International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 3, June 2022, pp. 2885~2893
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i3.pp2885-2893

Analysis of earthquake hazards prediction with multivariate adaptive regression splines

Dadang Priyanto[1], Muhammad Zarlis[2], Herman Mawengkang[2], Syahril Efendi[2]
[1]Graduate Program of Computer Science, Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
[2]Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received May 1, 2021
Revised Jan 5, 2022
Accepted Jan 23, 2022

Keywords:

Conic multivariate
Multivariate
Non parametric
Peak ground acceleration
Prediction analysis

ABSTRACT

Earthquake research has not yielded promising results, either in the form of causes or revealing the timing of their future events. Many methods have been developed, one of which is related to data mining, such as the use of hybrid neural networks, support vector regressor, fuzzy modeling, clustering, and others. Earthquake research has uncertain parameters and to obtain optimal results an appropriate method is needed. In general, several predictive data mining methods are grouped into two categories, namely parametric and non-parametric. This study uses a non-parametric method with multivariate adaptive regression spline (MARS) and conic multivariate adaptive regression spline (CMARS) as the backward stage of the MARS algorithm. The results of this study after parameter testing and analysis obtained a mathematical model with 16 basis functions (BF) and 12 basis functions contributing to the model and 4 basis functions not contributing to the model. Based on the level of variable contribution, it can be written that the epicenter distance is 100 percent, the magnitude is 31.1 percent, the location temperature is 5.5 percent, and the depth is 3.5 percent. It can be concluded that the results of the prediction analysis of areas in Lombok with the highest earthquake hazard level are Malaka, Genggelang, Pemenang, Tanjung, Tegal Maja, Senggigi, Mangsit. Meninting, and Malimbu.

Corresponding Author:

Dadang Priyanto
Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
Medan, Indonesia
Email: dadangpriyanto@students.usu.ac.id

1. INTRODUCTION

Earthquakes are natural disasters that can cause moderate 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 be able to determine the factors causing and or when the earthquake occurred. Many methods have been developed in research related to earthquake prediction. In the field of computer science, research on earthquake prediction is included in the scope of data mining research. In 2012, Han classifies the data mining process into two groups, namely predictive data mining and descriptive data mining. Predictive data mining is in principle a process of finding certain patterns and knowledge from big data sources [1]–[3]. A mathematical function is needed in the data mining process, such as association, correlation, classification, regression, and clustering functions [1]–[3]. Many methods are used in the predictive data mining process, one of which is the multivariate adaptive regression spline (MARS) method [4], [5].

The MARS method is a non-parametric method that is very effective to overcome the problem of high-dimensional data that is used to determine the relationship between predictor variables and response variables. The problem in earthquake prediction is the existence of uncertain parameters and with the MARS method the function of the mathematical model is influenced by the number of predictor variables used and the maximum number of basis functions. Another factor is interactivity and minimum observations need to be tested on the data used. The use of models in one area with other areas has different mathematical models because in the analysis of earthquake predictions it is influenced by bedrock conditions, types of faults or others. Predictive research with a nonparametric approach is preferred and has the advantage that this model does not make specific assumptions regarding the underlying functional relationship between the responsive variable and the predictor variable to estimate the general function of the high-dimensional data argument. Prediction results are more effective even though the data set does not provide uniformity of information from each earthquake recording station [6]–[8].

Previous research that uses MARS and conic multivariate adaptive regression splines (CMARS) methods, such as the development of a robust computational method for data prediction problems with the help of convex optimization (convex) in the presence of outliers in the dataset. The results show that the optimal level of process parameters produces the desired response in the application. The research proposes a new approach to deal with outliers in the prediction of ground motion in a systematic and effective manner. The result is that there are no assumptions that must be validated for effective modeling in the presence of outliers [9], [10].

Another study describes the development of a simple approach to predicting the displacement of underground structures caused by earthquakes. The method used is the MARS model approach, to predict the lift displacement of underground structures and evaluate the buoyancy of underground structures in terms of earthquake parameters, structural characteristics, and soil properties [11]. Similar research on ground motion prediction, explains that ground motion prediction equations (GMPEs) are empirical relationships used to determine the response of the ground peak at a certain distance from the earthquake source. Research has correlated the response of the ground peak as a function of the type of earthquake source, local conditions of the location, distance from the source, depth, and magnitude of the earthquake strength. The method used is CMARS on available datasets to obtain new GMPE. In the CMARS model, peak ground acceleration (PGA) and peak ground velocity (PGV) values are used as dependent variables while three other parameters such as magnitude moment (Mw), station location conditions (Vs30) and distance from earthquake source (Rjb) are used as independent variables. This study shows that CMARS can be effectively used to predict PGA and PGV values at various distances from the earthquake source [12], [13].

The main objective of this study is to analyze earthquake predictions using MARS and CMARS involving 4 predictor variables with 16 maximum basis functions. This study will contribute to a mathematical model of predictive analysis of earthquakes that occur in Lombok, West Nusa Tenggara, Indonesia, which has different bedrock characteristics from other regions. This research using the MARS method is the first research conducted in Lombok because earthquake prediction research at the same case study location has never been done. The results of this study will classify areas that have a category prone to earthquake hazards based on the highest PGA value.

2. RESEARCH METHOD

2.1. Multivariate adaptive regression spline (MARS)

The MARS method is a nonparametric regression method that is used to overcome the problem of high-dimensional data, which is used to determine the pattern of the relationship between the response variable and the predictor variable whose regression curve is not known and the previous information is not complete enough [14]. Prediction data mining or called prediction analysis can be solved by two approaches, namely parametric regression and nonparametric regression. These two approaches are commonly used as statistical methods and are widely used as methods for investigating and modeling relationships between variables [10]. 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. The working system of the MARS method is a two-stage algorithm, namely the forward stepwise model and the backward stepwise model [10], [15]. The first stage is the forward stepwise algorithm which is used to combine the basis of function (BF), maximum interaction (MI), and minimum observation (MO) to find the relationship between the respond variable and the predictor variable. Furthermore, the second stage of the Backward Stepwise model is used as a simplification of 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 squares of the least residual. In general, the nonparametric regression model can be presented as in (1) [16]–[18]:

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

where $yi =$ the dependent variable on observation i, f(xi) = vector independent 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 (2) and (3):

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

and

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

As demonstrated (2) and (3) have almost the same function so they can be called the reflected pairs. The goal is the reflected pairs in each variable x_j for each observation $x_{i,j}$ on the knot of the variable. So that a truncated linear function is formed from the base function as in (4):

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

The MARS model starts from (5):

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

where M is the number of basis functions that make up the model function $\beta_m(x)$ is the basis of the function formed by a single element or by multiplying two or more elements contained in r, multiplied by the coefficient β_m. The basic function to m can be explained into the base function as shown in (6):

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

where K_m is the number of truncated linear functions times the base function to m. For $X_{k_j}^{\ m}$ is the input variable related to the truncated function in the base function to m. $\tau_{k_j}^{\ m}$ is the knot variable value $X_{k_j}^{\ m}$. Whereas $S_{k_j}^{\ m}$ is the +/- operator which has a value of 1 or -1. The MARS model is flexible and is used to overcome weaknesses in recursive partition regression by increasing the accuracy of the model. The MARS model is run with two stages of the algorithm, namely forward stepwise and backward stepwise. Furthermore, the algorithm will determine the knot value in a continuous model and drink the generalized cross validation (GCV) value to obtain the best model. GCV measurement can be seen in (7):

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

where: y_i = variabel dependent
x_i = variabel independent
N = the number of observations
$\hat{f}_M(x_i)$ = the estimated value of the dependent variable on the M basis function on x_i
M = maximum number of base functions
$\hat{C}(M)$ = C(M)+d.M
$C(M)$ = trace [$B\ (B^T B)^{-1} B^T$]+1; where B is a matrix of M basis functions
d = value when each base function reaches optimization ($2 \leq d \leq 4$)

2.2. Conic multivariate adaptive regression splines (C-MARS)

C-MARS was developed as an alternative to the Backward Stepwise algorithm for the MARS model. C-MARS with the letter "C" is a of "CONIC", convex or continuous, as shown in (8) [19]–[21]:

$$PRSS = \sum_{i=1}^{N} (y_i - f(x_i))^2 + \sum_{m=1}^{Mmax} \lambda m \sum^2_{\substack{|\alpha|=1 \\ \alpha=(\alpha1,\alpha2)^T}} \sum_{\substack{r<s \\ r,S\in Vm}} \int \beta^2_{m[D^{\alpha}_{r,s}\ \beta_m\ (t^m)]^2}\ dt^m \tag{8}$$

where $((\tilde{x}_i , y_i)\ i = (1,2,3 \dots . N)$ represents the data points used as the predictor p-dimensional vector variable $\tilde{x}_i = (\tilde{x}_{i\,1}, \tilde{x}_{i2}\ , \dots , \tilde{x}_{ip})^T\ (i = 1, 2, \dots , N)$ and response value N $(y_i, y_2, \dots , yN)$. Furthermore M_{max} is the number of BF achieved at the end of the Forward MARS algorithm stage. Where: $V_{(m)}=\{(K_j^m)|j = 1,2, \dots , K_m\}$ is the set of variables associated with $m^{th}\ BF. z^m = (z_{m1}, z_{m2}, \dots , z_{mk_m})^T$ represents the variables that contribute to $m^{th}\ BF$. $\lambda_m(m = 1,2, \dots , M_{max})$ the value is always non-negative and is used as a penalty parameter. $D^{\alpha}_{r,s}\ \beta_m(t^m) \coloneqq \frac{\partial|\alpha|\beta_m}{\partial\alpha1\ t_r^m\ \partial\alpha2\ t_s^m}\ (t^m)$, is the partial derivative for the basis function (BF) to m. For $\alpha = (\alpha_1, \alpha_2)^T$, $|\alpha| := \alpha_1 + \alpha_2$, where $\alpha_1, \alpha_2 \in \{0,1\}$

After simplifying the equation and adding the penalization of λ, for each derived term, using the Tikhonov regularization the equation changes to be as (9):

$$PRSS \approx \left\|y - \beta(\tilde{d})\theta\right\|_2^2 + \lambda\| L\theta\|_2^2, \tag{9}$$

Furthermore, it can be formulated into the conic quadratic problem (CQP) as shown in (10):

$$\begin{aligned} &\min_{t,\theta}\ t, \\ &subject\ to\ \left||\beta(d)\theta - y|\right|_2 \leq t, \\ &||L\theta||_2 \leq \sqrt{\widetilde{M}} \\ &\text{with } t \geq \end{aligned} \tag{10}$$

2.3. Peak ground acceleration (PGA)

PGA calculation is used to determine the maximum ground vibration acceleration that occurs in an area caused by an earthquake. The PGA value can be obtained by empirical calculations using the attenuation function of the Joyner and Boore attenuation equations [22], [23]. The first step is to determine the coordinates of the location of a city in the area where the prediction analysis will be carried out, and the vibration radius which is usually up to 500 km. The second step is to calculate the AVECOS value, which is a value as a correction number because the Longitude coordinates towards the poles will be increasingly different, and to calculate the AVECOS value with the formula as shown in formula (11) [23]:

$$AVECOS = Cos(REDCOM \times Average\ Latitude) \tag{11}$$

Next, calculate the distance from the epicenter with the formula as in (12):

$$R\_epi = \sqrt{x^2 + y^2} \tag{12}$$

The epicenter is a seismic wave that is above the earth's surface and then spreads out in all directions. The next step is to calculate using the Joyner and Boore attenuation function equations, as shown in (13):

$$PGA\ (gal) = 10^{(0{,}71+0{,}23(M-6)-Log(r)-0{,}0027.r)} \tag{13}$$

where M is the magnitude and r is the root of $(R\_epi^2+8^2)$

$$r = \sqrt{R\_epi^2 + 8^2} \tag{14}$$

by giving the value of M and the value of r in (13), the PGA value will be obtained.

2.4. Data set

The data used in this study were taken from the geophysics station Badan Meteorologi, Klimatologi, and Geofisika (BMKG) of Mataram City. Data in the form of an earthquake catalog in Lombok with positions (-4.0636°) south latitude (-13.0636°) south latitude and (111.5798°) east longitude (120.5798°) east longitude for a period of 10 years between 2010 and 2019. Total earthquake data earth has a total of 8,053 records and varies in magnitude from 1.6-9.0 Mw, and a depth of 0-500 km [24], [25]. Magnitude data with a value of less than 4.5 Mw and a depth of more than 300 km, is deleted because the data with this value does not have a damaging impact.

After selecting the data set based on magnitude 4.5 Mw and above and a depth of less than 300 kilometers, the results of data processing obtained a table of earthquake frequencies in Lombok with grouping based on magnitude as shown in Table 1. The results of data processing can be seen in the graph of the earthquake spread in Lombok based on the epicenter distance and magnitude strength as shown in Figure 1.

Table 1. Earthquake frequency in Lombok is based on magnitude

No	Magnitude (Mw)	Frequency
1	4.5-5	283
2	5-6	121
3	6-7	15

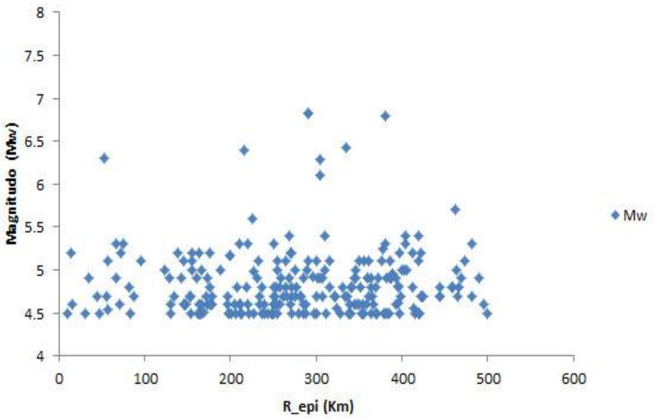

Figure 1. Distribution of earthquakes in Lombok with magnitude 4.5-7

3. RESULTS AND DISCUSSION

3.1. PGA value calculation results

PGA is the maximum ground vibration acceleration that occurs in an area caused by an earthquake. A large PGA value usually has a large impact or risk from an earthquake that occurs. The PGA value can be obtained by one of them performing empirical calculations with the attenuation function. The attenuation function is used to determine the relationship between the intensity of ground vibrations, the magnitude, and the distance of an area from the source of the earthquake. There are several factors that affect the function of attenuation, namely the earthquake mechanism, the distance of the epicenter and local soil conditions. This research is to get the PGA value using the attenuation function of the Joyner and Boore attenuation equations [22], [23]. From the results of data processing for earthquakes that occurred in Lombok from 2010 to 2019 the PGA values were obtained as shown in the example Table 2. After filtering and selecting the appropriate variable, for the respond variable, namely PGA and the predictor variable, the depth (depth), magnitude (Mw), epicenter distance (R-epi), and the addition of the variable temperature of the location of the incident (SUHU), obtained ready data used for the predictive analysis process as shown in Table 3.

Table 2. PGA value for earthquake in Lombok

Time	Long	Lat	Depth	Mw	AVECOS	x	y	R-epi	r	log PGA	PGA(g)
01-01-10	118.64	-9.5	99	3.8	0.987601365	281.1515923	-104.122929	299.812945	299.9196597	-3.082788016	0.000826441
01-01-10	119.06	-8.15	10	3.8	0.989382196	327.8645593	45.99022166	331.074417	331.1710581	-3.210214233	0.000616291
01-04-10	118.66	-11.21	69	4.4	0.98514891	282.6442954	-294.266254	408.020129	408.0985493	-3.370631134	0.00042596
01-04-10	117.23	-8.29	29	4.5	0.989203895	126.5156216	30.42293193	130.122086	130.367777	-2.102163258	0.007903815
01-04-10	120.55	-7.32	527	5.1	0.990408927	492.2961883	138.2820108	511.348659	511.4112353	-3.5865806	0.000259071
01-05-10	119.07	-8.76	277	3.9	0.988594515	328.7028021	-21.8386836	329.427473	329.5245973	-3.180604251	0.000659775
01-07-10	119.07	-7.83	24	4.5	0.989784195	329.0983649	81.57259819	339.057256	339.1516218	-3.081103277	0.000829653
01-07-10	120.46	-7.94	23	3.9	0.989646878	482.0134675	69.34115626	486.975542	487.0412496	-3.775577119	0.000167657
01-09-10	118.43	-7.84	10	4.3	0.98977175	258.6573647	80.46064892	270.882905	271.0010117	-2.845673644	0.001426679
01-12-10	116.35	-8.75	155	3.4	0.988607654	29.70258763	-20.7267343	36.2193488	37.09233381	-1.557433461	0.027705535
13/01/2010	119.05	-9.83	17	4.2	0.987145193	326.025599	-140.817255	355.136862	355.2269563	-3.213618697	0.000611479
13/01/2010	117.1	-7.87	282	4.4	0.989734367	112.2765199	77.12480112	136.213993	136.4487151	-2.161380981	0.006896346
14/01/2010	119.83	-9.61	10	3.5	0.987450218	411.7699143	-116.354371	427.893447	427.9682255	-3.651925735	0.000222882
15/01/2010	118.52	-7.85	18	4.2	0.989759296	268.5591694	79.34869965	280.036147	280.1503946	-2.907797304	0.001236524

Table 3. Data respond variable and the predictor variable earthquake in Lombok

No	Mw	Depth	SUHU (°)	R-epi	PGA(g)
4	4.5	29	27.3	130.1220861	0.007903815
7	4.5	24	25.8	339.0572556	0.000829653
18	4.6	15	27.3	284.6781717	0.001460629
52	4.7	23	27	321.4371625	0.001085476
56	6.1	52	27.3	305.1784169	0.002654803
83	4.6	74	27.2	354.8631328	0.000757626
103	4.9	47	27.2	393.1406573	0.000631921
104	5	25	27.2	402.1835925	0.000615714
105	5	45	27.2	165.6369759	0.006495246
113	4.6	27	25.8	239.3908901	0.002301091
143	5.1	217	27	57.34264291	0.038371768
145	4.8	10	27	264.0445075	0.001990106
152	4.8	60	27	276.5372938	0.001758335
157	4.7	26	27	153.2842779	0.006463695
171	4.6	51	27	205.2507607	0.003317323
186	4.8	109	27	237.4462852	0.002610475
203	4.7	10	27.3	55.42425365	0.032480759
213	4.7	77	25.8	250.9637935	0.002153882

3.2. Results of the development of an analysis model for earthquake prediction in Lombok

Selection of the best MARS model after going through the forward stepwise algorithm and backward stepwise algorithm based on a combination of BF, MI, and MO, the results of training data are obtained. The results of MARS regression based on training data are as shown in Table 4. The results of training data with a total of 16 basis functions BF after going through a process of elimination by applying the Backward Stepwise algorithm, a total of 12 BF are formed, namely BF 1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15, and 16. After doing elimination by eliminating the BF which does not have a contribution to the change in the dependent variable, namely the BF 4, 6, 8, and 12 in order to obtain the best MARS model as in (15):

$$\begin{aligned}Y_{(PGA)} &= -0.0175733 - 0.00211487 * BF1 + 0.0029936 * BF2 + \\ &0.000556472 * BF3 + 0.00172513 * BF5 + 0.000373726 * BF7 + \\ &0.000369563 * BF9 - 0.000160793 * BF10 - 0.000689482 * BF11 + \\ &0.000676173 * BF13 + 0.00329239 * BF14 - \\ &0.00125948 * BF15 + 6.46282e - 05 * BF16\end{aligned} \tag{15}$$

Model PGA_G_=BF1, BF2, BF3, BF5, BF7, BF9, BF10, BF11, BF13, BF14, BF15, BF16

Table 4. Results of training data from MARS

W: 442.00			SQUARED: 0.99723	
MEAN DEP VAR: 0.01402			ADJ R-SQUARED: 0.99715	
UNCENTERED R-SQUARED=R-0 SQUARED: 0.99785				
Parameter	Estimate	S.E.	T-Value	P-Value
Constant	-0.01999	0.00723	-2.76347	0.00597
Basis Function 1	-0.00219	0.00022	-9.93648	0.00000
Basis Function 2	0.00307	0.00022	14.14851	0.00000
Basis Function 3	0.00056	0.00004	14.16816	0.00000
Basis Function 5	0.00180	0.00022	8.17230	0.00000
Basis Function 7	0.00037	0.00001	61.36239	0.00000
Basis Function 9	0.00037	0.0000	35.98398	0.00000
Basis Function 10	-0.00016	0.0000	-28.32933	0.00000
Basis Function 11	-0.00069	0.00009	-7.50372	0.00000
Basis Function 13	0.00067	0.00004	16.66312	0.00000
Basis Function 14	0.00326	0.00032	10.15419	0.00000
Basis Function 15	-0.00129	0.00020	-6.47462	0.00000
Basis Function 16	0.00006	0.00001	5.58802	0.00000
F-STATISTIC=12850.73516			S.E. OF REGRESSION=0.00138	
P-VALUE=0.00000			RESIDUAL SUM OF SQUARES=0.00082	
[MDF, NDF]=[12, 429]			REGRESSION SUM OF SQUARES=0.29560	

Based on the best MARS model obtained, independent variable inference that affects PGA based on the MARS model according to the smallest GCV value in sequence based on the percentage of the contribution is R-epi, Mw, SUHU, and depth. As shown in Table 5 which describes the interactivity of the contribution of independent variables to the dependent variable. It can be seen in Table 5 that the variables that are very influential in the PGA value are the R-epi of 100% and the Mw of 31.08608%, while the SUHU is 5.48525% and the depth amounting to 3.52988%.

Table 5. The interactivity of the independent variable contribution

Variable	Importance	GCV
R-EPI	100.00000	0.00067
MW	31.08608	0.00007
SUHU	5.48525	0.00000
DEPTH	3.52988	0.00000

3.3. Testing and validation

In prediction analysis required statistical analysis test to obtain hypothesis test results and determine the level of significance. The significance level is intended to obtain the significance of the parameters [26]. Hypothesis testing is required to use statistical analysis to determine the significance of the parameters with the suitability of the obtained mathematical model. This research is in testing the analysis of a mathematical model using the partial regression coefficient test. In testing the partial regression coefficient, the following formulation is required:

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 $a_m \neq 0$;
m=1,2,3,5,7,9,10,11,13,14,15,16 (significant model)

- Significant level, α=0.05
- Statistic test: $t_{count}=\frac{\hat{a}_m}{Se(\hat{a}_m)}$ with $Se(\hat{a}_m) = \sqrt{var\ (\hat{a}_m)}$
- 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 4 (results of training data) that the P-value is less than 0.05, or in other words, every m<α or (m<0.05) so that the H0 status is rejected. This means that each coefficient $\alpha_1, \alpha_2, \alpha_3, \alpha_5, \alpha_7, \alpha_9, \alpha_{10}, \alpha_{11,}\ \alpha_{13}, \alpha_{14}, \alpha_{15,}\ \alpha_{16}$ has a significant effect on the mathematical model obtained. Based on the significance level of 5%, the mathematical model in formula (15) is significant, so it can be used in predictive analysis of the PGA value for earthquake data sets in Lombok. 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), temperature at the location of the incident (SUHU), and depth (Depth).

3.4. Potential areas with the highest earthquake danger in Lombok

After going through the testing and validation of the results of the prediction analysis, it can be seen that the areas in Lombok have the highest potential as shown in Table 6. The results of the calculation of the PGA value will be influenced by the magnitude, depth, distance to the location of the incident, and the temperature of the location where the earthquake occurred. In theory, based on a high PGA value, it will have a high impact on earthquake damage, although there are other factors that influence earthquake damage such as the condition of the bedrock of the location. Based on the results of predictive analysis by grouping areas that have the highest earthquake hazard in Lombok, this data can be used by government policy makers to make rules for infrastructure development with special specifications in earthquake-prone areas.

Table 6. Highest earthquake hazard potential areas in Lombok

No	Time	Lat	Long	Depth	Mw	R-epi	PGA(g)	PGV (cm/s)	emp (°)	Regional location
1	22-06-2013	-8.44	116.04	16	5.2	14.42381995	0.183715166	1.832770384	26.7	Malaka, Pemenang
2	09-08-2018	-8.36	116.22	12	6.2	27.39175594	0.16732396	1.814869245	24.9	Genggelang, Gangga
3	05-08-2018	-8.41	116.16	17	5.5	19.22254717	0.166068049	1.826111838	24.9	Tegal Maja, Tanjung
4	31/03/2016	-8.52	115.99	12	4.5	10.60647207	0.160605686	1.838096242	27.5	Senggigi, Meninting
5	06-08-2018	-8.42	116.03	23	5	16.8807379	0.143937973	1.829356187	24.9	Senggigi, Malimbu
6	04-05-15	-8.43	116.03	13	4.6	15.83302429	0.123357614	1.830810815	26.3	Mangsit, Senggigi

4. CONCLUSION

This research has obtained the result of a mathematical model involving a maximum number of basis functions (BF) as many as 16 with 12 basis functions (BF) having a close contribution to the model. The relationship between the predictor variable and the response variable has a contribution to the epicenter distance of 100 percent, magnitude) of 31.1 percent, the temperature of the incident location by 5.5 percent and the depth of 3.5 percent. Based on the highest PGA value, it can be concluded that the areas with the highest level of earthquake hazard in Lombok are Pemenang, Malacca, Genggelang, Tegal Maja, Tanjung, Senggigi, Mangsit. Meninting, and Malimbu. Earthquakes that have occurred in Lombok in the last 10 years that have a high damage impact are in the category of shallow earthquakes because they have a depth of less than 25 Kilometers.

ACKNOWLEDGMENT

Thank you very much to the Deputy for Strengthening Research and Development, Ministry of Research and Technology/National Research and Innovation Agency (DRPM Kemenristek/BRIN) for Fiscal Year 2020/2021. Thank you to the Chancellor of the University of North Sumatra, Dean and Head of Doctoral Program in Computer Science, University of North Sumatra. Thank you to the Chancellor of Bumigora Mataram University, who has provided full support for this research.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan Kaufmann, 2012.

[2] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning,” New York, NY: Springer New York, 2009.

[3] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley and Sons, 2005.

[4] F. Yerlikaya-Özkurt, İ. Batmaz, and G.-W. Weber, “A review and new contribution on conic multivariate adaptive regression splines (CMARS): a powerful tool for predictive data mining,” in *Springer Proceedings in Mathematics and Statistics*, Springer International Publishing, 2014, pp. 695–722.

[5] F. Yerlikaya, “A new contribution to nonlinear robust regression and classification with mars and its applications to data mining for quality control in manufacturing,” Middle East Technical University, 2008.

[6] F. Yerlikaya-Özkurt, A. Askan, and G.-W. Weber, “An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method,” *Engineering Optimization*, vol. 46, no. 12, pp. 1651–1668, Dec. 2014, doi: 10.1080/0305215X.2013.858141.

[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, Jul. 2018, doi: 10.1371/journal.pone.0199004.

[8] G. Asencio-Cortés, F. Martínez-Álvarez, A. Troncoso, and A. Morales-Esteban, “Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks,” *Neural Computing and Applications*, vol. 28, no. 5, pp. 1043–1055, May 2017, doi: 10.1007/s00521-015-2121-7.

[9] F. Yerlikaya-Özkurt, A. Askan, and G.-W. Weber, “A hybrid computational method based on convex optimization for outlier problems: application to earthquake ground motion prediction,” *Informatica*, vol. 27, no. 4, pp. 893–910, Jan. 2016, doi: 10.15388/Informatica.2016.116.

[10] G.-W. Weber, İ. Batmaz, G. Köksal, P. Taylan, and F. Yerlikaya-Özkurt, “CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization,” *Inverse Problems in Science and Engineering*, vol. 20, no. 3, pp. 371–400, Apr. 2012, doi: 10.1080/17415977.2011.624770.

[11] G. Zheng *et al.*, “Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines,” *Computers and Geotechnics*, vol. 113, Sep. 2019, doi: 10.1016/j.compgeo.2019.103099.

[12] R. A. Johnson and D. W. Wichern, *Apllied Multivariate Statistical Analysis*, 6th ed. Pearson, 2007.

[13] I. Peruš and P. Fajfar, “Ground-motion prediction by a non-parametric approach,” *Earthquake Engineering & Structural Dynamics*, vol. 39, no. 12, 2010, doi: 10.1002/eqe.1007.

[14] R. L. Eubank, *Nonparametric regression and spline smoothing*. CRC Press, 1999.

[15] A. Özmen, “Robust conic quadratic programming applied to quality improvement a robustification of CMARS,” Middle East Technical University, 2010.

[16] 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, Mar. 2020, doi: 10.1080/17480930.2019.1577940.

[17] C. Yazıcı, F. Yerlikaya-Özkurt, and İ. Batmaz, “A computational approach to nonparametric regression: bootstrapping CMARS method,” *Machine Learning*, vol. 101, no. 1–3, pp. 211–230, Oct. 2015, doi: 10.1007/s10994-015-5502-3.

[18] J. H. Friedman, “Multivariate adaptive regression splines,” *The Annals of Statistics*, vol. 19, no. 1, Mar. 1991, doi: 10.1214/aos/1176347963.

[19] A. Özmen, İ. Batmaz, and G.-W. Weber, “Precipitation modeling by polyhedral RCMARS and comparison with MARS and CMARS,” *Environmental Modeling and Assessment*, vol. 19, no. 5, pp. 425–435, Oct. 2014, doi: 10.1007/s10666-014-9404-8.

[20] A. Çevik, G.-W. Weber, B. M. Eyüboğlu, and K. K. Oğuz, “Voxel-MARS: a method for early detection of Alzheimer’s disease by classification of structural brain MRI,” *Annals of Operations Research*, vol. 258, no. 1, pp. 31–57, Nov. 2017, doi: 10.1007/s10479-017-2405-7.

[21] J. F. H. Jr., W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, 7th ed., vol. 50, no. 181. Prentice Hall International, 2010.

[22] W. B. Joyner and D. M. Boore, “Prediction of earthquake response spectra.” US Geological Survey, pp. 359–375, 1982, doi: 10.3133/ofr82977.

[23] W. Joyner and D. M. Boore, “Measurement, characterization and prediction of strong ground motion,” in *Earthquake Engineering*

and Soil Dynamics II - Recent Advances in Ground-Motion Evaluation: Proceedings of the Specialty Conference, 1988, pp. 43–102.
[24] 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, Feb. 2018, doi: 10.30630/joiv.2.1.106.
[25] "Indonesia tsunami early warning system," *BMKG*, 2019. https://inatews.bmkg.go.id/new/view_event_eq.php?catid=5953 (accessed Nov. 16, 2019).
[26] D. R. Cox *et al.*, "The role of significance tests," *Scandinavian Journal of Statistics*, vol. 4, no. 2, pp. 49–70, 1977.

BIOGRAPHIES OF AUTHORS

Dadang Priyanto Degree (S3) Doctoral Program in Computer Science at the Universitas Sumatera Utara in 2021, Master Degree (S2) in Computer Science obtained at Universitas Gadjah Mada Yogyakarta in 2006, and Bachelor Degree (S1) in Informatics Engineering obtained at STMIK Akakom Yogyakarta in 1998. Currently works as a permanent lecturer at Bumigora University Mataram, in the Software Engineering study program. Research has been carried out since 2005 in the fields of Multimedia and Data Science. The latest publication is related to Data Mining for Earthquake Predictions in Lombok. He can be contacted by email: dadang.priyanto@universitasbumigora.ac.id.

Muhammad Zarlis is a Professor of Computer Science at the Universitas Sumatera Utara, A Doctoral Degree (S3) in Computer Science was obtained at the University Sains Malaysia, Malaysia and a Masters Degree (S2) in Computer Science was obtained at the University of Indonesia-Sandwich Program with the University of Maryland USA, and a Bachelor Degree (S1) in Physics was obtained at the University of North Sumatra. Research that has been pursued for a long time is related to the fields of Intelligent Systems, Computer Security, and Computation. He can be contacted by email: m.zarlis@usu.ac.id.

Herman Mawengkang is a Professor of Mathematics at the Universitas Sumatera Utara. Masters and Doctoral degrees were obtained at the University of New South Wales Australia, and Bachelor degrees (S1) in Mathematics were obtained at the Universitas Sumatera Utara. The research that has been done so far is related to the fields of Mathematics and Operations Research. He can be contacted by email: hmawengkang@yahoo.com.

Syahril Efendi is a Permanent Lecturer at the Computer Science Doctoral Program at the Universitas Sumatera Utara. A Doctorate Degree (S3) in Mathematics was obtained at the Universitas Sumatera Utara, a Masters Degree (S2) in Computer Science was obtained at the National University of Malaysia, Malaysia, and a Bachelor Degree (S1) in Mathematics was obtained at the Universitas Sumatera Utara. The research he has been pursuing to date is related to the fields of research operations and information science. He can be contacted by email: syahril1@usu.ac.id.