Journal of Theoretical and Applied Information Technology
15th November 2020. Vol.98. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ALTERNATIVE APPROACH TO SEISMIC HAZARDS PREDICTION USING NON PARAMETRIC ADAPTIVE REGRESSION METHOD

[1]DADANG PRIYANTO, [2*]MUHAMMAD ZARLIS, [3]HERMAN MAWENGKANG, [4]SYAHRIL EFENDI

[1] Graduate Program Of Computer Science, Department Of Computer Science, Faculty Of Computer Science And Information Technology, Universitas Sumatera Utara, Medan, Indonesia.

[2,3,4] Department of Computer Science, Faculty of computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia.

E-mail: [1] dadangpriyanto@students.usu.ac.id ; [2*] m.zarlis@usu.ac.id ;
[3] hmawengkang@yahoo.com ; [4] syahril1@usu.ac.id

ABSTRACT

Research with data mining processes to find certain patterns related to mathematical functions such as Correlation, Classification and regression associations, Clustering and others are grouped into two categories namely Descriptive data mining and Predicted data mining. Data mining process Prediction to find out the relationship between variables can be used Parametric and Non-Parametric methods. Many non-parametric methods used one of them is the Multivariate Adaptive Regression Spline (MARS) method. The flexible nature of MARS modeling can be applied to various fields of application including earthquake prediction research. Research on earthquakes contains many parameters that are definitely necessary to get optimal results with cone optimization models difficult to do this research was conducted to complete research on earthquake predictions with uncertain parameters. This study uses a non-parametric method with MARS and to improve its ability to use the CMARS model which is the back of the MARS algorithm. The results of this study after observing the testing of parameters with a combination of basis functions (BF), Maximum Interaction (MI) and Minimum Observation (MO) obtained the results of predictive analysis with a mathematical model that has two basis functions (BF) namely MODEL (PGA) = BF1, BF2, BF3, BF5, BF7, BF9, BF10, BF11, BF13, BF14, BF15, and BF16. The model was obtained from trial and error observations with a combination of basis functions (BF) = 16, MI = 2, and MO = 2. Based on the level of importance of the independent variables on the dependent variable is the Epicenter Distance (R-epi), Magnitude (Mw), Temperature of the incident location (SUHU), and Depth (Depth). The results of the prediction analysis can reveal six areas that have the highest level of earthquake hazard in Lombok, namely the first area of Malacca, North Lombok Regency (KLU), first Genggelang, Ganga (KLU), Tegal Maja, Tanjung, Winner (KLU), Senggigi choice, Malimbu Regency, West Lombok (Lobar), Mataram, as many as Senggigi, Malimbu (Lobar), and the sixth are Mangsit, and Senggigi (Lobar).

Keywords: *Non Parametric, Prediction Analysis, MARS, C-MARS, PGA, Data Mining*

1. INTRODUCTION

Discussion of data mining is becoming very popular nowadays, with the support of abundant data. Data on social media, companies, governments and other organizations, can be accessed easily because it has adopted an online system. The available data can be processed so that it becomes more useful data such as aspects of user behavior, trends or product trends that are most preferred by users, level of user satisfaction, and others so that it can be used as an element of policy making for the future progress of the organization. Apart from that the data related to disaster problems can also be accessed easily by the users. Disaster data such as earthquake data that occurred in Indonesia can be accessed easily on the site of the Meteorological, Climatology and Geophysics Agency (BMKG) www.bmkg.go.id to obtain information on the latest earthquake data and recorded in the past. Earthquake data is one of the big data categories because data continues to grow recorded from time to time. Research related to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

earthquakes has been carried out by many experts, who use both parametric and non-parametric methods. Research that uses non-parametric methods is still a rare category due to the limitations of existing data.

Research with non-parametric methods is still wide open the opportunity to produce promising outputs, because this research usually uses earthquake precursor data (earthquake causes) with variables containing uncertain parameter elements. Non-parametric data are relatively more difficult to obtain so that many current studies are parametric using times series data. Data mining according to Turban is a process that uses statistical techniques, mathematics, artificial intelligence and machine learning to extract and identify useful information and related knowledge from large databases. [1] The occurrence of the earthquake in Lombok in 2018 became a special concern for research, because the earthquake had a devastating impact and caused a significant number of fatalities, namely 560 deaths and 1,469 injuries. The island of Lombok, which is in the area of the Three Active Plates meeting in Indonesia, makes the active category of areas affected by seismic activities. Generally large earthquakes will occur in a 100 year cycle. [2] [18] Recent research related to earthquake data using non-parametric methods such as, Yerlikaya, has developed a robust computational method for data prediction problems with the aid of convex optimization with Outliers. [3] [4] Other studies on earthquakes that utilize machine learning techniques and use Regression functions such as. [5] Predictions with Support Vector Regressor And Hybrid Neural Networks models. [6] and Predictions with Mathematical Innovation models. [7] Earthquake prediction with Neural Network based on automatic grouping in Indonesia. [8] This article is organized as follows, part 1 Introduction, which discusses the background to this research. Section 2. DATA SET, which provides an overview of the data set used in the study. Section 3 METHODOLOGY, namely the methods used in this study such as the MARS method and the use of the Joyner Boore Attenuation function. Section 4 RESULTS AND DISCUSSION, discussion of the results obtained. Section 5. is the last to discuss CONCLUSIONS.

2. DATA SET

The data mining process requires a complete data set, to obtain data characteristics in the prediction analysis process. [17] The data set used in this study is the earthquake data that occurred in Lombok in the span of time between 2010 and 2019. The data used in the form of catalog data is a series of data with coordinate position (-4.0636o) LS - (-13.0636 o) LS and (111.5798o) East BT - (120.5798o) East BT. Data obtained from the Geophysics Station (BMKG) of the city of Mataram with a total data recording of 8.053 records. The data cannot be used directly in this study, it requires data processing or preprocessing data to get the magnitude, distance location of the epicenter, depth of the epicenter, and the value of the maximum ground motion acceleration or (PGA). Data that has no significant contribution or value in the study will be deleted or deleted, such as magnitude magnitude data with an intensity of less than 4.5 Mw. Magnitude with such magnitude does not have an impact or may not be felt at all. This selection will be continued with other data sorting such as the distance of the location of the incident above 500 Km, and depths of more than 300 Km will be eliminated. The selection results will look like in Figure 1, the following Magnitude Magnitude Distribution Earthquake Distribution Chart in Lombok:

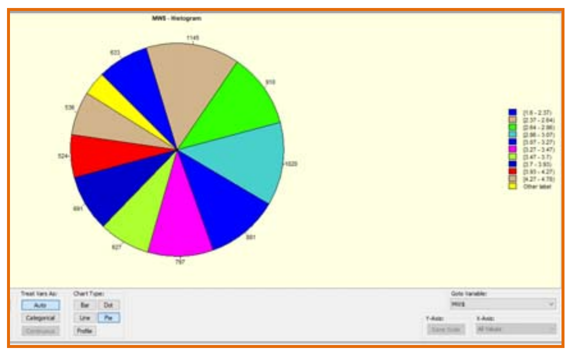

Figure 1. Graphic of Magnitude Magnitude Distribution in Lombok.

Seen in the graph 1 the Yellow color indicates for magnitude of earthquakes with a magnitude of more than 4.78 Mw, and this data will be used in this study, while Magnitude data below 4.5 Mw will be deleted. The results of the processing and selection of data obtained table Frequency of earthquake data in Lombok by grouping based on magnitude Magnitude as shown in table 1 below:

Table 1. Frequency of Earthquakes in Lombok based on magnitude

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

The pattern of earthquake spread in Lombok based on magnitude and epicenter distance gives a wide spread at varying depths as shown in Figure 2 below:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

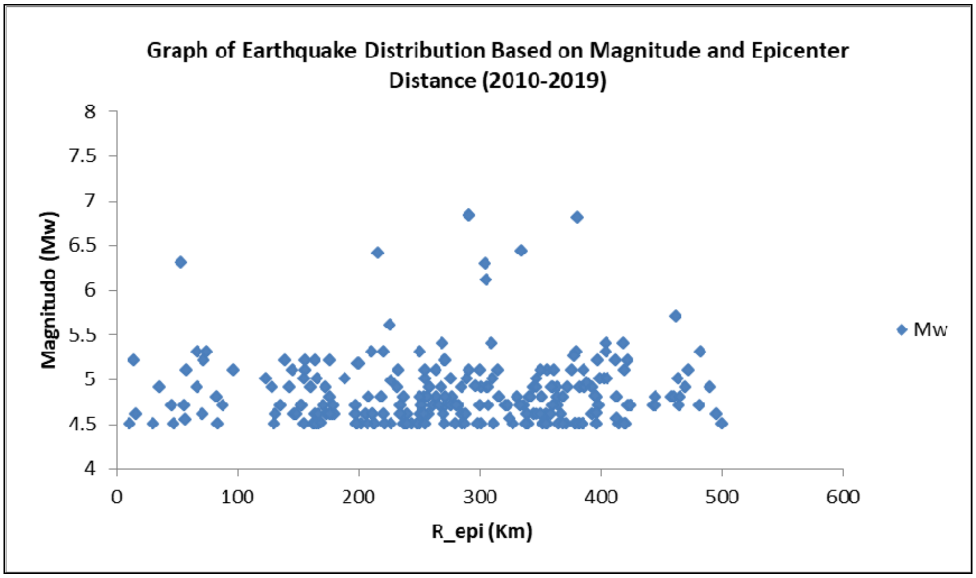

Figure 2. Graph of Distribution of Earthquakes in Lombok with Magnitude 4.5 - 7 from 2010 to 2019.

Peak Ground Acceleration (PGA) value data is obtained using empirical calculations using the Joyner and Boore Attenuation function equation. The Joyner and Boore Attenuation function as shown in equation (8) follows: [9], [10]

$$PGA\ (gal) = 10^{(0,71+0,23(M-6)-Log(r)-0,0027.r)} \qquad (1)$$

Furthermore, from the results of preprocessing data and calculations with the attenuation function as well as adding location temperature data, the data obtained in tabular form as shown in table 2 below:

Table 2. Dependent and independent variable data with the addition of the temperature 'SUHU' earthquake in Lombok

No	Mw	Depth	SUHU (^{o})	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
241	4.6	51	27.3	177.9184569	0.004533934
246	4.8	10	27.3	458.5500965	0.000342195
264	5.6	238	27.3	225.4855935	0.004522868
270	5.1	14	27.3	155.6102057	0.007756867
271	5.1	17	25.8	472.856902	0.000355887
272	4.5	67	25.8	238.9806738	0.002191718
277	5.2	15	27.3	270.4136053	0.002308582
290	4.5	11	27	239.1004258	0.002188993
291	4.7	106	26.9	251.9374108	0.002132626
295	4.5	37	26.3	356.0160844	0.000711107
300	5	86	26.3	289.1165369	0.001729209

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2 above gives an overview of the data from preprocessing, initial data processing before the next process is done for prediction analysis. Data displayed partially or not all of them due to quite a lot of data.

3. METHODOLOGY

3.1 Variate and Multivariate

The basic understanding of Multivariate analysis is variate, which is a linear combination of variables with empirically determined weights. The variables are determined by the researcher, while the weight is determined by the Multivariate technique, to meet certain objectives. A Variate of the variable weight "n" is $(x_1, x_2, \ldots, x_n)$ which can be expressed mathematically as in equation below:

$$\text{Variate Value} = w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n$$

Where xi is the observed variable and wi is the weight determined by the Multivariate technique. The result is a single value that represents the combination of the entire set of variables that best achieves a particular goal in the multivariate analysis. In multiple regression, the variate is determined in a way that maximizes the correlation between several independent variables and a single dependent variable.

3.2. Multivariate Analysis

The definition of Multivariate Analysis is still not an agreement among experts, some say that the use of the term Multivariate Analysis is only to test the relationship between or between more than two variables. Other experts say that Multivariate Analysis is based on all statistical techniques that simultaneously analyze multiple measurements on the individual or object under investigation.

Multivariate analysis is the development of Univariate Analysis (single distribution variable analysis) and Bivariate Analysis (cross classification, correlation, analysis of variance, and simple regression used to analyze two variables). For example, in the case of Multivariate, namely solving the problem of simple regression using a single predictor variable, it is expanded by using several predictor variables. There are still other Multivariate techniques for solving Multivariate problems such as Factor Analysis (which identifies the structure underlying a set of variables) and Discriminant Analysis (which differentiates between groups based on a set of variables).

Some experts state that the purpose of Multivariate analysis is to measure, explain, and predict the level of the relationship between variables (weight of the combination of variables). Thus, the Multivariate character lies in many variations (various combinations of variables), and not only in the number of variables or observations. So it can be concluded that understanding multivariate analysis requires understanding multivariable techniques and multivariate techniques so that understanding multivariable techniques is the first step in understanding multivariate analysis. (Hair et al, 2014)

3.3 Regression Analysis

The word regression was first introduced by Francis Galton, namely regression as the name of a common process for predicting one variable using another. According to its development Regression according to Levin and Rubin (1998: 648), is used in determining the nature and strength of the relationship between two variables and predicting the value of an unknown variable based on past observations of these variables and other variables. According to Gujarati, although regression analysis is related to the dependence of one variable on other variables, it does not necessarily imply a cause and effect relationship (causation). This has also been conveyed by Kendal and Stuart in their book entitled "The Advanced Statistics" published in 1961, saying that, however strong and suggestive a statistical relationship may never be able to establish a causal connection, the idea of cause and effect must comes from outside of statistics that comes from other theories. By definition Linear regression is a statistical method used to model the relationship between the dependent variable (dependent / response ('Y')) and one or more independent variables (independent / predictor ('X').

If the number of independent variables is one, it is called Simple Linear Regression, and if there are more than one independent variable, it is called Multiple Linear Regression. In general, regression analysis has 3 benefits, namely the first for the purpose of description of the data or cases being studied, the second for control purposes, and the third for predictive purposes. Regression can describe data phenomena through the formation of a numerical relationship model. In the case of prediction, the concept of regression should only be carried out within the data range of the independent variables used to form the regression model.

In regression, an estimating equation can be developed, which is a mathematical formula that connects known variables with unknown variables. Once the relationship pattern is known, correlation analysis can be applied, namely to determine the degree to which these variables are related. This means that correlation analysis reveals how true the estimation equation actually describes the relationship. Levin and Rubin explained that regression is related to correlation, but the two have differences. According to Gujarati (2009: 20) correlation analysis aims to measure the strength (strength) or degree (degree) of the linear relationship (linear association) between two variables. These variables are measured using correlation coefficient, whereas in regression does not measure variables.

The fundamental difference between regression and correlation is that in the regression there is an asymmetry (relationship) related to the treatment of the dependent variable / dependent and independent variables. The dependent variable is assumed to be statistical, random or stochastic, that is, it has a probability distribution and the independent variable / predictor is assumed to have a fixed value. Meanwhile in correlation the two variables are treated symmetrically, that is, there is no difference between the independent variable and the dependent variable.

3.4. Multivariate Adaptive Regression Spline (MARS)

MARS is one of the methods to overcome the problem of high-dimensional data, which is a nonparametric regression method used to determine the pattern of relationships between dependent variables and independent variables whose regression curve is unknown or that there is no complete past information. Nonparametric regression is one of the Regression analysis methods or can also be referred to as Prediction analysis which is grouped in two approaches, namely Parametric and Nonparametric Regression. Both are commonly used as statistical methods that are widely used to investigate and model relationships between variables. [11] The MARS model is used to overcome the weaknesses of Recursive Partitioning Regression (RPR) which is to produce a continuous model on the knots and can identify the existence of linear and additive functions. The MARS model works with two stages of the algorithm that must be passed namely the Forward Stepwise Model and the Backward Stepwise model. [12] [13] Forward Stepwise algorithm is used for a combination of basis functions (BF), maximum interactions (MI), and minimum observations (MO) to get the relationship between dependent variables and independent variables with the maximum number of base functions that have been set. Furthermore, the Backward Stepwise model is used to simplify the function basis (BF) of the BF results given at the Forward Stepwise stage. The basis function (BF) which does not contribute or makes a small contribution to the dependent variable in the backward stepwise model will be removed, and this deletion will result in a decrease in the sum of the least remaining squares. In general, the Nonparametric Regression model can be presented as in the following equation (2): [14], [15]

$$yi = f(xi) + \mathcal{E}i \qquad (2)$$

Where yi = the dependent variable on observation i,
$f(xi)$ = vector Independent variable function.
$\mathcal{E}i$ = is a free error i.

The MARS model is very flexible and to build the model is very dependent on the independent variables of the data entered and the basic functions can be explained in equations (2) and (3) below:

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

and

$$(x-r)_+ = \begin{cases} r-x, & if\ x < r, \\ 0, & otherwise \end{cases} \qquad (4)$$

In equations (3) and (4) similar functions are seen and are called reflected pairs. The goal is to reflect pairs for each variable x_j on each observation $x_{i,j}$ on the Knots of the variable. Furthermore, a linear function is formed that is truncated from the basis function as shown in equation (5) below

$$r = \left\{ (x_j - r)_+, (r - x_j)_+ \mid r \in \{x_{1j}, x_{2j}, \dots, x_{Nj}\}, j = 1,2,\dots,p \right\} \qquad (5)$$

The MARS model is built from the following equation (6): [16]

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m \beta_m(x), \qquad (6)$$

Where M is the number of basis functions that make up the function model. $\beta m\ (x)$ 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 βm. The basic function to m can be expressed in the basis function as in equation (7) follows:

$$\beta_m(x^m) = \prod_{j=1}^{K_m} \left[S_{k_j}{}^m \left(X_{k_j}{}^m - \tau_{k_j}{}^m \right) \right] +, \qquad (7)$$

Where Km is the number of truncated linear functions multiplied in the basis function to m. For $X_{k_j}^{m}$ is the input variable that is related to the truncated function in the basis function to m. $\tau_{k_j}^{m}$ is the knot value of the variable $X_{k_j}^{m}$. Whereas $S_{k_j}^{m}$ is the +/- operator, it usually takes the value 1 or -1.

MARS is used to overcome weaknesses in recursive partition regression by increasing the accuracy of the model. MARS model is run with two stages of the algorithm, namely Forward Stepwise and Backward Stepwise. The Forward Stepwise algorithm is used to find the basis function (BF) at each step by minimizing some criteria that are less feasible or inappropriate on each basis function (BF) selected. The process will stop when the M_{max} user specified value is reached or fulfilled. The next step is Backward Stepwise, the purpose of this step is to avoid or prevent overfting by reducing the complexity of the model without reducing the value of the data, as well as removing the base function model (BF) which has no contribution in forming the best model. This omission will result in a decrease in the number of the least remaining squares at each stage. This algorithm will then determine the value of knots in the continuous model and drink the value of Generalized Cross Validation (GCV) to obtain the best model. GCV measurements can be seen in the following equation (7):

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

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 ($2 \leq d \leq 4$)

3.5. Application of the MARS Method on the Problem of Prediction of Land Motion

One method of non-parametric approach to deal with data problems that have different and complex characteristics such as earthquake data can use an approach with the MARS model. Prediction analysis with MARS requires trial and error to get the pattern of the relationship between the dependent variable and the independent variable in the data set used. This study uses SPM 8 software using four independent variables namely Magnitude ('MW'), Depth ('Depth'), Epicenter Distance ('R_epi), and temperature of earthquake location (' SUHU '). While the dependent variable is defined as Peak Ground Acceleration ('PGA'). The base function (BF) is assigned a total of sixteen values (BF). Furthermore, it is necessary to do a combination of input with maximum interaction (MI) and also Minimum Observation (MO) to get the best MARS model. MI and MO input values are required by trial and error to enter values from 0 to 3 in combination with a BF value of sixteen. The best MARS model will be obtained by a combination of BF, MI, and MO by selecting the smallest Generalized Cross Validation (GCV) value. Trials were conducted to obtain a combination table of BF, MI, and MO as follows:

Table 3. Combination Results of BF, MI, and MO data input based on the smallest GCV

No	BF	MI	MO	GCV	MSE	R^2
1	8	1	0	0.00005	0.00004	
2	8	1	1	0.00005	0.00004	
3	8	1	2	0.00005	0.00004	
4	8	1	3	0.00005	0.00004	
5	8	2	0	0.00002	0.00001	
6	8	2	1	0.00002	0.00001	
7	8	2	2	0.00002	0.00001	
8	8	2	3	0.00002	0.00001	
9	8	3	0	0.00002	0.00001	
10	8	3	1	0.00002	0.00001	
11	8	3	2	0.00002	0.00001	
12	8	3	3	0.00002	0.00001	
13	12	1	0	0.00004	0.00004	
14	12	1	1	0.00004	0.00004	
15	12	1	2	0.00004	0.00004	
16	12	1	3	0.00004	0.00004	
17	12	2	0	0.00000	0.00000	0.99664
18	12	2	1	0.00000	0.00000	0.99598
19	12	2	2	0.00000	0.00000	0.99546
20	12	2	3	0.00000	0.00000	0.99546
21	12	3	0	0.00000	0.00000	0.99664
22	12	3	1	0.00000	0.00000	0.99598
23	12	3	2	0.00000	0.00000	0.99555
24	12	3	3	0.00000	0.00000	0.99662
25	16	1	0	0.00004	0.00004	
26	16	1	1	0.00004	0.00004	
27	16	1	2	0.00004	0.00004	
28	16	1	3	0.00004	0.00004	
29	16	2	0	0.00000	0.00000	0.9972
30	16	2	1	0.00000	0.00000	0.9971
31**	16	2	2	0.00000	0.00000	0.99723
32	16	2	3	0.00000	0.00000	0.99722
33	16	3	0	0.00000	0.00000	0.99722
34	16	3	1	0.00000	0.00000	0.9972
35	16	3	2	0.00000	0.00000	0.99717
36	16	3	3	0.00000	0.00000	0.99717

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Provisions in the selection of the best MARS model is by choosing the smallest GCV, if the GCV has the same value then the selection is based on the smallest MSE value. If the MSE value is also the same then the next election is based on the largest R^2 value. Seen in table 3 number 31 ** is the result of a trial and error combination of BF, MI, and MO with a minimum GCV value of 0.00000, minimum MSE of 0.00000, and R^2 the largest (Square) of 0.99723 , so it can be concluded number 31 ** in accordance with the provisions is the best model of MARS.

4. RESULTS AND DISCUSSION

Results of training data with a maximum number of base functions (BF) of 16 after going through the process of elimination by applying the Backward Stepwise algorithm will form a number of 12 basis functions (BF), namely BF 1, 2,3,5,7,9,10,11,13, 14.15 and 16. After eliminating by eliminating the basis function (BF) which has no contribution in changing the dependent variable, namely the basis function (BF) 4, 6, 8, and 12 so that the best MARS model is obtained as in equation (8) below :

$$Y_{(PGA)} = -0.0175733 - 0.00211487 * BF1 + 0.0029936 * BF2 + 0.000556472 * BF3 + 0.00172513 * BF5 + 0.000373726 * BF7 + 0.000369563 * BF9 - 0.000160793 * BF10 - 0.000689482 * BF11 + 0.000676173 * BF13 + 0.00329239 * BF14 - 0.00125948 * BF15 + 6.46282e\text{-}05 * BF16 \quad (8)$$

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

Where Y (PGA) is the result of PGA prediction from the MARS model with the contribution of each basis function (BF) is as follows:

BF1 = max(0, R_EPI - 64.642);

BF2 = max(0, 64.642 - R_EPI);

BF3 = max(0, MW - 4.7) * BF2;

BF5 = max(0, R_EPI - 31.8373);

BF6 = max(0, 31.8373 - R_EPI);

BF7 = max(0, R_EPI - 153.284);

BF8 = max(0, 153.284 - R_EPI);

BF9 = max(0, MW - 5.8) * BF8;

BF10 = max(0, 5.8 - MW) * BF8;

BF11 = max(0, SUHU__O_ - 27.4) * BF2;

BF13 = max(0, SUHU__O_ - 24.2) * BF6;

BF14 = max(0, MW - 5.1) * BF6;

BF15 = max(0, 5.1 - MW) * BF6;

BF16 = max(0, DEPTH - 1) * BF6;

4.1. The Level of Interactive Interdependence of Independent Variables

Variables 'R_epi', 'Mw', 'SUHU', and 'Depth' are independent variables that influence each other because they have a correlation to the dependent variable to get the best model. Based on the ANOVA test results can be obtained independent table variables that affect the dependent variable as shown in table 4. below:

Table 4. Percentage of Interest of independent variables

Variable	Importance	GCV
R_EPI	100.00000	0.00067
MW	31.08608	0.00007
SUHU	5.48525	0.00000
DEPTH	3.52988	0.00000

It can be seen in table 4 that the variables that are very influential in the PGA value are the epicenter distance (R_epi) by 100% and the Magnitude (Mw) by 31.08608%, while the location temperature (SUHU) is 5.48525% and depth (Depth) amounted to 3,52988%.

Furthermore, plots can be made in the three dimensions of the dependent variable relationship with the independent variable to determine the relationship pattern and the contribution of each independent variable. The three dimensions plot as shown in Figure 3,4 and 5 below:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

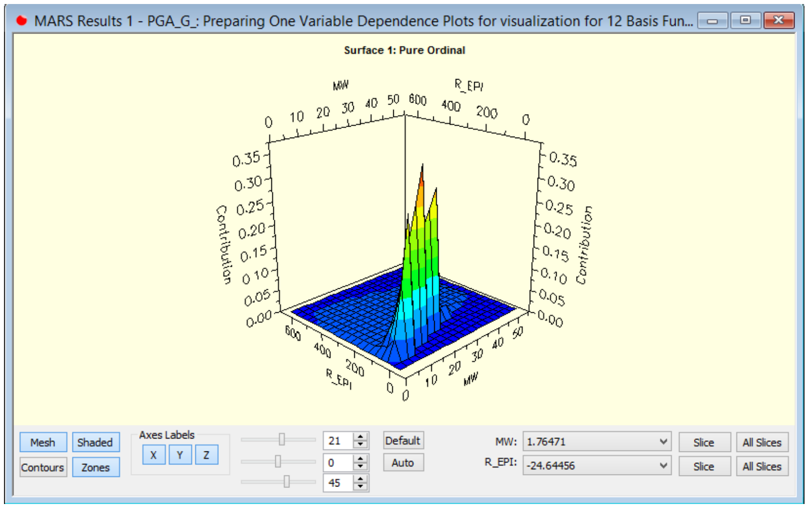

Figure 3 Plot of the contribution of variable magnitude and epicenter distance to PGA

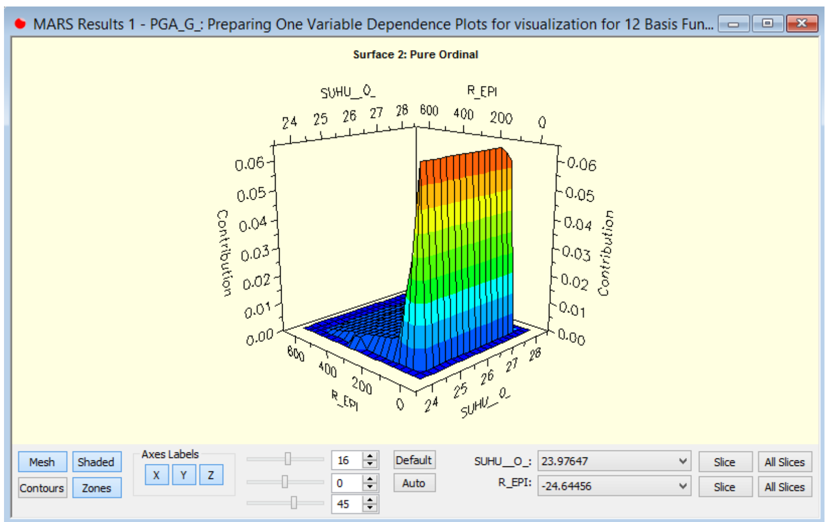

Figure 4. Plot of Contribution of Temperature and Epicenter Distance variables to PGA

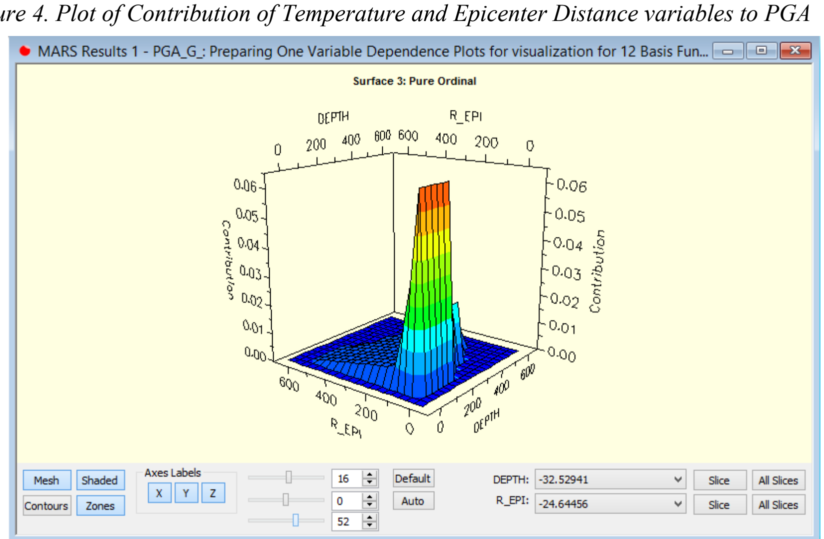

Figure 5. Plot of Contribution of Depth and Distance of Epicenter variables to PGA

4.2 Validity Test

Hypothesis testing is required to use statistical analysis to determine statistical significance. A significance test is needed to determine the significance of the parameters to get the suitability or suitability of the model obtained. The following is testing the model with the Partial Regression Coefficient. The test is carried out with the following formula provisions:

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, $\alpha = 0{,}05$
- Statistic test : $t_{hitung} = \frac{\hat{a}_m}{Se(\hat{a}_m)}$ with $Se(\hat{a}_m) = \sqrt{var\ (\hat{a}_m)}$
- Critical area : Refuse H_0 Jika $t > t_{(\frac{\alpha}{2},61)}$ or P-Value < α

Based on the results of testing the data as shown in Table 5.The results of the P-Value values are obtained as follows:

Table 5. Results of training data

```
=============================================================================
 MARS Regression: Training Data
=============================================================================
 W: 442.00                                  R-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      0.00001     35.98398      0.00000
 Basis Function 10       |    -0.00016      0.00001    -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
 ----------------------------------------------------------------------------
```

P-value is the smallest opportunity value in a hypothesis test so that the observed statistical test value still has meaning. The p-value is used to express a probability magnitude, to reject the null hypothesis (H0) with the actual condition that H0 is true, to obtain a test statistic (statistical test). P-value as an approach in drawing conclusions on the status of 'reject' or 'fail to reject' the given hypothesis. The decisions that can be taken from the P-Value table 5 for each basis function (BF) are:

BF1 = max(0, R_EPI - 64.642);
BF2 = max(0, 64.642 - R_EPI);
BF3 = max(0, MW - 4.7) * BF2;
BF5 = max(0, R_EPI - 31.8373);
BF6 = max(0, 31.8373 - R_EPI);
BF7 = max(0, R_EPI - 153.284);
BF8 = max(0, 153.284 - R_EPI);
BF9 = max(0, MW - 5.8) * BF8;
BF10 = max(0, 5.8 - MW) * BF8;
BF11 = max(0, SUHU - 27.4) * BF2;
BF13 = max(0, SUHU - 24.2) * BF6;
BF14 = max(0, MW - 5.1) * BF6;
BF15 = max(0, 5.1 - MW) * BF6;
BF16 = max(0, DEPTH - 1) * BF6;

It can be seen that the P-value at each m < α or (m < 0,05) thus H0 is rejected, which means that the coefficients α_1, α_2, α_3, α_5, α_7, α_8, α_9, α_{11} has a significant effect on the model.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The conclusion is that based on the 5% significance level the model is significant, so it can be used or used in predicting the PGA value. Based on the significance test of the MARS model by using the Partial Regression Coefficient (t test) on the best model, the independent variables that affect the PGA value are Epicenter Distance (R_epi), Magnitude (Mw), Temperature at the scene (SUHU), and depth (Depth).

4.3 Potential Areas with the Highest Earthquake Hazards in Lombok

Results of Analysis Predictions and empirical calculations with the Joyner-Boore attenuation function obtained areas that have the highest potential for earthquake hazard in Lombok as shown in table 6 below.

Table 6. The Highest Potential Areas of Earthquake Hazards in Lombok

No	Lat	Long	Depth	Mw	R-epi	PGA(g)	SUHU (^{o})	Area location
1	-8.44	116.04	16	5.2	14.42381995	0.183715166	26.7	Malaka, Pemenang
2	-8.36	116.22	12	6.2	27.39175594	0.16732396	24.9	Genggelang, Gangga KLU
3	-8.41	116.16	17	5.5	19.22254717	0.166068049	24.9	Tegal Maja, Tanjung, Pemenang KLU
4	-8.52	115.99	12	4.5	10.60647207	0.160605686	27.5	Senggigi, Meninting, Mataram.
5	-8.42	116.03	23	5	16.8807379	0.143937973	24.9	Senggigi, Malimbu
6	-8.43	116.03	13	4.6	15.83302429	0.123357614	26.3	Mangsit, Senggigi

There are six author ranking areas based on areas with high potential for earthquake hazard in Lombok based on the results of Prediction Analysis with the MARS-PGA model. PGA value calculation is influenced by the magnitude, depth, distance of the location of the incident, and the temperature of the location of the earthquake. In theory, a high magnitude value may not necessarily have a high damage effect, but a tendency for a high magnitude of a PGA value is also high and a damage impact is also high. The impact of damage is influenced by other factors such as the location of bedrock conditions. Six areas in Lombok that have the potential to have high earthquake hazards are the first region of Malacca, North Lombok Regency Winner (KLU), second Genggelang, Ganges (KLU), Third Tegal Maja, Tanjung, Winner (KLU), fourth Senggigi, Malimbu, West Lombok Regency (Lobar), Mataram, the five Senggigi, Malimbu (Lobar), and the sixth is Mangsit, and Senggigi (Lobar). These regions are presented based on the coordinate points issued by BMKG Mataram. The condition of these areas can be used as a solution in developing policies, especially in the construction of infrastructure or building structures that consider aspects of earthquake hazards.

5. CONCLUSION

This research has provided the results of empirical calculations with the Joyner-Boore attenuation function and based on prediction analysis with MARS the following conclusions are obtained:

1. A mathematical model can be obtained by considering the combination of BF, MI, and MO that can be used in the calculation of prediction analysis of PGA values.
2. Contributions of the influence of independent variables on the PGA value are the epicenter distance (R_epi) by 100% and the Magnitude Magnitude (Mw) by 31.08608%, while the location temperature (SUHU) is 5.48525% and the depth (Depth) is 3.52988% .
3. It was concluded that the areas with the highest levels of earthquake hazard in Lombok were Malacca, Pemenang, Genggelang, Tegal Maja, Tanjung, which were included in North Lombok District. West Lombok regencies such as Senggigi, Mangsit. Meninting and Malimbu. and part of Mataram City.
4. Data shows that areas with a high level of danger of earthquake damage have an earthquake epicenter depth of less than 24 Km so that the earthquake events in that area are included in the shallow earthquake category.

REFERENCES :

[1] Turban, Efram, Aronson, Jay E, and Peng-Liang, Ting , “Decision Support Systems and Intelligent Systems”, Pearson, 2005.
[2] Teti Zubaidah, Monika Korte, Mioara Mandea, Mohamed Hamoudi, “New insights into regional tectonics of the

Sunda–Banda Arcs region from integrated magnetic and gravity modelling"2014, Journal of Asian Earth Sciences 80 (2014) 172–184

[3] F. Yerlikaya, "A New Contribution to Nonlinear Robust Regression and Classification with MARS and Its Application to Data Mining for Quality Control in Manufacturing", MSc., Middle East Technical University, 2008.

[4] Fatma Yerlikaya-Özkurt, Aysegul Askan, GerhardWilhelm Weber, " A Hybrid Computational Method Based on Convex Optimization for Outlier Problems: Application to Earthquake Ground Motion Prediction" Informatica, 2016, Vol. 27, No. 4, 893–910

[5] Panakkat, A. and Adeli, H. "Neural Network Models For Earthquake Magnitude Prediction Using Multiple Seismicity Indicators", 2007, vol. 17, no. 1, pp. 13–33

[6] K. M. Asim, A. Idris, and T. Iqbal, , "Earthquake prediction model using support vector regressor and hybrid neural networks", 2018, pp. 1–22.

[7] Kannan, A., "Innovative Mathematical Model For Earthquake Prediction", 2015, https://www.researchgate.net/publication/280764638.

[8] Shodiq, M.N, Kusuma, D.H., Rifqi, M.G, "Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia", 2018,Vol 2, pp.37-43.

[9] Joyner, W.B. and Boore, D.M,"Prediction of earthquake response sprectra.", *Proceedings 51st Annual Convention, Structural Enggineers Association of California (Also U.S. Geological Survey Open File Report* 1982, 82-977).

[10] Joyner, W.B. and Boore, D.M,"Measurement, Characterization and Prediction of Strong Ground Motion.", *Proceedings of Conference on Earthquake Enggineering and soil Dynamics, II. Park City, Utah, ASCE, Washington, DC.* 1988.

[11] Weber, G.W., ˙I. 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,* 20 (3) 2012: 371–400.

[12] Yerlikaya, F., Batmaz, I., Weber, G.W, "A Review and New Contribution on Conic Multivariate Adaptive Regression Splines (CMARS): A Powerful Tool for Predictive Data Mining*", In book: Modeling, Dynamics, Optimization and Bioeconomics I Edition: Springer Proceedings in Mathematics & Statistics}*, Volume 73, 2014 Chapter: 38 Publisher: Springer Verlag

[13] Yerlikaya, F., Askan, A., Weber, G.W., "An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression method", *Engineering Optimization,* Vol. 46, No. 12, 2014, pp 1651–1668.

[14] Yazici C, Yerlikaya, F.O, Batmaz I,"A computational approach to nonparametric regression: bootstrapping CMARS method" 2015 , https://www.researchgate.net/publication/277907238.

[15] Friedman, J.H., "Multivariate Adaptive Regression Spline (With Discussion)", The Annals of Statistics, Vol. 19, 1991, pp. 1-141.

[16] Ozmen, A., Batmaz, I., Weber, G.-W., "Precipitation modeling by polyhedral RCMARS and comparison with MARS and CMARS". *Environ. Model. Assess*. 19 (5), 2014, pp 425–435.

[17] Han, J. Kamber, M. Pei, J., "Data mining : concepts and techniques", Morgan Kaufmann, 225Wyman Street,Waltham, MA 02451 USA, 2012.

[18] Teti Zubaidah, Bulkis Kanata, Paniran, Misbahuddin, Rosmaliati, Made Sutha Yadnya, Susilawati Riskia, "Earth Magnetic Fields Evolution over Nusa Tenggara Region from Declination and Inclination Changes on Lombok Geomagnetic Observatory" Electrical Engineering Department University of Mataram, Mei, 2018, pp 85-91