Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 22, No. 3, July 2023, pp. 583∼592
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v22i3.3061

Data Mining Earthquake Prediction with Multivariate Adaptive Regression Splines and Peak Ground Acceleration

Dadang Priyanto[1], Bambang Krismono Triwijoyo[1], Deny Jollyta[2], Hairani Hairani[1], Ni Gusti Ayu Dasriani[1]
[1]Universitas Bumigora, Mataram, Indonesia
[2]Institut Bisnis dan Teknologi Pelita Indonesia , Riau, Indonesia

Article Info

Article history:

Received June 11, 2023
Revised June 28, 2023
Accepted July 14, 2023

Keywords:

Earthquake hazard prediction
Function base
Predictor variable
Peak ground acceleration
Nonparametric regression

ABSTRACT

Earthquake research has not yielded promising results because earthquakes have uncertain data parameters, and one of the methods to overcome the problem of uncertain parameters is the nonparametric method, namely Multivariate Adaptive Regression Splines (MARS). Sumbawa Island is part of the territory of Indonesia and is in the position of three active earth plates, so Sumbawa is prone to earthquake hazards. Therefore, this research is important to do. This study aimed to analyze earthquake hazard prediction on the island of Sumbawa by using the nonparametric MARS and Peak Ground Acceleration (PGA) methods to determine the risk of earthquake hazards. The method used in this study was MARS, which has two completed stages: Forward Stepwise and Backward Stepwise. The results of this study were based on testing and parameter analysis obtained a Mathematical model with 11 basis functions (BF) that contribute to the response variable, namely (BF) 1,2,3,4,5,7,9,11, and the basis functions do not contribute 6, 8, and 10. The predictor variables with the greatest influence were 100% Epicenter Distance and 73.8% Magnitude. The conclusion of this study is based on the highest PGA values in the areas most prone to earthquake hazards in Sumbawa, namely Mapin Kebak, Mapin Rea, Pulau Panjang, and Pulau Saringi.

Corresponding Author:

Dadang Priyanto, +62818362756
Faculty of Engineering and Computer Science Study Program,
Universitas Bumigora, Mataram, Indonesia,
Email: dadang.priyanto@universitasbumigora.ac.id

How to Cite:
D. Priyanto, B. Triwijoyo, D. Jollyta, H. Hairani, and N. Dasriani, "Data Mining Earthquake Prediction with Multivariate Adaptive Regression Splines and Peak Ground Acceleration", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 22, no. 3, pp. 583-592, Jul. 2023.

1. INTRODUCTION

Earthquakes are natural disasters that can cause minor to severe damage. Many lives and property were lost as a result of the earthquake. Research on earthquakes to date has not provided significant results to determine the causative factors or when the earthquake occurred. Many studies have been carried out, but the problem is that the data related to earthquakes is uncertain and involves big data as a result of recording the accelerograph machine, so an appropriate method is needed to perform predictive analysis based on past data. Research has been carried out using various methods, such as classification and regression methods, Artificial Neural Networks (ANN), Support Vector Regression (SVR), Hybrid Neural Networks (HNN), and others, but these methods use a parametric approach, while earthquake data The earth is uncertain, so an appropriate method is needed, namely a nonparametric approach such as the Multivariate Adaptive Regression Spline (MARS) method. Earthquakes can occur anywhere, including on Sumbawa Island, West Nusa Tenggara. Sumbawa Island is part of the Indonesian archipelago and is positioned west of the Alas Strait. Sumbawa, with an area of 15,448 square km, has an active volcano that once erupted violently in 1815 and had an impact on the whole world with changes in weather and the distribution of volcanic ash up to 1,300 km. The existence of Sumbawa is geologically in the position of the western and eastern island arcs due to the subduction of the Australian plate at the continental boundary of the Indo-Pacific plate, which is to the south of Sumbawa Island. Because Sumbawa is at a plate-meeting position, Sumbawa Island is an area that is prone to tectonic earthquakes. History shows that Sumbawa often experiences earthquakes with a magnitude scale of more than 5 with a depth of less than 70 Kilometers. Many methods have been developed, and research on earthquake prediction is included in the scope of data mining research [1]. Data mining is grouped into two, namely predictive data mining and descriptive data mining. Data mining, in principle, is the process of finding certain patterns and knowledge from large data sources [1]. In the data mining process, mathematical functions are needed, such as Association, Correlation, Classification, Regression, and Clustering functions [1–3]. Many methods are used in the data mining process, and one of them is the Multivariate Adaptive Regression Spline (MARS) method [4–6]. MARS is a nonparametric method for solving high-dimensional data problems and is used to find the relationship between predictor variables and response variables. MARS is very effectively used in processing earthquake prediction data mining, as will be done in this study.

Research related to earthquake prediction using various methods has been carried out, such as using the Artificial Neural Network method [7]. The study calculated sixty seismic features with seismological concepts such as the Gutenberg-Richter law, foreshock frequency, seismic energy release, seismic change rate, and total repeat time. Furthermore, Maximum Relevance and Minimum Redundancy (MRMR) are used to extract the relevant features. Classification methods are used with Vector Regressor (SVR) and Hybrid Neural Networks (HNN) in making earthquake predictions. The numerical results obtained after being compared with previous prediction studies show an increase in prediction performance for all regions considered [7]. This study evaluates earthquake events for 50 years with a magnitude range of 0 to 8 Mw. Artificial Neural Networks are used to analyze earthquake data to produce predictions of the next earthquake magnitude [8]. Comparison Three methods, namely Levenberg Marquardt Backpropagation (LMBP), Recurrent Neural Network (RNN), and Radial Basis Function (RBF), have also been carried out to predict and evaluate four different statistical measures. The results of the three RNN methods provide better predictive accuracy [9, 10]. The Artificial Neural Network (ANN) is used to predict aftershocks for the next five days after earthquakes occur in several areas in Indonesia. Six clusters were used for analysis using Valley Tracing and Hill Climbing algorithms, while Hierarchical K-means were applied to the group data set. The evaluation results give better results in predicting the occurrence of aftershocks of or greater than 6 magnitudes [11]. Another technique that uses the Fuzzy Logic System method to model mathematically in predicting earthquakes, and the results of evaluating MAPE and MSE values for the best model for predicting earthquakes is to provide 7 inputs and 1 output from the fuzzy model [12]. The Innovative Mathematical Model (IMM) is used to analyze earthquake events in the last 20 years, and the Poisson distribution and Spatial Connection methods for each earthquake zone and identify patterns of random earthquake events [13].

Earthquake data is random and contains a high element of uncertainty, so choosing the right method for carrying out predictive analysis is necessary. Hence, the MARS method is very suitable for predicting earthquake data [14]. Another study with MARS and C-MARS explains that Ground Motion Prediction Equations (GMPEs) is an empirical relationship used to determine the response of the ground peak at a certain distance from the earthquake source. Conic Multivariate Adaptive Regression Splines (CMARS) method on available data sets to obtain new GMPE. The CMARS model uses PGA and PGV values as dependent variables. In contrast, three other parameters, such as moment magnitude (Mw), station location conditions (Vs30), and distance from earthquake source (Rjb), are used as independent variables. The findings of this study indicate that CMARS can be used effectively to predict PGA and PGV values at various distances from the earthquake source. The results were compared with the other three GMPEs, and CMARs were more effective for ground motion prediction purposes [15–18].

This study aims to develop research using the MARS method in a case study of the earthquake that occurred on Sumbawa Island, Indonesia. This research differs from earthquake prediction research conducted by other researchers because the basic nature of earthquake research is influenced by bedrock conditions where the earthquake's location varies from one area to another, with different soil and rock structures. This research was conducted by developing the function of the Mathematical model formed by MARS according to the condition of the regional bedrock on Sumbawa Island. This researcher will use three predictor variables to find correlations in predictive analysis. This study will analyze predictions of earthquake-prone areas on Sumbawa Island based on Peak Ground Acceleration data with the highest value.

2. RESEARCH METHOD

2.1. Multivariate Adaptive Regression Spline (MARS)

The MARS method is a nonparametric regression method used to overcome the problem of high-dimensional data, which is used to determine the relationship pattern between the response variable and the predictor variable whose regression curve is not known [19]. In data mining management, predictions can be completed in two ways: Parametric Regression and Nonparametric Regression. These two approaches are commonly used as statistical methods and widely used for investigating and modeling relationships between variables [20]. The MARS method can overcome the shortcomings of Recursive Partitioning Regression (RPR) by producing a continuous model at knots and identifying the presence of an additive linear function. Two stages of the algorithm can solve the MARS method, namely the Forward Stepwise model and the Backward Stepwise model [18, 21, 22]. The first stage, namely the Forward Stepwise Algorithm, is used for a combination of basis functions (BF), maximum interaction (MI), and minimum observation (MO). to find the relationship between the response variables and predictor variables. This research has determined that the response variable is Peak Ground Acceleration (PGA), and the predictor variables are depth, magnitude (Mw), and epicenter distance (Repi). Furthermore, the Backward Stepwise model's second stage is used to simplify the basis function (BF) obtained from the Forward Stepwise stage. The basis function (BF), which has no contribution or makes a small contribution to the response variable, will be eliminated at the backward stepwise model stage. This deletion process will have the effect of decreasing the number of least squares of the remainder. In general, the Nonparametric Regression model can be presented as in Equation (1) [23–25].

$$yi = f(xi) + \mathcal{E}i \tag{1}$$

Where yi is the response variable on observation I, $f(xi)$ is the vector predictor variable function, and $\mathcal{E}i$ is a free error i. The determination of the independent variable greatly determines the results of the model built using the MARS method so that the MARS model is flexible, and its basic functions can be explained in Equations (2) and (3). Equations (2) and (3) seem almost the same function, so they can be called reflected pairs. The goal is reflected pairs on each variable x j on each observation xi, j on the knots of the variable so that a truncated linear function is formed from the basis function as in Equation (4). The MARS model starts from Equation (5).

$$(x-r)_+ = \begin{cases} x-r, & \text{if } x > r \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

and

$$(x-r)_+ = \begin{cases} x-r, & \text{if } x \geq r \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$r = \left\{(x_j - r)_+ , \ (r - x_j)_+ | \ r \ \in \{x_{1j}, x_{2j}, \ldots, x_{Nj}\} , \ j = 1, 2, \ldots, p\right\} \tag{4}$$

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m \beta_m(x) \tag{5}$$

Where M is the number of basis functions that make up the function model. $\beta_m\,(x)$ is a basic function formed by a single element or by multiplying two or more elements contained in r, multiplied by the coefficient β_m. The m basic function can be explained into the basis function as shown in Equation (6). Where K_m is the number of truncated linear functions times the basis function to m. For $X_{k_j}^{m}$ is the input variable associated with the truncated function in the mth basis function. $\tau_{k_j}^{m}$ is the value of the knot variable $\tau_{k_j}^{m}$. While $S_{k_j}^{m}$ is operator +/-, which is worth 1 or -1.

$$\beta_m\,(x^m) = \prod_{j=1}^{K_m} \left[S_{k_j}^{m} \left(X_{k_j}^{m} - \tau_{k_j}^{m}\right)\right] +, \tag{6}$$

The MARS model is flexible and can be used to overcome the weaknesses of recursive partition regression by increasing the accuracy of the model. The MARS model is run with a two-stage algorithm: Forward Stepwise and Backward Stepwise. Then the algorithm will determine the value of knots in the continuous model and minimize the value of Generalized Cross Validation (GCV) to obtain the best model. GCV measurement can be seen in Equation (7). Where y_i is Variabel response, x_i is Variable predictor, N is the number of observations, $\hat{f}_M(x_i)$ is the estimated value of the dependent variable on the M basis function on x_i, M is the maximum number of base functions, $\hat{C}(M)$ is $C(M) + d.M$, $C(M)$ is $Trace\ [B(BTB) - 1BT] + 1$, where B is a matrix of M basis functions, and d is the value when each base function reaches optimization $(2 \leq d \leq 4)$.

$$GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{f}_M(xi)]^2}{\left[1 - \frac{\hat{C}(M)}{N}\right]^2} \tag{7}$$

2.2. Peak Ground Acceleration (GPA)

Maximum Ground Acceleration (PGA) is the maximum ground vibration acceleration that occurs in an area caused by an earthquake. A large PGA value in an area usually has a large damage impact on the area at the center of the earthquake. The unit of PGA value is usually expressed in units of Gravitational Acceleration "gal." One way to get the PGA value is by using the empirical calculation of the Attenuation function. The attenuation function determines the relationship between ground vibration intensity, magnitude, and distance from an area to the earthquake's epicenter. Several factors affect the attenuation function, namely the earthquake mechanism, the epicenter's distance, and the ground location's condition. This research is to get the PGA value using the Attenuation function of the Joyner and Boore Attenuation equations as in the following Formula (8) [26, 27].

$$PGA\ (gal) = 10(0,71 + 0,23(M - 6) - Log(r) - 0,0027.r) \tag{8}$$

$$r = \sqrt{R\_epi^2 + 8^2} \tag{9}$$

Where M is the magnitude, and r is the root of (R_epi2+82). The PGA value will be obtained by assigning a value to 'M' and the value of 'r' in Equation (9). Furthermore, the research was conducted with Prediction analysis with previous selection/separation and selection of appropriate variables for Responsive and Predictor variables. This study uses the response variable 'PGA,' and the predictor variables are depth, magnitude (Mw), and epicenter distance (Repi).

2.3. Data Collection

This study uses earthquake catalog data taken from sources on the USGS website (https://earthquake.usgs.gov/earthquakes/search/). The data was accessed on May 31, 2022, at 10:47 AM and has been filtered with a magnitude of more than 4 Mw. This is done because a Magnitude of less than 4 Mw does not have a significant impact or may not be felt at all. The coordinate position taken is -8,59491 South Latitude and 117,26121 East Longitude. Earthquake catalog data was obtained over 20 years with a total of 105 records in the range of Magnitude 4 to the highest of 5.5 Mw. The data is processed by a selection system with a magnitude of more than 4 Mw, a depth of less than 250 Km, and an earthquake center distance of less than 300 Km. Data that is not in the provisions or ring will be deleted or not used because it does not cause damage. The data is processed for use in predictive analysis. Three data variables have been determined: Magnitude, Epicenter Distance, and depth of the center of the earthquake location.

The earthquake prediction analysis process used the Multivariate Adaptive Regression Spline (MARS) method using the equation function number 6, and to get the minimum value of Generalized Cross Validation (GCV) used equation number 7. The SPM 8 software was used to predict earthquakes by analyzing the parameter factor of the relationship between the predictor variable and the response variable. MARS works with two algorithms, the Forward Stepwise and the Backward Stepwise algorithms. The Forward Stepwise algorithm determines the combination of the maximum basis function with maximum interaction and minimum observation (MO).

Maximum basis function for cross multiplying between variables that have linkage and correlation. Maximum Interaction (MI) to describe the maximum line in the basis function (BF) that can be traversed or past the knot point, and the minimum observation to obtain a minimum smoothing parameter value or, in other words, the minimum observation between knots. Furthermore, the Backward Stepwise algorithm is used to simplify the complexity of the formed mathematical model functions. This algorithm uses a regularization technique to minimize generalization errors by using the Tikhonov Regularization technique, which gives a penalty if the function of the formed mathematical model is too complex. Peak Ground Acceleration (PGA) is used to determine whether an area is categorized as prone or not to earthquake hazards. A high PGA value in an area will have a high impact due to the occurrence of an earthquake. The PGA value is obtained from recording using an Accelerograph machine or by empirical calculations, and this study uses empirical calculations using the Joyner and Boore Attenuation functions, as in the equations of functions number 8 and 9.

3. RESULT AND ANALYSIS

3.1. Results

The study's results began with preprocessing the data to find the value of the epicenter distance and Maximum Ground Acceleration (PGA). The Joyner and Boore attenuation functions were used to find the PGA value. After knowing the PGA value, the calculation and prediction analysis can be continued using the MARS method. At this stage, to get the best MARS model, it is necessary to test the data and determine the best model by selecting the minimum GCV value. Peak Ground Acceleration (PGA) is the maximum ground vibration acceleration that occurs in an area caused by an earthquake. A large PGA value in an area usually has a large damage impact on the area at the center of the earthquake. The unit of PGA value is usually expressed in units of Gravitational Acceleration "gal." One way to get the PGA value is by using the empirical calculation of the Attenuation function. The attenuation function determines the relationship between ground vibration intensity, magnitude, and distance from an area to the earthquake's epicenter. Several factors affect the attenuation function, namely the earthquake mechanism, the epicenter's distance, and the ground location's condition. This research is to get the PGA value using the Attenuation function of the Joyner and Boore Attenuation equations as in the following Formula (8) and (9) [26]. The PGA value was obtained from the results of processing earthquake data in Sumbawa from 2000 to 2021, as shown in Table 1.

Table 1. PGA Value for Earthquake in Sumbawa

Long	Lat	Depth	Mw	AVECOS	Repi	r	log PGA	PGA(g)
117.8202	-8.9048	44.32	5.1	0.9883619	70.457604	70.910323	-1.5391673	0.0288957
117.8051	-8.8713	36.3	4.6	0.9884063	67.234924	67.709195	-1.6254625	0.0236885
117.8057	-8.9356	49.79	5.3	0.988321	70.840676	71.290963	-1.4965201	0.0318772
117.4461	-9.1528	100.58	4.7	0.9880304	65.282689	65.771038	-1.5846165	0.0260246
116.8967	-8.3735	40.91	4.7	0.9890566	47.024816	47.700454	-1.3963137	0.0401501
117.5666	-8.5657	171.14	4.5	0.9888077	33.757461	34.692451	-1.2689046	0.0538388
116.7862	-8.3791	10	4.6	0.9890494	57.46731	58.021477	-1.5322468	0.0293598
117.7632	-9.0262	110.96	4.8	0.9882002	73.110058	73.546452	-1.6311371	0.023381
117.7632	-8.361	10	5.1	0.9890727	61.049659	61.571591	-1.4526237	0.0352676
117.7632	-8.4768	10	4.9	0.9889232	56.76373	57.324698	-1.4561185	0.034985
117.7632	-8.3893	10	4.6	0.9890362	59.774942	60.307907	-1.5552056	0.027848
117.7632	-8.422	11.58	5.5	0.988994	58.478568	59.02324	-1.3353858	0.046197
117.7632	-8.543	10	5.2	0.9888373	55.519608	56.09302	-1.37436	0.0422318
117.7632	-8.4278	9.61	4.5	0.9889866	58.269313	58.815924	-1.5632979	0.0273339
117.7632	-8.4038	10	4.5	0.9890175	59.176037	59.714348	-1.5723074	0.0267727
117.7632	-8.4087	10	4.5	0.9890112	58.982238	59.522302	-1.5703899	0.0268912
117.7632	-8.446	10	4.8	0.988963	57.654607	58.20699	-1.488134	0.0324987
117.7632	-8.3828	10	5.2	0.9890446	60.05553	60.586027	-1.4199547	0.0380229
117.7632	-8.377	10	5.3	0.9890521	60.312111	60.840371	-1.3994609	0.0398602
117.7632	-8.415	10	5	0.9890031	58.739555	59.281829	-1.4529825	0.0352385
117.7632	-8.8919	127.07	4.6	0.988379	64.318381	64.813997	-1.5986666	0.0251961
117.7632	-8.9508	100.5	4.9	0.9883008	67.910506	68.380091	-1.5625559	0.0273807
117.7632	-8.3939	168.87	4.6	0.9890303	59.580871	60.115557	-1.5532989	0.0279706
117.7632	-8.7612	117.24	4.8	0.9885517	58.217277	58.764371	-1.4937779	0.0320791
117.7632	-9.037	64.81	4.6	0.9881857	73.90276	74.334501	-1.6838936	0.0207065
117.7632	-8.966	44.7	5.1	0.9882805	68.908192	69.371023	-1.5254799	0.0298209
117.7632	-8.954	50	4.7	0.9882965	68.118274	68.586437	-1.6104216	0.0245233
117.7632	-8.971	44.9	5	0.9882739	69.242253	69.702867	-1.5514484	0.02809
117.7632	-8.882	35	4.6	0.9883921	63.760801	64.260717	-1.5934495	0.0255006
117.7632	-8.524	146.7	4.5	0.988862	55.780214	56.350974	-1.5380491	0.0289702
117.7632	-8.931	35	4.5	0.9883271	66.652781	67.131164	-1.6431783	0.0227416
117.7632	-8.847	30.6	5.2	0.9884385	61.90634	62.42111	-1.4378685	0.0364864

Furthermore, the research was conducted with Prediction analysis with previous selection/separation and selection of appropriate variables for Responsive and Predictor variables. This study uses the response variable 'PGA,' and the predictor variables are depth, magnitude (Mw), and epicenter distance (Repi). The results of selecting the appropriate type of variables in the prediction analysis data can be obtained, as shown in Table 2. The results of the prediction analysis using the MARS method using the Forward Stepwise algorithm and the Backward Stepwise algorithm based on a combination of BF, MI, and MO are in the form of training data. The results of the MARS regression based on the training data are shown in Table 3.

Table 2. Data on Response Variables and Predictors of Earthquakes in Sumbawa

Depth	Mw	Repi	PGA(g)
44.32	5.1	70.4576041	0.028895663
36.3	4.6	67.2349243	0.023688498
49.79	5.3	70.8406764	0.031877182
100.58	4.7	65.2826885	0.026024566
40.91	4.7	47.0248155	0.040150066
171.14	4.5	33.7574611	0.053838803
10	4.6	57.4673104	0.02935981
110.96	4.8	73.1100578	0.023380988
10	5.1	54.7251628	0.040821878
10	4.9	36.7876717	0.060202688
10	4.6	37.7199042	0.049861564
11.58	5.5	41.526454	0.071544947
10	5.2	34.4625147	0.076161062
9.61	4.5	50.3673126	0.03309364
10	4.5	45.9733817	0.03715543
10	4.5	47.2993799	0.035850709
10	4.8	38.4930653	0.054110685
10	5.2	47.2897638	0.051952829
10	5.3	51.5476395	0.049065729
10	5	51.8219065	0.041571507
127.07	4.6	60.4038662	0.027456473
100.5	4.9	41.4528397	0.052181561
168.87	4.6	50.3650706	0.034895496
117.24	4.8	40.4900619	0.050921282
64.81	4.6	77.3671974	0.019369949
44.7	5.1	46.9598263	0.049709907
50	4.7	44.9069075	0.042534853
44.9	5	45.8472392	0.048586533
35	4.6	34.5821119	0.055206542
146.7	4.5	50.4304764	0.033040392
35	4.5	45.2402482	0.037909235
30.6	5.2	64.7445074	0.034305771

Table 3. Results of Training Data

Parameter	Estimate	SE.	T-Value	P-Value
Constant	0.04286	0.00029	148.3594	0
Basis Function 1	-0.00131	0.00007	-18.98361	0
Basis Function 2	0.00151	0.00003	58.67444	0
Basis Function 3	0.02211	0.0011	20.06196	0
Basis Function 4	-0.02334	0.00054	-43.43563	0
Basis Function 5	0.00107	0.00008	13.60181	0
Basis Function 7	0.00038	0.00003	10.96367	0
Basis Function 9	0.00031	0.00004	6.99231	0
Basis Function 11	0.00039	0.00006	6.58793	0
F-STATISTIC = 5977.78679			S.E. OF REGRESSION = 0.00035	
P-VALUE = 0.00000			RESIDUAL SUM OF SQUARES = 0.00000	
[MDF,NDF] = [8, 23]			REGRESSION SUM OF SQUARES = 0.00577	

3.2. Testing and Analysis

In predictive analysis, a statistical analysis test is needed to obtain the hypothesis testing results and determine the significance level. The significance level is meant to get the parameter significance. Hypothesis testing is required to use statistical analysis to determine the significance of parameters with the suitability of the mathematical model obtained. This research tests mathematical model analysis using a partial regression coefficient test. In testing the partial regression coefficient, the following Formula is needed:

$H_0 : a_1 = a_2 = a_3 = a_5 = a_7 = a_8 = a_9 = a_{11} = 0$

H_1 : there is at least one am $\neq$ 0;

$m = 1, 2, 3, 4, 5, 7, 9, 11$ (significant model)

- Significant level, α = 0,05
- Statistic test: $tcount = \frac{\hat{a}_m}{Se(\hat{a}_m)} with Se\left(\hat{a}_m\right) = \sqrt{var\left(\hat{a}_m\right)}$
- Critical Area: refuse H_0 if $t > t_{(\frac{\alpha}{2},61)} or P - value < \alpha$

P-value in statistical tests used to determine the magnitude of the opportunity, to state the status Reject the null hypothesis or (H0) with the actual condition (H0) is true. As shown in Table 3 (results of training data) that the P-value is less than 0.05, or in other words, every m $<\alpha$ or (m $<$ 0.05) so that the H0 status is rejected. This means that each coefficient $\alpha_1, \alpha_2,\ \alpha_3,\ \alpha_4\alpha_5, \alpha_7,\ \alpha_9,\ \alpha_{11}$ has a significant effect on the mathematical model obtained. Based on the significance level of 5%, the mathematical model in Formula (10) is significant. It can be used in predictive analysis of the PGA value for earthquake data sets in Sumbawa. Furthermore, after knowing the suitability of the parameters and mathematical models obtained based on testing, it is concluded that the variables that affect the PGA value are epicenter distance (R-epi), magnitude (Mw), and depth (Depth).

3.3. Discussion

It can be seen in Table 3 that the parameters formed with 11 basis functions that contribute to the response variable are Basis functions 1, 2, 3, 4, 5, 7, 9 and 11. Several basis functions do not contribute to the response variable, namely base functions 6, 8, and 10, then the basic function is omitted or deleted. The results of testing the data at the Backward Stepwise stage by simplifying the function can be obtained from a Mathematical model as in Formula (10).

$$\begin{aligned} Y \ = \ & 0.042863 - 0.00130501 * BF1 + 0.00151234 * BF2 + 0.0221103 * BF3 - 0.0233377 * BF4 + \\ & 0.00106639 * BF5 + 0.000377886 * BF7 + 0.000305277 * BF9 + 0.000391561 * BF11; \end{aligned} \tag{10}$$

MODEL PGA = BF1 BF2 BF3 BF4 BF5 BF7 BF9 BF11;

Where $Y_{(PGA)}$ is the result of PGA Prediction analysis with the MARS model with the contribution of each basis function (BF) as follows:

BF1 = max(0, REPI - 50.3651);
BF2 = max(0, 50.3651 - REPI);
BF3 = max(0, MW - 5.1);
BF4 = max(0, 5.1 - MW);
BF5 = max(0, MW - 4.8) * BF2;
BF7 = max(0, REPI - 41.4528) * BF4;
BF9 = max(0, REPI - 64.7445);
BF11 = max(0, REPI - 44.9069);

Based on the best MARS model, the predictor variable inference that affects PGA is obtained based on the MARS model according to the smallest GCV value sequentially based on the percentage of its contribution, namely the distance of the epicenter (Repi), the magnitude (Mw), and the depth (Depth) as shown in Table 4, which describes the interactivity of the predictor variable's contribution to the response variable.

Table 4. The Interactivity of Predictor Variable Contributions

Variable	Importance	-gcv
REPI	100.00000	0.00023
MW	73.80473	0.00012
DEPTH	0.00000	0.00000

It can be seen in Table 4 that the most influential variables in the PGA value are the Epicenter Distance (Repi) of 100% and the Magnitude (Mw) of 73.8%, while the depth (Depth) does not contribute at all or 0%. The test results to clarify the description of the interactivity of the variable contributions of each predictor variable can be seen in Figure 1 of the Three Dimensional graphs of the contribution of the predictor variable to the response variable.

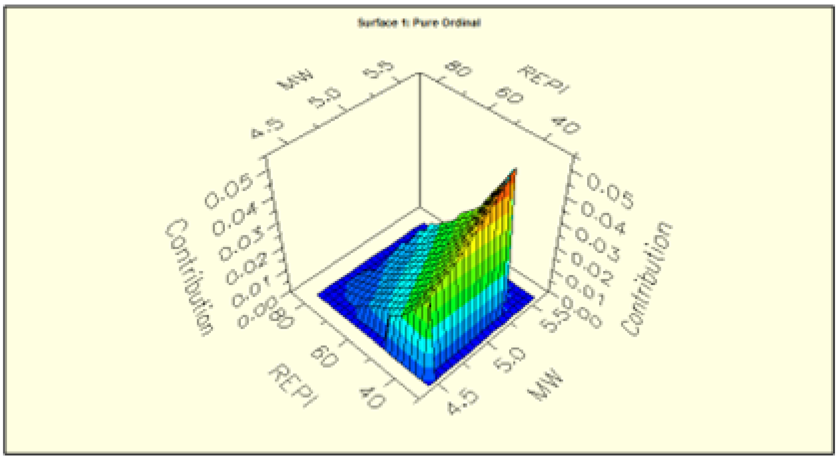

Figure 1. Graph of the Contribution of the Predictor Variable to the Response Variable

As seen in Figure 1, the three-dimensional graph shows that the lower the value of the epicenter distance (Repi), the higher the contribution value to the Response variable, and this means that the closer the epicenter distance, the higher the impact of damage caused by earthquakes. Likewise, it can be seen that the larger the Magnitude (Mw) variable value, the higher the contribution value to the Response variable, meaning that the greater the magnitude value, the greater the damage caused by the earthquake. After going through the testing and validation of the Prediction Analysis results, the Regions in Sumbawa with the Highest Potential for Earthquake Hazards can be identified based on the highest PGA values referring to Table 1, namely Mapin Kebak, Mapin Rea, Pulau Panjang, and Pulau Saringi. Based on the calculation of the PGA value, which is influenced by the magnitude, depth, and distance of the earthquake location. In theory, based on a high PGA value will have a high impact on earthquake damage, although other factors affect earthquake damage, such as the condition of the bedrock of the location. Based on the results of the prediction analysis by grouping the areas with the highest earthquake vulnerability in Sumbawa, policymakers can use it to make rules in infrastructure development with special specifications in earthquake-prone areas.

Based on a literature search, no earthquake prediction research was found that specifically mapped areas in Sumbawa prone to earthquakes. However, other studies discuss, in general, that Sumbawa Island is an earthquake-prone area, as explained by Haryadi, that the potential for an earthquake on Sumbawa Island is very likely to occur because in the northern part of Sumbawa Island, there are micro tectonic plates that extend from Singaraja Bali to Dompu Regency and there is a hemisphere fracture. This threat originates from the south, which is at the bottom of the Indian Ocean because of the Indo-Australian oceanic plate [28]. This is reinforced by the results of research conducted by Sabtaji, who stated that the results of his research in West Nusa Tenggara Province, including the island of Sumbawa, have a number of monthly earthquakes the most, namely the seismicity that occurred in August 2018 as many as 1,658 earthquake events [29]. Another research by Hidhajah stated that major earthquakes occurred from July to August 2018, which impacted food poisoning among refugees in the Alas area, Sumbawa district [30]. Based on the results of previous research, it can be concluded that the authors' research gave the same results that Sumbawa Island is included in areas prone to earthquake hazards. The authors have been able to cluster Sumbawa Island, which areas have the greatest risk of earthquake hazards.

4. CONCLUSION

Based on earthquake catalog data from 2000 to 2021, this study analyzes earthquake hazard predictions in Sumbawa using the MARS method, which involves 11 basic functions. There is a close relationship between the predictor variable and the response variable, with a percentage of 100% epicenter distance and 73.8% magnitude. Based on PGA data, the Potential Areas with a great earthquake hazard in Sumbawa are Mapin Kebak, Mapin Rea, Pulau Panjang, and Pulau Saringi. The analysis of earthquake hazard predictions in Sumbawa can be used as a consideration in infrastructure development in Sumbawa to minimize the risk of earthquake hazards. Furthermore, this research can be developed by adding the number of predictor variables and the number of basis functions to provide more accurate prediction results.

5. ACKNOWLEDGEMENTS

Praise be to God Almighty, who has given His Grace so that this paper can be completed for publication. Many thanks to the Rector of Bumigora University, the Head of LPPM, and all those who directly or indirectly helped in the preparation and publication of this paper.

6. DECLARATIONS

AUTHOR CONTIBUTION

The authors compiled this study, which is divided into their respective tasks. Dadang Priyanto compiled and analyzed the methods used, system testing, and provision of software in research. Bambang Krismono designed interfaces and created data structures, making tables and graphs. Deny Jollyta conducted a theoretical study and comparison of methods, as well as an analysis of the results of PGA and Hairani collected data and the initial process of processing the data in calculating the PGA value with the Attenuation function.

FUNDING STATEMENT

This research did not receive grants from anywhere, including the public, commercial, or non-profit sectors.

COMPETING INTEREST

This study has no reserves related to competing financial, public, or institutional interests.

REFERENCES

[1] R. Kumar, "Data Mining Techniques for Weather Forecasting," *The Visual Computer*, vol. 15, no. 5, pp. 56–63, 2023.

[2] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Cham: Springer International Publishing, 2022, no. March.

[3] P. Dubey, "Analysis of Clustering-Algorithms for Efficient Data Mining," no. April, 2023.

[4] A. Özmen, "Multi-objective regression modeling for natural gas prediction with ridge regression and CMARS," *International Journal of Optimization and Control: Theories and Applications*, vol. 12, no. 1, pp. 56–65, 2022.

[5] A. Özmen, G. W. Weber, and I. Batmaz, "The new Robust CMARS (RCMARS) method," in *24th Mini EURO Conference on Continuous Optimization and Information-Based Technologies in the Financial Sector, MEC EurOPT 2010*, no. June, 2010, pp. 362–368.

[6] J. Zhen, F. J. de Ruiter, E. Roos, and D. den Hertog, "Robust Optimization for Models with Uncertain Second-Order Cone and Semidefinite Programming Constraints," *INFORMS Journal on Computing*, vol. 34, no. 1, pp. 196–216, 2022.

[7] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks," *PLOS ONE*, vol. 13, no. 7, pp. 1–22, jul 2018.

[8] T. A. Adagunodo, S. Lüning, A. M. Adeleke, J. O. Omidiora, A. P. Aizebeokhai, K. D. Oyeyemi, and O. S. Hammed, "Evaluation of 0 M 8 earthquake data sets in African Asian region during 19662015," *Data in Brief*, vol. 17, no. April, pp. 588–603, apr 2019.

[9] B. Sadhukhan, S. Chakraborty, S. Mukherjee, and R. K. Samanta, "Climatic and seismic data-driven deep learning model for earthquake magnitude prediction," *Frontiers in Earth Science*, vol. 11, no. February, pp. 1–24, feb 2023.

[10] X. Zhang, W. ReichardFlynn, M. Zhang, M. Hirn, and Y. Lin, "Spatiotemporal Graph Convolutional Networks for Earthquake Source Characterization," *Journal of Geophysical Research: Solid Earth*, vol. 127, no. 11, nov 2022.

[11] 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," *International Journal on Informatics Visualization*, vol. 2, no. 1, pp. 37–43, 2018.

[12] R. Zakaria, A. N. Jifrin, S. N. Jaman, and R. Roslee, "Fuzzy Interpolation Curve Modelling of Earthquake Magnitude Data," *IOP Conference Series: Earth and Environmental Science*, vol. 1103, no. 1, pp. 1–12, nov 2022.

[13] X. Jianjun, Y. Bingjie, and W. Rongji, "Study on China's Earthquake Prediction by Mathematical Analysis and its Application in Catastrophe Insurance," *IOP Conference Series: Earth and Environmental Science*, vol. 128, no. 1, pp. 1–7, mar 2019.

[14] J. Fayaz, Y. Xiang, and F. Zareian, "Generalized ground motion prediction model using hybrid recurrent neural network," *Earthquake Engineering and Structural Dynamics*, vol. 50, no. 6, pp. 1539–1561, 2021.

[15] D.-C. Feng, X.-Y. Cao, D. Wang, and G. Wu, "A PDEM-based non-parametric seismic fragility assessment method for RC structures under non-stationary ground motions," *Journal of Building Engineering*, vol. 63, no. January, pp. 1–17, jan 2023.

[16] G. Zheng, P. Yang, H. Zhou, C. Zeng, X. Yang, X. He, and X. Yu, "Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines," *Computers and Geotechnics*, vol. 113, p. 103099, sep 2019.

[17] D. Priyanto, M. Zarlis, H. Mawengkang, and S. Efendi, "Analysis of earthquake hazards prediction with multivariate adaptive regression splines," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 2885–2893, 2022.

[18] C. K. Arthur, V. A. Temeng, and Y. Y. Ziggah, "Multivariate Adaptive Regression Splines (MARS) approach to blast-induced ground vibration prediction," *International Journal of Mining, Reclamation and Environment*, vol. 34, no. 3, pp. 198–222, 2020.

[19] S. D. Prihastuti Yasmirullah, B. W. Otok, J. D. Trijoyo Purnomo, and D. D. Prastyo, "Modification of Multivariate Adaptive Regression Spline (MARS)," *Journal of Physics: Conference Series*, vol. 1863, no. 1, pp. 1–12, mar 2021.

[20] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning," *Frontiers in Neuroscience*, vol. 14, no. May, pp. 1–19, may 2020.

[21] F. Ferraccioli, L. M. Sangalli, and L. Finos, "Nonparametric tests for semiparametric regression models," *TEST*, no. June, pp. 1–25, jun 2023.

[22] D. Priyanto, M. Zarlis, H. Mawengkang, and S. Efendi, "Alternative approach to seismic hazards prediction using non parametric adaptive regression method," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 21, pp. 3425–3435, 2020.

[23] S. Xiang and W. Yao, "Nonparametric statistical learning based on modal regression," *Journal of Computational and Applied Mathematics*, vol. 409, no. January, pp. 1–15, aug 2022.

[24] E. Koken, "Yatay Milli Krclarda Krma Kapasitesinin Regresyon, Yapay Sinir Alar ve Çok Deikenli Uyarlamal Regresyon Analizi Kullanlarak Modellenmesi," *Afyon Kocatepe University Journal of Sciences and Engineering*, vol. 22, no. 5, pp. 1193–1203, 2022.

[25] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, "RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 5059–5074, sep 2022.

[26] M. Rajif and S. Syafriani, "Hazard Seismic Zonation Analysis Of West Sumatra Region Using Probabilistic Hazard Seismic Analysis (Phsa) Method," *Pillar or Physics*, vol. 14, no. 1, pp. 8–17, jul 2021.

[27] A. Akinci, I. Munafò, and L. Malagnini, "S-Wave Attenuation Variation and its Impact on Ground Motion Amplitudes During 20162017 Central Italy Earthquake Sequence," *Frontiers in Earth Science*, vol. 10, no. July, pp. 2011–2015, 2022.

[28] W. Haryadi, "Gempa Tektonik Di Pulau Sumbawa Dan Dampaknya Terhadap Bangunan Sipil (Suatu Kajian Geologis)," *Ganec Swara*, vol. 6, no. 2, pp. 13–19, 2012.

[29] A. Sabtaji, "Statistics of Tectonic Earthquake Events Each Proovince in Indonesia Territory for 11 Years of Observation (2009-2019)," *Buletin Meteorologi, Klimatologi, dan Geofisika*, vol. 1, no. 7, pp. 31–46, 2020.

[30] A. C. Hidajah, F. Febriyanti, and D. R. Faisal, "Faktor Risiko KLB Keracunan Makanan Pasca Gempa Bumi di Kabupaten Sumbawa," *Jurnal Kesehatan*, vol. 14, no. 2, pp. 65–70, 2021.