 10.30630/joiv.2.1.106.

[24] S. Marhain, A. N. Ahmed, M. A. Murti, P. Kumar, and A. El-Shafie, "Investigating the application of artificial intelligence for earthquake prediction in Terengganu," *Natural Hazards*, vol. 108, no. 1, pp. 977–999, Aug. 2021, doi: 10.1007/s11069-021-04716-7.

[25] K. Weiqi, W. Weisong, and Z. Maoxing, "Integrated Learning Algorithms with Bayesian Optimization for Mild Steel Mechanical Properties Prediction," *Knowledge-based Engineering and Sciences*, vol. 3, no. 3, pp. 101–112, 2022, doi: 10.51526/kbes.2022.3.3.101-112.

[26] W. Astuti, R. Akmeliawati, W. Sediono, and M. J. E. Salami, "Hybrid Technique Using Singular Value Decomposition (SVD) and Support Vector Machine (SVM) Approach for Earthquake Prediction," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 7, no. 5, pp. 1719–1728, May 2014, doi: 10.1109/JSTARS.2014.2321972.

[27] S. Karimzadeh, M. Matsuoka, J. Kuang, and L. Ge, "Spatial Prediction of Aftershocks Triggered by a Major Earthquake: A Binary Machine Learning Perspective," *ISPRS Int J Geoinf*, vol. 8, no. 10, p. 462, Oct. 2019, doi: 10.3390/ijgi8100462.

[28] Md. H. Al Banna *et al.*, "Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges," *IEEE Access*, vol. 8, pp. 192880–192923, 2020, doi: 10.1109/ACCESS.2020.3029859.

[29] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl Soft Comput*, vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[30] V. Tschannen, A. Ghanim, and N. Ettrich, "Partial automation of the seismic to well tie with deep learning and Bayesian optimization," *Comput Geosci*, vol. 164, p. 105120, Jul. 2022, doi: 10.1016/j.cageo.2022.105120.

[31] S. Panigrahi and H. S. Behera, "Effect of Normalization Techniques on Univariate Time Series Forecasting using Evolutionary Higher Order Neural Network," 2013.

[32] A. Aprianti, A. Jufriansah, B. Donuata, A. Khusnani, and J. Ayuba, "Comparison of K-Means Algorithm and DBSCAN on Aftershock Activity in the Flores Sea: Seismic," *Journal of Novel Engineering Science and Technology*, vol. 2, no. 3, pp. 77–82, 2019, doi: 10.56741/jnest.v2i03.393.

[33] B. B. Odevci, E. Emsen, and M. N. Aydin, "Machine learning algorithms for lamb survival," *Comput Electron Agric*, vol. 182, p. 105995, Mar. 2021, doi: 10.1016/j.compag.2021.105995.

[34] A. Yudhana, D. Sulistyo, and I. Mufandi, "GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia,"

DOI: https://doi.org/10.29207/resti.v7i6.1495

Sens Biosensing Res, vol. 33, p. 100435, Aug. 2021, doi: 10.1016/j.sbsr.2021.100435.

[35] M. S. Abdalzaher, S. S. R. Moustafa, M. Abd-Elnaby, and M. Elwekeil, "Comparative Performance Assessments of Machine-Learning Methods for Artificial Seismic Sources Discrimination," *IEEE Access*, vol. 9, pp. 65524–65535, 2021, doi: 10.1109/ACCESS.2021.3076119.

[36] M. S. Abdalzaher, M. S. Soliman, and S. M. El-Hady, "Seismic Intensity Estimation for Earthquake Early Warning Using Optimized Machine Learning Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023, doi: 10.1109/TGRS.2023.3296520.

[37] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification," *Appl Soft Comput*, vol. 70, pp. 1000–1009, Sep. 2018, doi: 10.1016/j.asoc.2017.07.027.

[38] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[39] A. Lulli, L. Oneto, and D. Anguita, "Mining Big Data with Random Forests," *Cognit Comput*, vol. 11, no. 2, pp. 294–316, Apr. 2019, doi: 10.1007/s12559-018-9615-4.

[40] F. Soleimani, "Analytical seismic performance and sensitivity evaluation of bridges based on random decision forest framework," *Structures*, vol. 32, pp. 329–341, Aug. 2021, doi: 10.1016/j.istruc.2021.02.049.

[41] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "Identifying channel sand-body from multiple seismic attributes with an improved random forest algorithm," *J Pet Sci Eng*, vol. 173, pp. 781–792, Feb. 2019, doi: 10.1016/j.petrol.2018.10.048.

[42] J. Roiz-Pagador, A. Chacon-Maldonado, R. Ruiz, and G. Asencio-Cortes, "Earthquake Prediction in California Using Feature Selection Techniques," 2022, pp. 728–738. doi: 10.1007/978-3-030-87869-6_69.

[43] R. Mallouhy, C. A. Jaoude, C. Guyeux, and A. Makhoul, "Major earthquake event prediction using various machine learning algorithms," in *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, IEEE, Dec. 2019, pp. 1–7. doi: 10.1109/ICT-DM47966.2019.9032983.

[44] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, p. 1331, Jun. 2020, doi: 10.12928/telkomnika.v18i3.14756.

[45] D. Tao, Q. Ma, S. Li, Z. Xie, D. Lin, and S. Li, "Support Vector Regression for the Relationships between Ground Motion Parameters and Macroseismic Intensity in the Sichuan–Yunnan Region," *Applied Sciences*, vol. 10, no. 9, p. 3086, Apr. 2020, doi: 10.3390/app10093086.

[46] M. Wieland, W. Liu, and F. Yamazaki, "Learning Change from Synthetic Aperture Radar Images: Performance Evaluation of a Support Vector Machine to Detect Earthquake and Tsunami-Induced Changes," *Remote Sens (Basel)*, vol. 8, no. 10, p. 792, Sep. 2016, doi: 10.3390/rs8100792.

[47] I. Aljarah, A. M. Al-Zoubi, H. Faris, M. A. Hassonah, S. Mirjalili, and H. Saadeh, "Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm," *Cognit Comput*, vol. 10, no. 3, pp. 478–495, Jun. 2018, doi: 10.1007/s12559-017-9542-9.

[48] P. Hajikhodaverdikhan, M. Nazari, M. Mohsenizadeh, S. Shamshirband, and K. Chau, "Earthquake prediction with meteorological data by particle filter-based support vector regression," *Engineering Applications of Computational Fluid Mechanics*, vol. 12, no. 1, pp. 679–688, Jan. 2018, doi: 10.1080/19942060.2018.1512010.

[49] A. Jufriansah, A. Anggraini, Z. Zulfakriza, A. Khusnani, and Y. Pramudya, "Forecast earthquake precursor in the Flores Sea," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 3, p. 1825, Dec. 2023, doi: 10.11591/ijeecs.v32.i3.pp1825-1836.

DOI: https://doi.org/10.29207/resti.v7i6.1495