ANOVA Decomposition and Importance Variable Process in Multivariate Adaptive Regression Spline Model

Bambang Widjanarko Otok [a,*], Agnes Tuti Rumiati [a], Ayub Parlin Ampulembang [b], Harun Al Azies [a]

[a] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[b] Badan Pusat Statistika Provinsi Sulawesi Selatan, Makassar, 90125, Indonesia
Corresponding author: *dr.otok.bw@gmail.com

Abstract— **This article reviews one of the non-parametric functions, namely the MARS (Multivariate Adaptive Regression Spline) method, a complex combination of recursive partitioning and spline regression. The many advantages of the MARS function over other non-parametric regression functions are of interest to researchers. One of them is it can accommodate the additive and interaction model to improve the prediction and interpretation of the data. There are some important things in the MARS method, namely, ANOVA decomposition and Importance Variable. Decomposition ANOVA is a technique in MARS that is useful for grouping basis function based on variables engagement, whether they enter by one variable or interactions with other variables, making it easier to interpret in graphical form. In comparison, the important variables are a technique that can be used to determine which predictor variables most influence the MARS modeling. This study assesses ANOVA decomposition, and the important variables process in MARS modeling based on GCV and MSE criteria. We use the poverty rate modeling data on Java Island to implement the study results. The results show that the MARS model's interpretation of the poverty rate can be better done through ANOVA decomposition. Besides that, based on GCV and MSE criteria, the result also shows that the biggest variable importance in poverty rate modeling on Java Island is the percentage of per capita expenditure for food, while the smallest is the economic growth variable.**

Keywords— **MARS; ANOVA decomposition; importance variable; MSE; GCV.**

I. INTRODUCTION

Regression analysis is a method that aims to model the relationship pattern between the response and predictor variable [1]. Several approaches in regression analysis can be used to model the relationship above, and o of them is non-parametric regression [2]. In some cases, regression modeling often cannot be solved with a parametric regression approach and must be solved with a non-parametric regression approach, namely in cases where the shape of the regression curve is/is not still known. Compared to other regression approaches, non-parametric regression has many advantages. This is because the approach is not limited by the form of a certain function, such as linear, quadratic, or cubic [3].

This case can be solved because non-parametric regression can find the shape of the regression curve model, which is not/not yet known. This ability is supported by the presence of parameters in each non-parametric regression method, which makes the regression curve model more flexible. One method in the non-parametric regression approach is

Multivariate Adaptive Regression Spline (MARS) [4]. This method was developed first by Friedman in 1991. This method has many advantages because MARS models not only the additives effect but also the interaction effect in data modeling [5], so it allows better prediction and better interpretation than the other function in non-parametric regression [6].

There are some important things in the MARS method, namely ANOVA decomposition [7] and Importance Variable [8]. Decomposition ANOVA is a technique in MARS useful for grouping basis functions based on variables engagement [9], whether they enter by one variable or interactions with other variables, making it easier to interpret in graphical form. In comparison, the important variables are a technique that can be used to determine which predictor variables most influence MARS modeling [10]. During this time, the research of MARS was limited to data modeling. Several studies applying MARS principles that focus only on data modeling include 2019 by Kumar and Samui [11], who wanted to predict and compare the bearing capacity performance of piles embedded in soil without cohesion. In

addition, a study to obtain an accurate model for predicting nanofluid features was carried out by Maleki using three methods, including MARS [12]. Other than that, other research is to predict flow patterns in semi-arid areas with a new hybrid model called Multivariate Adaptive Regression Spline, which is integrated with Differential Evolution (MARS-DE). The results obtained in this study are the excellent capability of the MARS-DE model for monthly runoff fluxes in the semi-arid region [13]. Some of the results of the research above and most of the research related to the MARS method are still limited to data modeling while discussing how many studies have not developed the process of obtaining the ANOVA decomposition and importance variable, respectively. Based on those descriptions, this research aims to study the process of the ANOVA decomposition and importance variable in MARS modeling.

This paper is divided into four parts. The first part is an explanation related to the introduction, where the problems and objectives of this research are proposed. Furthermore, the methodology section, the second section, explains the conceptual analysis method used and the data sources as a case study in this research. The third section is the analysis section and the results of the process to get the importance of the variables studied by the researcher, and finally, the last section reveals the conclusions reached.

II. MATERIALS AND METHOD

This chapter describes several concepts, research findings, and theorems that serve as a review of the literature and a theoretical basis to support the research. The study begins by explaining the regression analysis theory, including non-parametric regression. The regression study focused on discussing one of the non-parametric regression methods used in this study, namely MARS. The MARS method, ANOVA decomposition, and processes of variables of importance in MARS modeling are also described in this chapter. The MARS model and the process of variable importance in MARS modeling developed in this study will then be applied to the modeling of data on the level of well-being of the inhabitants of the island of Java, Indonesia, from the database of Statistics Indonesia data.

A. Conceptual of Non-parametric Regression Model

Regression analysis in statistics explains a causal relationship between responses and predictors [14]. Two approaches are often used in regression analysis: parametric and non-parametric regression. The main difference between these two regression approaches lies in the assumption of the shape of the regression curve pattern used in the data modeling [15]. Regression studies are focused on this research, namely the non-parametric regression method.

The parametric regression approach cannot be used when the model of the relationship between the response and the predictor is not known for the shape of the regression curve because if it is forced and the shape of the curve n is not appropriate, it will produce a large error variance. The approach that should be used in this condition is a non-parametric regression approach [16]. Regression studies discuss one non-parametric regression method used in this study, Multivariate Adaptive Regression Spline (MARS). The

MARS method is an extension of the truncated spline method which is also described in this chapter.

B. Truncated Spline in Non-parametric Regression

The spline is a very popular method in the non-parametric regression approach. The spline as a data model approach was introduced by Whittaker [17], while the spline as an optimization problem was developed by Reinsch [18]. Spline methods in non-parametric regression can be found in many forms, including smoothing splines [19], [20], [21], and truncated splines [22], [23], [24].

The two spline methods use different parameters to estimate the regression curve more flexibly, namely the smoothing parameter for the smoothing spline and the knot point on the truncated spline [25]. The difference in the types of parameters causes the optimization to get an estimator for the two spline methods is also different. The difference in the types of parameters causes the optimization to get an estimator for the two spline methods is also different. A smoothing spline estimator that depends on the smoothing parameter is obtained by optimization of penalized least squares. The difference in the types of parameters causes the optimization to get an estimator for the two spline methods is also different. The smoothing spline estimator that depends on the smoothing parameter is obtained by optimization of penalized least squares (PLS) [20], while the estimator of truncated spline that depends on knot point is obtained by optimization of ordinary least squares (OLS) [26].

C. Conceptual of Multivariate Adaptive Regression Spline (MARS)

The truncated spline method, in its application, often has limitations in determining the position and number of node points used when regression modeling involves many predictors. Indeed, the combination of nodes to be chosen is very large and complex, namely from the combination of the number of predictors, the position of the nodes, and the number of nodes [27]. The MARS method makes it possible to overcome the weakness of the truncated spline in this case because the determination of the nodes in MARS is not sought individually from the combination but by an adaptive process. The adaptive process in MARS is performed with a stepping algorithm including forward and backward stepwise [28].

The multivariate adaptive regression spline (MARS) method developed by Friedman in 1991 is part of a non-parametric regression model useful for solving high-dimensional data problems [29]. This method can produce accurate response variable predictions and produce a continuous model in knots based on the smallest Generalized cross-validation (GCV) value [30]. GCV is a method for obtaining optimal knots. At each knot, there is expected to be continuity of the base of function from one region to another. The function's basis explains the relationship between the response variable and the predictor. It was suggested that the maximum number of basic functions (BF) is two to four times the number of predictor variables while the maximum number of interactions (MI) is one, two, or three, knowing that more than three results in a very complex model and a minimal distance between knots or minimum observation (MO) of zero, one, two, three, five, and ten [31]. Multivariate Adaptive Regression Spline (MARS) is a non-parametric function that

is a complex combination between spline regression and recursive partitioning (RPR) [32]. According to Friedman [33], MARS model can be expressed in equation form as follows.

$$y_i = f(x_{1i}, x_{2i}, \ldots, x_{pi}) + \varepsilon_i; \ \ i = 1, 2, \ldots, n \quad (1)$$

where:

$$f(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} [(s_{km}(x_{ji(k,m)} - t_{j(k,m)})]$$
$$= \alpha_0 + \sum_{m=1}^{M} \alpha_m B_m \quad (2)$$

The variable notation in eq. (1) and (2) is expressed as follows: y and x are the response and predictor variable, respectively, f is MARS function, ε is random error, p is the number of predictors, α is the coefficient of basis function, M is the number of basis function, K_m is the kth interaction, $s_{km} = \pm 1$ is the sign of a pair of basis function, and B_m is the mth basis function. In simple form, MARS model in eq. (1) can be expressed in matrix form as follows:

$$\underset{\sim}{y} = \underset{\sim}{f}(\underset{\sim}{x}) + \underset{\sim}{\varepsilon} = B\underset{\sim}{\alpha} + \underset{\sim}{\varepsilon} \quad (3)$$

where:
$$\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)^T, \ \underset{\sim}{f} = (f_1, f_2, \ldots, f_n)^T, \ \underset{\sim}{\alpha}$$
$$= (\alpha_0, \alpha_1, \ldots, \alpha_M)^T$$

$$B = \left(\underset{\sim}{1} \quad \prod_{k=1}^{K_1} [s_{k1}(\underset{\sim}{x}_{j(k,1)} - t_{j(k,1)})] \quad \cdots \quad \prod_{k=1}^{K_M} [s_{kM}(\underset{\sim}{x}_{j(k,M)} - t_{j(k,M)})] \right)$$
$$\underset{\sim}{1} = (1 \quad 1 \quad \cdots \quad 1_n)^T; \ \underset{\sim}{x}_j$$
$$= (x_{j1} \quad x_{j2} \quad \cdots \quad x_{jn})^T; \ j$$
$$= 1, 2, \ldots, p$$

In order to get parameter estimation in the MARS model, we use the ordinary least square (OLS) method by assuming random error as a normal distribution with mean $\underset{\sim}{0}$ and variance $\sigma^2 I$ with result as follows:

$$\hat{\underset{\sim}{f}}(\underset{\sim}{x}) = B(B^T B)^{-1} B^T \underset{\sim}{y} \quad (4)$$

According to Friedman, the best model in MARS is obtained by stepwise procedure, including forward and backward stepwise [34]. The model was built forward stepwise by adding the basis function until we obtained the model with the maximum number of basic functions. In comparison, the backward stepwise aims to get the simplified model (parsimonious) [35]. In this stage, we eliminate basis functions with a small contribution to the model controlled by minimum GCV value [36]. The value of GCV in MARS is expressed in the formula as follows [33]

$$GCV = \frac{n^{-1}((\underset{\sim}{y} - \hat{\underset{\sim}{f}}(\underset{\sim}{x}))^T (\underset{\sim}{y} - \hat{\underset{\sim}{f}}(\underset{\sim}{x})))}{(1 - c(M)/n)^2} \quad (5)$$

Where $c(M)$ is the complexity penalty with the formula: $c(M) = M((d/2) + 1) + 1$, for M is the number of basis functions while d is factor penalty which has the best value on interval 2 to 4 [33].

D. ANOVA Decomposition in MARS Modeling

MARS model in eq. (2) can be classified into some functions based on predictor variables that enter into the model, whether they contain one variable or interaction between variables. The classification makes MARS model can be easier to explain the relationship between the response variable and predictor variables. ANOVA Decomposition can be formulated as follows [37].

$$f(x) = \alpha_0 + \sum_{K_m=1} f_a(x_{ai}) + \sum_{K_m=2} f_{ab}(x_{ai}, x_{bi})$$
$$+ \sum_{K_m=3} f_{abc}(x_{ai}, x_{bi}, x_{ci}) + \ldots \quad (6)$$

Eq. (5) shows that the sum of the first function covers all basic functions for one predictor, the sum of the second function covers all basic functions for the interaction between two predictors, the sum of the third function covers all basic functions for the interaction between three variables, and so on. The contribution of interaction between two variables in Eq. (5) can be expressed in the following equation.

$$f_{ab}^*(x_{ai}, x_{bi}) = f_a(x_{ai}) + f_b(x_{bi}) + f_{ab}(x_{ai}, x_{bi}) \quad (7)$$

While on the higher level, it can be expressed in the same manner as shown in eq. (6).

Eq. (5) in MARS modeling is known as ANOVA decomposition techniques. This technique can be used to determine the contribution of each function (ANOVA function), which is grouped by one predictor or interaction between predictors. The contribution of the ANOVA function can be visualized in graphic form. For example, the contribution of one predictor can be visualized by plotting $f_a(x_a)$ to x_a, whereas the contribution of interaction between two predictors can be visualized by plotting $f_{ab}(x_a, x_b)$ to x_a and x_b.

The ANOVA decomposition in the MARS model can be obtained by using algorithm stages as follows [9]
1. Getting the optimal basis function in MARS model.
2. Grouping the optimal basis function obtained in first stage, based on predictor's involvement in model, whether it involves one predictor or interaction between predictos, thus forming some group function called ANOVA function.
3. Calculate the standard deviation and the contribution of each ANOVA function to MARS model based on GCV value.
4. Presenting the ANOVA function in graphical form so the model interpretation becomes easier.

E. Importance Variable in MARS Modeling

One of the important things in MARS modeling is the important variable. Based on the important variable, we can know which predictor variables provide the biggest impact on the data modeling. In order to get the value of the important variable in MARS, we need the algorithm proceeds as follows [12].
1. Getting the optimal basis function from the MARS modeling.
2. Remove basis functions from the set of optimal basis functions obtained from step 1. We start from the basis function with the smallest contribution to the model

according to the Mean Squared Error (MSE) value and GCV.

3. Calculate the difference in mse value (dif-mses) and GCV value (dif-gcvs) respectively from each step in the elimination process in step 2 compared to the condition of mse and GCV value before the elimination is done.

4. Calculate the cumulative value from the dif-mses and dif-gcvs obtained in step 3, for each predictor variable involved in the elimination step.

5. Determine the maximum value of the cumulative dif-mses and dif-gcvs obtained in step 4.

6. Determine the value of the importance variable by obtaining the square root of the division between each value in dif-mses and dif-gcvs with the maximum value of dif-mses and dif-gcvs, respectively, obtained in step 5 for each predictor variable.

F. Data Analysis Steps

The steps followed to analyze the method used are as follows:

1. Form a plot matrix between the response variables (Y) and each predictor variable (X) which is used as the initial detection of the pattern of relationships between these variables.

2. Develop data modeling on modeling well-being indicators in Java using the MARS method. The steps of analysis using the MARS method are the following:
   a. Combination of basis function (BF), maximum interaction (MI), and minimum observation (MO)
   b. Determine the best model with the minimum GCV value. Based on the results of the best model in step (2b). Furthermore, the analysis steps use the ANOVA decomposition method and the important variable.

3. The analysis steps using the ANOVA decomposition method are as follows to obtain the value of the ANOVA decomposition.
   a. Obtain the optimal basis function from the MARS modeling in step (2b). The MARS model equation in step (2b) can be grouped into several ANOVA functions based on the involvement of the variables that enter the model.
   b. Group the basis of the optimal functions obtained in step 3a, according to the involvement of the variables in the model, involving either one variable or interactions between variables, to form several groups of functions called ANOVA functions.
   c. Calculation of the standard deviation and contribution of each ANOVA function to the MARS model based on the GCV value.
   d. Presentation of the ANOVA function in a graphical view for easier interpretation. In addition to the ANOVA decomposition, the MARS model also provides functionality for the important variable which is used to determine which predictor variable has the greatest influence on the MARS model.

4. To obtain the variable importance value of each predictor as follows:
   a. Obtain the optimal basis function from the MARS modeling in step (2b).
   b. Remove the basic functions one by one from the set of optimal basis functions obtained in step (4a), starting with the basic functions that have the smallest contribution to the model according to the Mean Squared Error (MSE) and generalized cross-validation (GCV) values.
   c. Calculate the difference in the value of MSE (dif-mses) and the difference in the value of GCV (dif-gcvs), from each step of removing the basis function performed in step (4b) with the condition of the MSE values and GCV before the deletion step is performed.
   d. Calculate the cumulative value of the dif-mses and dif-gcvs values obtained in step (4c), for each predictor variable involved in the base function deletion phase.
   e. Determine the maximum value of the cumulative dif-mses and dif-gcvs obtained in step (4d).
   f. Determining the importance variable value of each predictor variable by obtaining the square root of dividing each value of dif-mses and dif-gcvs with the maximum value of dif-mses and maximum dif-gcvs obtained at step (4e) for each variable predictor.

The stages of data analysis in this study can be described in a flow chart, as shown in Figure 1 below.
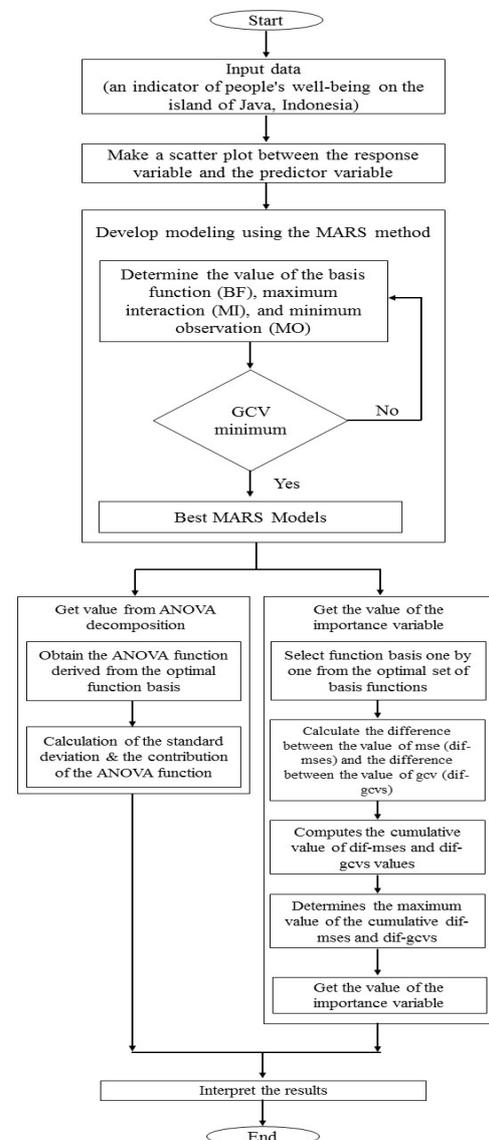

Fig. 1 Research flow chart ANOVA decomposition and importance variable process in multivariate adaptive regression spline model

G. Data Source

To implement the ANOVA decomposition and importance variables process in MARS, then this research uses the data about welfare indicators modeling in Java Island, Indonesia, sourced from the Statistics Indonesia database. As a response variable (y), we use the data of poverty rate, while for the predictor variable (x), we use the data of percentage per capita expenditure for food (x1), economic growth (x2), and unemployment rate (x3), respectively. Furthermore, to identify the relationship pattern between the response and predictor variable, then we present a scatter plot of the research variables as follows.
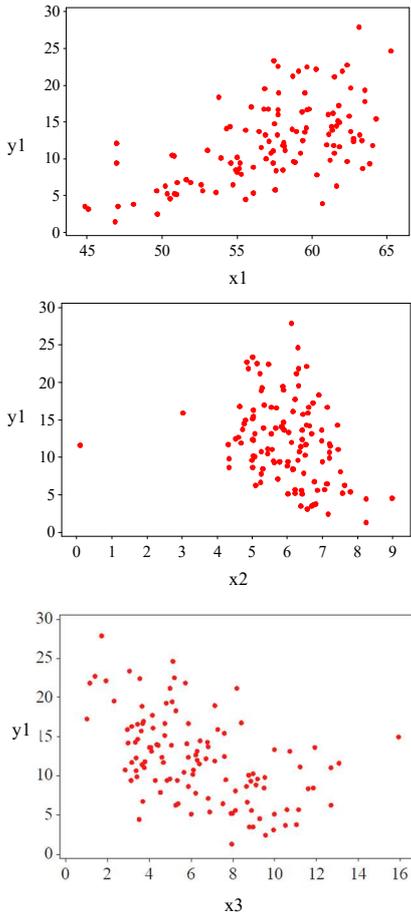

Fig. 2 Scatter plot between the response variable "y" and the predictor variable "x" (a) variable "y" with variable "x1" (b) variable "y" with variable "x2" (c) variable "y" with variable "x3"

Based on Fig.2 (a-c), the relationship pattern between the response and predictor variables appears to show an unknown pattern, or there is a tendency to have behavior change in some sub-interval data. These conditions cause non-parametric regression, as MARS is more appropriately used to model the poverty rate data than parametric regression.

III. RESULTS AND DISCUSSION

Based on data processing in MARS, the estimate of the optimal function basis is obtained in the regression equation as follows.

$$\hat{f} = 9.93 - 0.53 * BF_1 + 3.49 * BF_2 - 1.58 * BF_3 \\ + 8.89 * BF_4 - 2.44 * BF_5 \qquad (8) \\ + 0.93 * BF_6$$

The MARS function in eq. (8) can be grouped into several ANOVA functions based on predictor involvement which enter into the model, whether they enter as one predictor or interaction between predictors. Contributions of the ANOVA function can be expressed in the ANOVA decomposition table as follows.

TABLE I
VARIABLE ANOVA DECOMPOSITION IN MARS MODELING

ANOVA Function	Std.dev	GCV	The Number of Basis	Variable
f1	32.60	12.71	1	x1
f2	32.15	9.34	1	x2
f3	29.42	30.89	2	x3
f4	31.73	8.97	1	x1, x2
f5	32.09	9.94	1	x2, x3

Table 1 shows that the largest contribution to the MARS model (poverty rate) is given by ANOVA function f3 with impairment contributions to GCV model is 30.89. While the smallest contribution is given by ANOVA function f4 with impairment contribution to GCV model is 8.97. Both plot representations by the additive and interaction are presented in Fig. 3 and Fig. 4.

Interpretations that can be obtained from the plot of the ANOVA function in Fig. 3 are as follows: The additive model (ANOVA function f3) shows that the lower unemployment rate (x3) tends to impact the increasing poverty rate. This phenomenon usually occurs in the district/city where their jobs are dominated by the agricultural sector, where their labor involves almost all the family members but with low income. So even though the unemployment is low in this region, they are still poor.
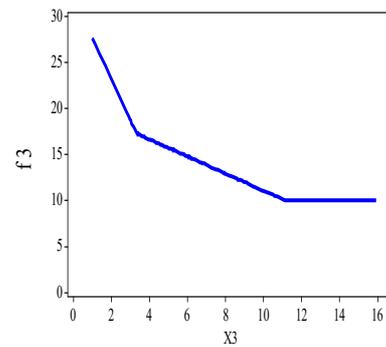

Fig. 3 Plotting the ANOVA function with one predictor additively (ANOVA function f3)

The additional interpretation is shown by ANOVA function f5 (function of interaction), which shows that the unemployment rate does not always affect the poverty rate increase under certain conditions. However, it rather affects the poverty rate decrease with high economic growth in this region. Interpretation for this interaction is not found in other non-parametric models, which model data by additive only, so the interpretation of the model is incomplete.
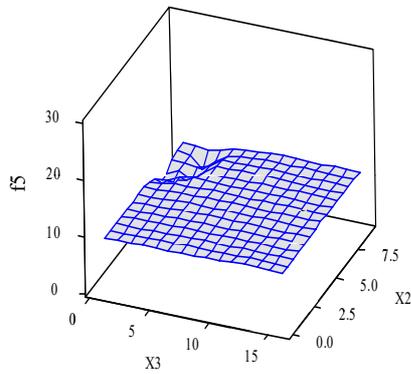
Fig. 4 Plotting the ANOVA function f5 (function of interaction) with interaction between two predictors in poverty rate modeling.

TABLE II
THE VALUE OF MSES , DIF(MSES), GCVS AND DIF (GCVS) ON EACH STAGE OF BASIS FUNCTION ELIMINATION WHICH CONTAIN THE CORRESPONDING PREDICTOR VARIABLE

Source	The stage of basis function elimination						
	1st	2nd	3rd	4th	5th	6th	7th
Basis Function	constant	constant	constant	constant	constant	constant	constant
	x1	x1	x1	x1	x1	x1	-
	x3	x3	x3	x2,x3	x3	-	-
	x2	x2,x1	x2,x3	x3	-	-	-
	x2,x1	x2,x3	x3	-	-	-	-
	x2,x3	x3	-	-	-	-	-
	x3	-	-	-	-	-	-
mses	5.35	5.67	5.91	6.44	7.36	8.84	18.64
dif (mses)	0.32	0.24	0.53	0.92	1.48	9.80	-
gcvs	7.17	7.23	7.19	7.48	8.17	9.39	18.96
dif (gcvs)	0.06	-0.04	0.29	0.69	1.22	9.57	-

remark: dif(mses(i)) = mses(stage-i + 1) – mses(stage-i)
 dif(gcvs(i)) = gcvs(stage-i + 1) – gcvs(stage-i)

In addition to ANOVA decomposition, MARS model also provides important variable technique which serves to determine the predictor variables which have the greatest influence on the MARS model. To determine the variable importance from its equation model, we can start by removing the basis function one by one until the only remaining basis function constant is reached. We use mse and GCV criteria to remove basis function above. Furthermore, in each elimination of the basis function, we calculate the difference value of mse (dif-mses) and GCV (dif-gcvs) compared to the condition before the elimination is done, as shown in Table 2.

The next step is calculating the cumulative value of dif-mse and dif-GCV for each elimination stage containing the corresponding predictor variables. Based on this stage, we obtain the importance variable value from each predictor, as shown in Table 3. Furthermore, based on Table 3, we can determine which of predictor variables which have highest to lowest value in variable importance according to mses and gcvs criteria, variable x1 (percentage of per capita expenditure for food) is a predictor variable which has the highest importance value in modeling data response, namely poverty rate (y) variable with gcvs and mses value are 100 percent respectively. While x2 (economic growth) is a predictor variable with the lowest level with gcvs and mses value of 12.29 percent and 38.89 percent respectively. Besides, based on nsubsets criteria shows that x1 variable is still survive (not removed) until the 6th stage of basis function

elimination. It means that the contribution of these variables is the most influence for the model formation. While for x2 variable is only survive until the 4th stage of basis function elimination and is removed on the 5th elimination. It shows that the contribution of this variable to the overall model is the smallest than the other predictor variables.

TABLE III
CUMULATIVE VALUE OF DIF (MSES) ON EACH STAGE OF BASIS FUNCTION ELIMINATION IN CALCULATING IMPORTANCE VARIABLE ON CORRESPONDING PREDICTOR

The stage of basis function elimination	Variable of dif(mses)		
	x1	x2	x3
1st	0.32	0.32	0.32
2nd	0.24	0.24	0.24
3rd	0.53	0.53	0.53
4th	0.92	0.92	0.92
5th	1.48	-	1.48
6th	9.80	-	-
7th	-	-	-
Total	13.29	2.01	3.49
importance	100.00	38.89	51.24

remark: importance = sqrt(./max)*100

IV. CONCLUSION

Based on results and discussion, it is known that the ANOVA decomposition and importance variable has respective roles in MARS modeling. ANOVA decomposition is good for interpreting the MARS model, whereas the importance variable is good for determining each predictor variable's contribution. The results on real data (poverty rate data) show that the interpretation of MARS model through decomposition ANOVA is advantageous because MARS explains the additive effect and the predictors' interaction effect on the response. It is not found in other non-parametric methods because other non-parametric only interpret data for additives alone. Based on the implementation of MARS on real data, it can be seen that the largest important variable in poverty rate modeling is the percentage per capita food expenditure (x1), whereas the smallest importance is the economic growth (x2).

REFERENCES

[1] G. K. Kate, C. B. Nayak and S. B. Thakare, "Optimization of sustainable high-strength–high-volume fly ash concrete with and without steel fiber using Taguchi method and multi-regression analysis," *Innovative Infrastructure Solutions,* vol. 6, no. 2, pp. 1-18, 2021.
[2] T. B. Armstrong and M. Kolesar, "Simple and honest confidence intervals in non-parametric regression," *Quantitative Economics,* vol. 11, no. 1, pp. 1-39, 2020.
[3] D. P. Rahmawati, I. N. Budiantara, D. D. Prastyo and M. A. D. Octavanny, "Mixed Spline Smoothing and Kernel Estimator in Biresponse Non-parametric Regression," *International Journal of Mathematics and Mathematical Sciences,* 2021.
[4] M. A. Sahraei, H. Duman, M. Y. Codur and E. Eyduran, "Prediction of transportation energy demand: Multivariate Adaptive Regression Splines," *Energy,* p. 224, 2021.
[5] W. Zhang, C. Wu, Y. Li, L. Wang and P. Samui, "Assessment of pile drivability using random forest regression and multivariate adaptive

regression splines," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards,* vol. 15, no. 1, pp. 27-40, 2021.

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

[7] D. H. W. Li, W. Chen, S. Li and S. Lou, "Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) – A case study of Hong Kong," *Energy,* vol. 186, 2019.

[8] N. B. Serrano, A. S. Sanchez, f. S. Lasheras, F. J. Iglesias-Rodriguez and G. F. Valverde, "Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS)," *Applied Ergonomics,* vol. 82, 2020.

[9] K. Sun, M. Rajabtabar, S. Samadi, M. Rezaie-Balf, A. Ghaemi, S. S. Band and A. Mosavi, "An integrated machine learning, noise suppression, and population-based algorithm to improve total dissolved solids prediction," *Engineering Applications of Computational Fluid Mechanics,* vol. 15, no. 1, pp. 251-271, 2021.

[10] Y. Liu, W. Tian and X. Zhou, "Energy and carbon performance of urban buildings using metamodeling variable importance techniques," *Building Simulation,* vol. 14, pp. 535-547, 2021.

[11] M. Kumar and P. Samui, "Reliability Analysis of Pile Foundation Using ELM and MARS," *Geotechnical and Geological Engineering,* vol. 37, pp. 3447-3457, 2019.

[12] A. Maleki, M. Elahi, M. E. H. Assad, M. A. Nazari, M. S. Shadloo and N. Nabipour, "Thermal conductivity modeling of nanofluids with ZnO particles by using approaches based on artificial neural network and MARS," *Journal of Thermal Analysis and Calorimetry,* vol. 143, pp. 4261-4272, 2021.

[13] Z. A. Al-Sudani, S. Q. Salih, A. Sharafati and Z. M. Yaseen, "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation," *Journal of Hydrology,* vol. 573, pp. 1-12, 2019.

[14] S. C. Gupta and V. K. Kapoor, Fundamentals of mathematical statistics, New-Delhi: Sultan Chand & Sons, 2020.

[15] A. E. Marques, P. A. Prates, A. F. G. Pereira, M. C. Oliveira, J. V. Fernandes and B. M. Ribeiro, "Performance Comparison of Parametric and Non-Parametric Regression Models for Uncertainty Analysis of Sheet Metal Forming Processes," *Metals,* vol. 10, no. 4, p. 457, 2020.

[16] D. J. Henderson and A.-C. Souto, "An Introduction to Non-parametric Regression for Labor Economists," *Journal of Labor Research,* vol. 39, pp. 355-382, 2018.

[17] D. P. Rahmawati, I. N. Budiantara, D. D. Prastyo and M. A. D. Octavanny, "Modeling of Human Development Index in Papua Province Using Spline Smoothing Estimator in Non-parametric Regression," in *3rd International Conference on Statistics, Mathematics, Teaching, and Research ,* Makassar, 2021.

[18] M. Kobayashi, K. Hoshina, Y. Nemoto, S. Takagi, M. Shojima, M. Hayakawa, S. Yamada and M. Oshima, "A penalized spline fitting method to optimize geometric parameters of arterial centerlines extracted from medical images," *Computerized Medical Imaging and Graphics,* vol. 84, 2020.

[19] N. P. A. M. Mariati, I. N. Budiantara and V. Ratnasari, "The Application of Mixed Smoothing Spline and Fourier Series Model in Non-parametric Regression," *Symmetry,* vol. 13, no. 11, 2021.

[20] R. Hidayat, I. N. Budiantara, B. W. Otok and V. Ratnasari, "The regression curve estimation by using mixed smoothing spline and kernel (MsS-K) model," *Communications in Statistics - Theory and Methods,* vol. 50, no. 17, pp. 3942-3953, 2019.

[21] R. Hidayat, I. N. Budiantara, B. W. Otok and V. Ratnasari, "An Extended Model of Penalized Spline with The Addition of Kernel Functions in Non-parametric Regression Model," *Applied Mathematics & Information Sciences,* vol. 13, no. 3, pp. 453-460, 2019.

[22] M. A. D. Octavanny, I. N. Budiantara, H. Kuswanti and D. P. Rahmawati, "Non-parametric Regression Model for Longitudinal Data with Mixed Truncated Spline and Fourier Series," *Abstract and Applied Analysis,* 2020.

[23] D. A. Widyastuti, A. A. R. Fernandes and H. Pramoedyo, "Spline estimation method in non-parametric regression using truncated spline approach," in *1 st International Conference on Mathematics and its Applications (ICoMathApp) ,* Malang, 2020.

[24] S. D. P. Yasmirullah, B. W. Otok, J. D. T. Purnomo and D. D. Prastyo, "Modification of Multivariate Adaptive Regression Spline (MARS)," in *International Conference on Mathematics, Statistics and Data Science (ICMSDS),* Bogor, 2020.

[25] D. D. Prawanti, I. N. Budiantara and J. D. T. Purnomo, "Parameter Interval Estimation of Semiparametric Spline Truncated Regression Model for Longitudinal Data," *IOP Conference Series: Materials Science and Engineering,* vol. 546, no. 5, 2019.

[26] M. Ramli, V. Ratnasari and I. N. Budiantara, "Estimation of Matrix Variance-Covariance on Non-parametric Regression Spline Truncated for Longitudinal Data," in *The 15th International Symposium on Geometric Function Theory and Applications ,* Malang, 2020.

[27] M. P. Wand, "A Comparison of Regression Spline Smoothing Procedures," *Computational Statistics,* vol. 15, pp. 443-462, 2000.

[28] A. Ghaemi, M. Rezaie-Balf, J. Adamowski, O. Kisi and J. Quilty, "On the applicability of maximum overlap discrete wavelet transform integrated with MARS and M5 model tree for monthly pan evaporation prediction," *Agricultural and Forest Meteorology,* vol. 278, 2019.

[29] S. K. Depren and M. T. Kartal, "Prediction on the volume of non-performing loans in Turkey using multivariate adaptive regression splines approach," *International Journal of Finance & Economics,* vol. 26, no. 4, pp. 6395-6405, 2021.

[30] B. W. Otok, R. Y. Putra, Sutikno and S. D. P. Yasmirullah, "Bootstrap Aggregating Multivariate Adaptive Regression Spline for Observational Studies in Diabetes Cases," *Systematic Reviews in Pharmacy,* vol. 11, no. 8, pp. 406-413, 2020.

[31] N. Santoso and S. P. Wulandari, "Hybrid Support Vector Machine to Preterm Birth Prediction," *Indonesian Journal of Electronics and Instrumentation Systems (IJEIS),* vol. 8, no. 2, pp. 191-200, 2018.

[32] P. J. García-Nieto, E. García-Gonzalo, J. R. A. Fernández and C. D. Muñiz, "Modeling eutrophication risks in Tanes reservoir by using a hybrid WOA optimized SVR-relied technique along with feature selection based on the MARS approximation," *Stochastic Environmental Research and Risk Assessment,* 2021.

[33] J. H. Friedman, "Multivariate Adaptive Regression Splines," *Ann. Statist,* vol. 19, no. 1, pp. 1-67, 1991.

[34] M. H. Ahmadi, B. Mohseni-Gharyehsafa, M. Farzaneh-Gord, R. D. Jilte, R. Kumar and K.-w. Chau, "Applicability of connectionist methods to predict dynamic viscosity of silver/water nanofluid by using ANN-MLP, MARS and MPR algorithms," *Engineering Applications of Computational Fluid Mechanics,* vol. 13, no. 1, pp. 220-228, 2019.

[35] S. D. P. Yasmirullah, B. W. Otok, J. D. T. Purnomo and D. D. Prastyo, "Multivariate adaptive regression spline (MARS) methods with application to multi drug-resistant tuberculosis (MDR-TB) prevalence," in *AIP Conference Proceedings,* 2021.

[36] J. Diaz, F. J. Fernandez and M. M. Prieto, "Hot Metal Temperature Forecasting at Steel Plant Using Multivariate Adaptive Regression Splines," *Metals,* vol. 10, no. 1, p. 41, 2020.

[37] G. E. B. Archer, A. Saltelli and I. M. Sobol, "Sensitivity measures,anova-like Techniques and the use of bootstrap," *Journal of Statistical Computation and Simulation,* vol. 58, no. 2, pp. 99-120, 1997.