

# Analysis of Earthquake Hazards Prediction with Multivariate Adaptive Regression Splines

*By* dadang Prianto

# Analysis of Earthquake Hazards Prediction with Multivariate Adaptive Regression Splines

Dadang Priyanto<sup>1</sup>, Muhammad Zarlis<sup>2\*</sup>, Herman Mawengkang<sup>3</sup>, Syahril Efendi<sup>4</sup>

<sup>1</sup> Graduate Program Of Computer Science, Department Of Computer Science, Faculty Of Computer Science And Information Technology, Universitas Sumatera Utara, Medan, Indonesia.

<sup>2,3,4</sup> Department of Computer Science, Faculty of computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia.

---

## Article Info

## ABSTRACT

---

### Keywords:

Non Parametric  
Prediction Analysis  
MARS, C-MARS  
PGA  
Data Mining

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 non-parametric methods with MARS and 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. Han (2012) 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, 2, 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 Response 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.[8]

Previous research that uses MARS and 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 valid for effective modeling in the presence of outliers [9,10]. Another study describes the development of a simple approach 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 Conic Multivariate Adaptive Regression Splines (CMARS) on available datasets to obtain new GMPE. In the CMARS model, PGA and 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 [15]. 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 [15,16,24,25]. 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 Response 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 equation (1) below : [17,18,19]

$$y_i = f(x_i) + \epsilon_i \quad (1)$$

Where  $y_i$  = the dependent variable on observation  $i$   
 $f(x_i)$  = vector Independent variable function.  
 $\epsilon_i$  = 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) as follows:

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

and

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

Equations (2) and (3), as shown, 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 equation (4) below :

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

The MARS model starts from equation (5) below:

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

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  $\beta_m$ . The basic function to  $m$  can be explained into the base function as shown in equation (6) below:

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

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  $i$  and 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 equation (7) below:

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

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 ( $2 \leq d \leq 4$ )

## 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 equation (8) below [20,21]:

$$PRSS = \sum_{i=1}^N (y_i - f(x_i))^2 + \sum_{m=1}^{Mmax} \lambda m \sum_{|\alpha|=1}^2 \sum_{\alpha=(\alpha1, \alpha2)^T} \sum_{r, S \in V^m} \int \beta_m^{2} [D_{r,s}^{\alpha} \beta_m(t^m)]^2 dt^m \quad (8)$$

Where  $((\tilde{x}_i, y_i) \quad i = (1,2,3 \dots N)$  represents the data points used as the predictor p-dimensional vector variable  $\tilde{x}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip})^T$  ( $i = 1, 2, \dots, N$ ) and response value  $N$  ( $y_1, y_2, \dots, y_N$ ). Furthermore  $M_{max}$  Is the number of basis functions (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.  $D_{r,s}^\alpha \beta_m(t^m) := \frac{\partial |\alpha| \beta_m}{\partial \alpha_1 t_1^\alpha \partial \alpha_2 t_2^\alpha} (t^m)$ , 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  $\lambda$ , for each derived term, using the Tikhonov regularization the equation changes to be as follows in equation (9) below.

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

Furthermore, it can be formulated into the Conic Quadratic Problem (CQP) as shown in equation (10) as follows:

$$\begin{aligned} & \min_{t, \theta} \quad t, \\ & \text{subject to } \|\beta(\tilde{d})\theta - y\|_2 \leq t, \\ & \|L\theta\|_2 \leq \sqrt{M} \\ & \text{with } t \geq 0. \end{aligned} \quad (10)$$

### 2.3. Peak Ground Acceleration (PGA)

Peak Ground Acceleration (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. then calculate the AVECOS value with the formula as shown in the following formula (11): [27]

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

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

$$R_{epi} = \sqrt{x^2 + y^2} \quad (12)$$

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 equation (13).

$$\text{PGA (gal)} = 10^{(0.71+0.23(M-6)-\text{Log}(r)-0.0027.r)} \quad (13)$$

where M is the magnitude and r is the root of  $(R_{epi}^2+8^2)$

$$r = \sqrt{R_{epi}^2 + 8^2} \quad (14)$$

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

### 2.4. Data set

The data used in this study were taken from the Geophysics Station (BMKG) of Mataram city. Data in the form of an earthquake catalog in Lombok with positions (-4.0636o) South Latitude – (-13.0636o) South Latitude and (111.5798o) East Longitude – (120.5798o) 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 Mw to 9.0 Mw, and a depth of 0 Km to 500 Km. [24, 26]. 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 below:

Table 1. Earthquake Frequency in Lombok is based on Magnitude

No	Magnitudo (Mw)	Jumlah/Frekuensi
1	4.5 – 5	283
2	5 – 6	121
3	6 – 7	15

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 below:

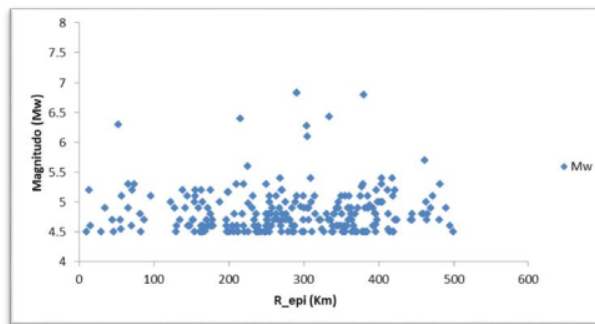


Figure 1. Distribution of Earthquakes in Lombok with Magnitude 4.5 - 7

### 3. RESULTS AND DISCUSSIONS

#### 3.1. PGA value calculation results

Maximum Ground Acceleration (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 following :

Table 2. PGA Value for Earthquake in Lombok

Time	Long	Lat	Depth	Mw	AVECOS	x	y	Repi	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
18/01/2010	114.02	-7.89	13	2.6	0.989709408	-226.6823599	74.90090259	238.736335	238.8703361	-3.095112128	0.000803319
18/01/2010	114.89	-8.81	89	3.8	0.988528709	-130.782075	-27.3984299	133.6212	133.8604688	-2.284075608	0.005199055
18/01/2010	114.02	-9.68	10	4.4	0.987353559	-226.1427778	-124.138016	257.974423	258.0984366	-2.766651153	0.001711389
18/01/2010	118.63	-8.13	15	4.6	0.989407547	280.5656074	48.21412019	284.678172	284.7905571	-2.835460089	0.001460629
21/01/2010	114.14	-9.64	10	3.2	0.987408838	-212.9800564	-119.690219	244.307701	244.4386486	-2.982154225	0.001041947
22/01/2010	117.5	-9.3	75	4.1	0.987873847	156.0040845	-81.883944	176.188123	176.3696534	-2.449621926	0.003551224
22/01/2010	113.98	-9.57	10	3.1	0.987505286	-230.5697522	-111.906574	256.29181	256.4166375	-3.058271122	0.000874438
23/01/2010	114.18	-9.67	11	3.1	0.98736739	-208.5795065	-123.026067	242.158675	242.2907833	-2.995522009	0.001010364
25/01/2010	113.9	-9.61	10	2.7	0.987450218	-239.3408509	-116.354371	266.12475	266.2449673	-3.193142819	0.000640999
26/01/2010	120.29	-7.95	10	3.6	0.98963435	463.3001712	68.22920699	468.297206	468.3655338	-3.77717187	0.000167043
26/01/2010	121.5	-7.39	220	4.5	0.990324341	596.8672339	130.4983659	610.966708	611.0190819	-4.070806294	8.49559E-05
26/01/2010	115.47	-10.54	10	3.2	0.986135982	-66.86659344	-219.765653	229.713046	229.8523082	-2.916050101	0.001213249
27/01/2010	119.07	-9.46	76	4.3	0.987656102	328.3907845	-99.6751322	343.184556	343.2777874	-3.143495728	0.000718628
27/01/2010	118.43	-9.4	10	3.9	0.987737982	258.1258796	-93.0034366	274.369475	274.4860815	-2.952632747	0.001115237
27/01/2010	114.02	-9.77	10	3	0.987228742	-226.1141899	-134.14556	262.911883	263.0335682	-3.110201811	0.000775886
28/01/2010	116.12	-8.25	142	2.6	0.989254989	4.422005462	34.870729	35.1499911	36.04888172	-1.726223778	0.018783487
29/01/2010	114.52	-8.99	10	3	0.988290247	-171.4108854	-47.4135167	177.8475	178.0273383	-2.7111160512	0.001944641

... etc

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 'SUHU', obtained ready data. used for the predictive analysis process as shown in Table 3 the following.

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
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

... etc

### 3.3. 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 following:

Table 4. 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
-----

```

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 basis functions (BF) are formed, namely BF 1, 2, 3, 5, 7, 9, 10, 11, 13, 14, 15 and 16. After doing elimination by eliminating the basis function (BF) which does not have a contribution to the change in the dependent variable, namely the basis function (BF) 4, 6, 8, and 12 in order to obtain the best MARS model as in equation (15) below. :

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

MODEL PGA\_G\_ = BF1, BF2, BF3, BF5, BF7, BF9, BF10, BF11, BF13, BF14, BF15, BF16

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 Epicenter distance (R\_epi), Magnitude (Mw), temperature of the location of the incident (SUHU), and depth (Depth). As shown in the following table 5 which describes the interactivity of the contribution of independent variables to the dependent variable:

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

It can be seen in Table 7 that the variables that are very influential in the PGA value are the Epicenter Distance (R\_epi) of 100% and the Magnitude (Mw) of 31.08608%, while the temperature of the location of the incident (SUHU) is 5.48525% and the depth (Depth), amounting to 3.52988%.

### 3.4. 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.



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:

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

$$H_1 : \text{there is at least one } a_m \neq 0 ;$$

$$m = 1,2,3,5,7,9,10,11,13,14,15,16 \text{ (significant model)}$$

- 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  $< \alpha$

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  $m < \alpha$  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).

### 3.5. Potential Areas with the Highest Earthquake Danger 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 below.

Table 6. Highest Earthquake Hazard Potential Areas in Lombok

No	Time	Lat	Long	Depth	Mw	R-epi	PGA(g)	PGV (cm/s)	SUHU (°)	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, Pemenang
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

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.

## 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.

## REFERENCES

- [1] Han, J. Kamber, M. Pei, J., "Data mining : concepts and techniques", Morgan Kaufmann, 225Wyman et, Waltham, MA 02451, USA, 2012.
- [2] Hastie, T., Tibshirani R., Friedman J., "The Elements of Statistical Learning, Data Mining, Inference and Prediction", Springer, New York, 2001.

- [3] Larose, D.T. *29* *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley-Interscience, John Wiley & Sons, Inc. *2* Hoboken, 2005.
- [4] Yerlikaya, F., Batmaz, I., Weber, G.W., "A Review and New Contribution on Conic Multivariate Adaptive Regression Splines (CMARS): A Powerful *33* 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
- [5] Yerlikaya, F. 2008. "A New Contribution to Nonlinear Robust Regression and Classification with MARS and Its Application to Data Mining for Quality Control in Manufacturing." MSc diss., Middle East Technical University, Turkey *1*
- [6] Yerlikaya, F., Askan, A., Weber, G.W., "An alternative approach to the ground motion prediction problem by a *25* non-parametric adaptive regression method", *Engineering Optimization*, Vol. 46, No. 12, 1651–1668. 2014.
- [7] K. M. Asim, A. Idris, and T. Iqbal, , "Earthquake prediction model using support vector regressor and hybrid neural networks", 2018, pp. *18*–*22*.
- [8] Asencio. C.G, Martu N. L. F, Troncoso A, Morales E. A. 2015, "*5* Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks". *Neural Computing and Applications*.:1–13. <https://doi.org/10.1007/s00521-015-2121-7>.
- [9] Ozkurt, Y., Askan, A., Weber, G.-W., "A Hybrid Computational Method Based on Convex Optimization for *40* Outliers: Application to Earthquake Ground Motion Prediction", *INFORMATICA*, Vol. 27, No. 4, 893–910, DOI: <http://dx.doi.org/10.15388/Informatica.2016.116>, 2016.
- [10] Weber, G.W., Batmaz, I., Koksals, G., Taylan, P., Yerlikaya, F., 2012, CMARS: A New Contribution to Nonparametric Regression with Multivariate Adaptive Regression Splines Supported by Continuous Optimisation, *Inverse Problems in Science and Engineering* : 1–33, <https://www.researchgate.net/publication/233429692>
- [11] Zheng G, Yang P, Zhou H, Zeng C, Yang X, He X, Yu X, 2019 "Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines", journal homepage: [www.elsevier.com/locate/compegeo](http://www.elsevier.com/locate/compegeo)
- [12] Yerlikaya, F., Askan, A., Weber, G.W., "An alternative approach to the ground motion prediction problem by a non-parametric adaptive regression *50* method", *Engineering Optimization*, Vol. 46, No. 12, 1651–1668. 2014.
- [13] Peru's, I., and P. Fajfar. 2010. "Ground-Motion Prediction by a Non-Parametric Approach." *Earthquake Engineering and Structural Dynamics* 39 (12): 1395–1416.
- [14] Eubank, R.L., 1999, *Nonparametric Regression and Spline Smoothing, Second Edition*, Marcel Dekker, New York. *6*
- [15] 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 *2*" *Inverse Problems in Science and Engineering*, 20 (3) 2012: 371–400.
- [16] Yerlikaya, F., Batmaz, I., Weber, G.W., "A Review and New Contribution on Conic Multivariate Adaptive Regression Splines (CMARS): A Powerful *33* 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
- [17] Yerlikaya, F., Askan, A., Weber, G.W., "An alternative approach to the ground motion prediction problem by a *5* non-parametric adaptive regression method", *Engineering Optimization*, Vol. 46, No. 12, 2014, pp 1651–1668.
- [18] Yazici C, Yerlikaya, F.O, Batmaz I,"A computational approach to nonparametric regression: bootstrapping *28* ARS method", <https://www.researchgate.net/publication/277907238>. 2015.
- [19] Friedman, J.H., "Multivariate Adaptive Regression Spline (With Discussion)", *The Annals of Statistics*, Vol. 19, *16* 1, pp. 1-141.
- [20] Ozmen, A., Batmaz, I., Weber, G.-W., 2014a. Precipitation modeling by polyhedral RCMARS and comparison with MARS and CMARS. *Environ. Model. Assess.* 19 (5), 425–435.
- [21] Weber, G.W., Cevik, A., " Voxel-MARS and CMARS: Methods for Early Detection of Alzheimer's Disease by *3* Classification of Structural Brain MRI", <https://www.researchgate.net/publication/327338165>, 2018.
- [22] Joyner, W.B. and Boore, D.M,"Prediction of earthquake response spectra.", *Proceedings 51<sup>st</sup> Annual Convention, Structural Engineers Association of California (Also U.S. Geological Survey Open File Report 1982, 82-977)*.
- [23] Joyner, W.B. and Boore, D.M,"Measurement, Characterization and Prediction of Strong Ground Motion.", *Proceedings of Conference on Earthquake Engineering and Soil Dynamics, II, Park City, Utah, ASCE, Washington, DC.* 1988.
- [24] *19* KG, 2019, [https://inatews.bmkg.go.id/new/view\\_event\\_eq.php?catid=5953](https://inatews.bmkg.go.id/new/view_event_eq.php?catid=5953)
- [25] Arthur .C.K, Temeng. V. A, Ziggah. Y *51* "Multivariate Adaptive Regression Splines (MARS) approach to blast-induced ground vibration Prediction", <https://doi.org/10.1080/17480930.2019.1577940>, 2019.
- [26] Hair J.F., Anderson R.E., Babin B.J, Black W.C. 2014. *Multivariate Data Analysis 7th Edition Pearson Education.* *22* Prentice Hall, inc
- [27] Shodiq, M.N, Kusuma, D.H., Rifqi, M.G., "Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia", Vol 2, pp.37-43. 2018. *22*

## BIOGRAPHIES OF AUTHORS



Dadang Priyanto, M.Kom, is now a student of the Doctoral Program (S3) in Computer Science at the Universitas Sumatera Utara, a Masters Degree (S2) in Computer Science was obtained at Gajah Mada University Yogyakarta in 2006, and a Bachelor Degree (S1) in Informatics Engineering was obtained at STMIK Akakom Yogyakarta in 1998. Currently working as a permanent lecturer at the University of Bumigora Mataram, in the Software Engineering study program. Research has been carried out since 2005 in the field of Multimedia and Data Science. The latest publication is related to Earthquake Prediction Data Mining in Lombok.



Prof. DR. 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.



Prof. Dr. 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.



Dr. Syahril Efendi S.Si., M.IT. 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.

# Analysis of Earthquake Hazards Prediction with Multivariate Adaptive Regression Splines

---

ORIGINALITY REPORT

---

# 18%

SIMILARITY INDEX

---

## PRIMARY SOURCES

---

- 1** [etd.lib.metu.edu.tr](http://etd.lib.metu.edu.tr)  
Internet 95 words — 2%
- 2** Prateek Mittal, Kishalay Mitra. "chapter 8 Decomposition-Based Multi-Objective Optimization of Energy Noise Trade-Off in a Wind Farm", IGI Global, 2018  
Crossref 62 words — 1%
- 3** Gregory P. Tsinker. "Handbook of Port and Harbor Engineering", Springer Science and Business Media LLC, 1997  
Crossref 54 words — 1%
- 4** [mafiadoc.com](http://mafiadoc.com)  
Internet 49 words — 1%
- 5** [www.springerprofessional.de](http://www.springerprofessional.de)  
Internet 47 words — 1%
- 6** Mohammad Ali Sahraei, Hakan Duman, Muhammed Yasin Çodur, Ecevit Eyduran. "Prediction of Transportation Energy Demand: Multivariate Adaptive Regression Splines", Energy, 2021  
Crossref 46 words — 1%
- 7** Ananto Wibowo, M. Rismawan Ridha. "Comparison of Logistic Regression Model and MARS Using 45 words — 1%

---

8 V V Bukhtoyarov, V S Tynchenko, E A Petrovsky, S G Dokshanin, V V Kukartsev. "Research of methods for design of regression models of oil and gas refinery technological units", IOP Conference Series: Materials Science and Engineering, 2019 34 words — 1%

Crossref

---

9 S D P Yasmirullah, B W Otok, J D T Purnomo, D D Prastyo. "Parameter Estimation of Multivariate Adaptive Regression Spline (MARS) with Stepwise Approach to Multi Drug-Resistant Tuberculosis (MDR-TB) Modeling in Lamongan Regency", Journal of Physics: Conference Series, 2021 33 words — 1%

Crossref

---

10 china.iopscience.iop.org 33 words — 1%

Internet

---

11 www.researchgate.net 30 words — < 1%

Internet

---

12 www.tandfonline.com 30 words — < 1%

Internet

---

13 dergipark.org.tr 29 words — < 1%

Internet

---

14 Gang Zheng, Pengbo Yang, Haizuo Zhou, Chaofeng Zeng, Xinyu Yang, Xiaopei He, Xiaoxuan Yu. "Evaluation of the earthquake induced uplift displacement of tunnels using multivariate adaptive regression splines", Computers and Geotechnics, 2019 27 words — < 1%

Crossref

---



- 15 "Mathematical Methods in Engineering", Springer Science and Business Media LLC, 2019  
Crossref 26 words — < 1%
- 
- 16 [link.springer.com](http://link.springer.com)  
Internet 25 words — < 1%
- 
- 17 [garuda.ristekbrin.go.id](http://garuda.ristekbrin.go.id)  
Internet 24 words — < 1%
- 
- 18 "Prediction of Seismic zone in India using Neural Network Algorithms", International Journal of Innovative Technology and Exploring Engineering, 2019  
Crossref 21 words — < 1%
- 
- 19 Yi Luo, Xiaoqing Wei, Junhong Huang, Guang Zhang, Xing Bian, Xinping Li. "PPV distribution of sidewalls induced by underground cavern blasting excavation", Scientific Reports, 2021  
Crossref 21 words — < 1%
- 
- 20 [moon.bme.ym.edu.tw](http://moon.bme.ym.edu.tw)  
Internet 21 words — < 1%
- 
- 21 Heru Mardiansyah, Rahmat Widia Sembiring, Syahril Efendi. "Handling Problems of Credit Data for Imbalanced Classes using SMOTEXGBoost", Journal of Physics: Conference Series, 2021  
Crossref 20 words — < 1%
- 
- 22 [produvia.com](http://produvia.com)  
Internet 20 words — < 1%
- 
- 23 [repository.unpar.ac.id](http://repository.unpar.ac.id)  
Internet 19 words — < 1%
- 
- 24 [www.scilit.net](http://www.scilit.net)  
Internet

19 words — < 1%

25 Mihaela Oprea. "A general framework and guidelines for benchmarking computational intelligence algorithms applied to forecasting problems derived from an application domain-oriented survey", Applied Soft Computing, 2020  
Crossref

18 words — < 1%

26 [jurnal.iainkediri.ac.id](http://jurnal.iainkediri.ac.id)  
Internet

18 words — < 1%

27 Serhat Yüksel, Shahriyar Mukhtarov, Ceyhun Mahmudlu, Jeyhun Mikayilov, Anar Iskandarov. "Measuring International Migration in Azerbaijan", Sustainability, 2018  
Crossref

17 words — < 1%

28 [www.scribd.com](http://www.scribd.com)  
Internet

17 words — < 1%

29 [www.journalglobal.in](http://www.journalglobal.in)  
Internet

16 words — < 1%

30 [id.123dok.com](http://id.123dok.com)  
Internet

15 words — < 1%

31 [bearworks.missouristate.edu](http://bearworks.missouristate.edu)  
Internet

14 words — < 1%

32 Pakize Taylan, Gerhard-Wilhelm Weber, Fatma Yerlikaya Özkurt. "A new approach to multivariate adaptive regression splines by using Tikhonov regularization and continuous optimization", TOP, 2010  
Crossref

13 words — < 1%

- 
- 33 "Modeling, Dynamics, Optimization and Bioeconomics II", Springer Science and Business Media LLC, 2017  
Crossref 12 words — < 1%
- 
- 34 Alper Çevik, Gerhard-Wilhelm Weber, B. Murat Eyüboğlu, Kader Karlı Oğuz. "Voxel-MARS: a method for early detection of Alzheimer's disease by classification of structural brain MRI", Annals of Operations Research, 2017  
Crossref 12 words — < 1%
- 
- 35 repository.its.ac.id  
Internet 11 words — < 1%
- 
- 36 blog.umy.ac.id  
Internet 10 words — < 1%
- 
- 37 Ulusay, R.. "An attenuation relationship based on Turkish strong motion data and iso-acceleration map of Turkey", Engineering Geology, 200408  
Crossref 9 words — < 1%
- 
- 38 stat.metu.edu.tr  
Internet 9 words — < 1%
- 
- 39 theses.ubn.ru.nl  
Internet 9 words — < 1%
- 
- 40 www.coursehero.com  
Internet 9 words — < 1%
- 
- 41 Annisa Nur Insany, Nur'eni Nur'eni, Mohammad Fajri. "Pemodelan IPM Di Kawasan Timur Indonesia Menggunakan Multivariate Adaptive Regression Spline (MARS)", Natural Science: Journal of Science and Technology, 2019  
Crossref 8 words — < 1%

42 Ceyda Yazıcı, Fatma Yerlikaya-Özkurt, İnci Batmaz. "A computational approach to nonparametric regression: bootstrapping CMARS method", Machine Learning, 2015 8 words — < 1%  
Crossref

43 Ezgi Ayyıldız, Vilda Purutçuoğlu, Gerhard Wilhelm Weber. "Loop-based conic multivariate adaptive regression splines is a novel method for advanced construction of complex biological networks", European Journal of Operational Research, 2017 8 words — < 1%  
Crossref

44 cocukergen2020.com 8 words — < 1%  
Internet

45 dokumen.pub 8 words — < 1%  
Internet

46 jurnalsaintek.uinsby.ac.id 8 words — < 1%  
Internet

47 repositorium.sdum.uminho.pt 8 words — < 1%  
Internet

48 www.mdpi.com 8 words — < 1%  
Internet

49 Maylita Hasyim, Dwi Shinta Rahayu, Novita Eka Muliawati, Diesty Hayuhantika et al. "Bootstrap Aggregating Multivariate Adaptive Regression Splines (Bagging MARS) to Analyse the Lecturer Research Performance in Private University", Journal of Physics: Conference Series, 2018 7 words — < 1%  
Crossref

50 Önder Nazım Onak, Yesim Serinagaoglu Dogrusoz, Gerhard Wilhelm Weber. "Evaluation of 7 words — < 1%

multivariate adaptive non-parametric reduced-order model for solving the inverse electrocardiography problem: a simulation study", Medical & Biological Engineering & Computing, 2018

Crossref

51

Bruno Perez, Christophe Lang, Julien Henriet, Laurent Philippe, Frédéric Auber. "A cluster-based approach to predict serious adverse events in surgery", Expert Systems with Applications, 2021

6 words — < 1%

Crossref

EXCLUDE QUOTES OFF

EXCLUDE MATCHES OFF

EXCLUDE BIBLIOGRAPHY OFF