VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590** (Online)
Vol. 7 No. 3 (2025), 228-240

DOI: 10.35580/variansiunm518

APPLICATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) TO MODEL THE FACTORS AFFECTING THE PERCENTAGE OF POOR POPULATION IN INDONESIA

Nur Shanty, Ruliana*, Muhammad Kasim Aidid

Department of Statistics, Faculty of Mathematics and Natural Sciences, State University of Makassar, Indonesia

Keywords: GCV, MARS, Nonparametric Regression, Poverty.

Abstract:

Poverty is one of the social and economic problems that Indonesia continues to face today. The Multivariate Adaptive Regression Spline (MARS) is a nonparametric regression model that estimates the functional relationship between the response variable and predictor variables when the relationship form is unknown. This study aims to estimate the parameters of the Multivariate Adaptive Regression Spline (MARS) method for the percentage of poor population in Indonesia and to identify the factors that significantly affect the percentage of poor population. The results of this study found that the best model was obtained with a combination of BF = 21, MI = 1, and MO = 3, with GCV = 0,3102717. Based on the MARS model, the variables that significantly affect the percentage of the poor population are the percentage of formal workers (x_3), percentage of households with access to proper sanitation (x_4), and Gini Ratio (x_7) with a coefficient of determination (R^2) of 81,44%.

1. Introduction

Regression analysis is a method used to study the relationship between predictor variable X and response variable Y (Tiro, 2010). The relationship between the response variable and predictor variables can be estimated using three approaches parametric, nonparametric, and semiparametric (Meilinda et al., 2021).

One of the approaches used in nonparametric regression is the spline estimator. Splines are pieces of polynomial that have a segmented (piecewise polynomial) property at points called knots (Eubank, 1999). These knot points serve as focal points in the spline function, as the resulting curve is divided into segments at these points (Astiti et al., 2016)

Multivariate Adaptive Regression Spline (MARS) is one type of approach in nonparametric regression. When the functional form is unknown and there is no clear relationship between the dependent and independent variables, nonparametric regression is used (Putra et al., 2021). According to Friedman (1991), in MARS there are several aspects that need to be considered in selecting the best model, particularly when the model has the lowest (minimum) Generalized Cross Validation (GCV) value among competing models (Anam et al., 2017).

Poverty is a condition in which individuals are unable to enjoy various choices and opportunities in fulfilling their basic needs, such as access to health, a decent standard of living, freedom, self-esteem, and the respect afforded to others (Wulandari et al., 2022). As a developing country, Indonesia still faces poverty as one of the main problems in its economy (Mahmud et al., 2020). Meanwhile, the goal of Indonesia's national development is to improve economic

* Corresponding author.
E-mail address: ruliana.t@unm.ac.id

performance by creating jobs and ensuring a better standard of living for its population, which ultimately leads to achieving one of its national development goals, namely reducing poverty (Iskandar & Subekan, 2016).

Statistics Indonesia (BPS) has reported that the poverty rate in Indonesia declined from 2020 to 2023, except in 2021 due to the COVID-19 pandemic. In 2020, the poverty rate was recorded at 9.78%, in 2021 at 10.14%, in 2022 at 9.54%, and in 2023 at 9.35% (BPS, 2023). This percentage is relatively high when compared to other countries in the world, even in Southeast Asia, which makes Indonesia 73rd as the poorest country in the world and makes it difficult for Indonesia to get out of the middle-income trap. To optimize the handling of this poverty problem, the government needs to prioritize the factors that affect poverty itself (Septa et al., 2025). Therefore, this study applies the MARS model to identify variables that have a significant effect on poverty.

2. Literature Review

2.1 Regression Analysis

Regression analysis is a statistical method that explores the relationship between one response variable and one or more explanatory variables (Tiro, 2008). Suppose y is the response variable and x is the predictor variable. For n observations, the general relationship between y and x can be expressed as follows (Budiantara et al., 2006):

$$y_i = g(x_i) + \varepsilon_i \; ; i = 1,2, \dots, n \,, \tag{1}$$

2.2 Nonparametric regression

Nonparametric regression is a method used to identify patterns of relationships between predictor variables and response variables when the form of the function between them is unknown (Nurdiani et al., 2017). The general form of a nonparametric regression model is as follows (Ulfitasari et al., 2025):

$$y_i = f(x_i) + \varepsilon_i \; ; i = 1,2, \dots, n, \tag{2}$$

Description:
y_i = the i-th response variable.
x_i = the i-th predictor variable.
$f(x_i)$ = regression curve to be estimated
ε_i = the i-th observation error is assumed to be identical, independent, and normally distributed, denoted by $\varepsilon_i \sim$ IID $(0,\sigma^2)$

2.3 Spline Regression

Nonparametric spline regression is a regression analysis method that can be used when the pattern of the relationship between the response variable and the predictor variables is unknown in terms of its regression curve form (Ni'matuzzahroh & Dani, 2022). Therefore, in the spline regression equation, knot points $k_{1,}k_{2.} \dots . k_n$are introduced. The function $f(x_i)$ can be expressed by the following equation (Mar'ah et al., 2024):

$$f(x_i) = \beta_0 + \sum_{j=1}^{m} \beta_j x_j^i + \sum_{k=1}^{n} \beta_{m+l} (x_i - k_l)_+^m, \tag{3}$$

where the function $(x_i - k_l)_+^m$ is a truncated (piecewise) function defined as follows (Mar'ah et al., 2024):

$$(x_i - k_l)_+^m = \begin{cases} (x_i - k_l)^m, x_i \geq k_l \\ 0 \qquad\qquad , x_i < k_l \end{cases}, \tag{4}$$

2.4 Multivariate Adaptive Regression Spline

Multivariate Adaptive Regression Spline (MARS) is a nonparametric regression model used to estimate the functional relationship between the response variable and predictor variables when the form of the relationship is unknown (Otok, 2010). Several aspects need to be considered in constructing a MARS model (Friedman, 1991):

a) Knots are points of the independent variables where the slope of the regression line changes, which can be interpreted as the boundary of one segment and the starting point of the next segment. The commonly used minimum number of observations (MO) between knots is 0, 1, 2, and 3 observations.

b) Basis functions (BF) represent the distance between successive knots. The maximum number of basis functions allowed is between 2 to 4 times the number of predictor variables.

c) Interaction is the result of multiplication between variables that have relationships with each other. The maximum allowable level of interaction is 1, 2, or 3. If the interaction level exceeds 3, the resulting model becomes increasingly complex and difficult to interpret.

In general, the MARS model can be written as follows (Mar'ah et al., 2024):

$$f(x_i) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} (S_{km}(x_{v(k,m)} - t_{km})), \tag{5}$$

Description:
a_0 = constant
a_m = coefficient of the m-th basis function
M = maximum number of basis functions (non-constant basis functions)
K_m = degree of interaction
S_{km} = equals 1 if $x$lies to the right of the knot and -1 if $x$lies to the left of the knot
$x_{v(k,m)}$ = the v-th predictor variable, k-th selection, and m-th subregion
t_{km} = knot value of the predictor variable $x_{v(k,m)}$
v = number of predictor variables
k = number of interactions
The MARS model can be simplified as follows (Mar'ah et al., 2024):

$$y_i = a_0 + a_1 BF_1 + a_2 BF_2 + \cdots + a_m BF_m, \tag{6}$$

where y_i is the response variable, a_0 is the constant, a_m adalah is the coefficient of the m-th basis function, and BF_m adalah is the m-th basis function (Mar'ah et al., 2024).

2.5 Parameter Estimation of Multivariate Adaptive Regression Spline

The estimation of the regression curve $f(x_i)$ is generally obtained using the Penalized Least Squares (PLS) approach, based on the regression model in Equation (2.5). The matrix notation of the MARS model can be expressed as follows (Otok et al., 2012):

$$Y = \mathbf{Ba} + \boldsymbol{\epsilon}, \tag{7}$$

2.6 Selection of the Best MARS Model

Model selection in MARS is performed using a stepwise method consisting of forward and backward procedures. The principle of parsimony states that a simpler model is generally preferred over a model with more parameters (Aswi & Sukarna, 2006). According to Friedman (1991), to satisfy the parsimony concept, a backward stepwise procedure is applied by selecting basis functions generated from the forward stepwise stage while minimizing the Generalized Cross Validation (GCV) criterion (Oktora, 2015).

a. Forward Stepwise

The forward stepwise procedure aims to obtain a model with the maximum number of basis functions. The steps of the forward stepwise method in MARS are as follows (Zhang & Singer, 2010):

1. Assume $B_0 = 1$, which is a constant basis function, as the initial basis function.
2. Determine a pair of basis functions $B_1 = (x_i - t)_+$ dan $B_2 = (t - x_i)_+$ a combination of the predictor variable x_i and knot t_1 to be added to the model. This step produces a MARS model with the minimum Average Square Residual (ASR) value.
3. The next step is to expand the MARS model by adding the product of existing basis functions B_m with each new basis function into the current model, resulting in several possible models. Therefore, the pair of products that produces the model with the smallest ASR value is selected.
4. Repeat step (3) until the number of basis functions in the model is greater than or equal to the predetermined maximum number of basis functions.

The basis function used is the hinge function. The hinge function has the following (Friedman, 1991):

$$h(x) = \max(0, x - c) \quad \text{or} \quad h(x) = \max(0, c - x) \tag{8}$$

Where:
x = predictor variable
c = specified cutoff (knot) point

b. Backward Stepwise

The backward stepwise procedure aims to obtain the simplest possible model (parsimony principle). This process starts from the model obtained in the forward stepwise stage, which contains M basis functions, with the following steps (Zhang & Singer, 2010):

1. Remove one non-constant basis function that has the smallest contribution, namely the basis function whose removal causes the smallest increase in ASR.
2. Repeat step (1) until the model contains only the constant basis function.

Next, the best model is selected based on the Generalized Cross Validation (GCV) value, where the best model is the one with the smallest (minimum) GCV among all candidate models (Pintowati & Otok, 2012). The GCV function can be expressed as follows (Pintowati & Otok, 2012):

$$GCV(M) = \frac{ASR}{\left[1-\frac{\tilde{C}(M)}{n}\right]^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}_M(x_i)]^2}{\left[1-\frac{\tilde{C}(M)}{n}\right]^2}, \tag{9}$$

2.7 Significance Testing of Model Parameters

a. Simultaneous (Overall) Test

The simultaneous test is conducted to determine whether all predictor variables in the model jointly influence the response variable (Rositawati & Budiantara, 2020). The hypotheses are as follows:

$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_i = 0$

H_1 : at least one $\alpha_i \neq 0,\ i = 1,2,\ldots,i$

The F-test statistic used is as follows (Wicaksono et al., 2014):

$$F_{\text{calculated}} = \frac{MS_{regression}}{MS_{residual}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 / M}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / N - M - 1}, \tag{10}$$

The rejection region is : reject H_0 if $F_{calculated} > F_{\alpha;(M;N-M-1)}$ or if the $p - value < \alpha$ (Wicaksono et al., 2014).

b. Partial (Individual) Testing

The partial test is used to determine whether individual parameters have a significant effect on the model using the t-test (Fajriyyah & Budiantara, 2015). The hypotheses are as follows:

$H_0 : \alpha_i = 0$

$H_1 : \alpha_i \neq 0, i = 1,2, \dots, i$

The test statistic used is (Wicaksono et al., 2014):

$$t_{\text{calculated}} = \frac{\hat{a}_m}{se(\hat{a}_m)}, \tag{11}$$

The rejection region is : reject H_0 if $t_{calculated} > F_{\frac{\alpha}{2};(N-M)}$ or if the $p - value < \alpha$ (Wicaksono et al., 2014).

2.8 Coefficient of Determination (R^2)

The coefficient of determination is a value that indicates the proportion of the total variability around the mean $\bar{y}$ yang that can be explained by the regression model. The formula for the coefficient of determination (R^2) is as follows (Darma et al., 2019):

$$R^2 = \frac{SS_{regression}}{SS_{total}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{12}$$

2.9 Poverty

Poverty is a condition in which an individual or group lacks sufficient material resources, such as money, food, shelter, or access to adequate healthcare and education, to meet their basic needs (Irfan et al., 2024). According to the World Bank, poverty should be understood as a multidimensional phenomenon by considering various factors contributing to inequality and marginalization (Purwanti, 2024). The World Bank defines the poverty line as an income of $2 per day, or approximately IDR 22,000 per day (Irfan et al., 2024).

3. Research Methods

This study uses secondary data obtained online from the official website of the Statistics Indonesia (BPS) (https://www.bps.go.id/id), in the form of data on the percentage of poor population in 2024 covering 38 provinces in Indonesia. The dependent variable used is the Percentage of Poor Population (y) and 7 independent variables are included: Open Unemployment Rate (x_1), Labor Force Participation Rate (x_2), Percentage of Formal Sector Workers (x_3), Percentage of Households with Access to Proper Sanitation(x_4), Mean Years of Schooling (x_5), Gross Regional Domestic Product (GRDP) Growth Rate (x_6), and Gini Ratio (x_7). The research procedure consists of literature review, data collection, data analysis, and drawing conclusions as well as preparing the research report.

The data analysis techniques used in this study are as follows:

1. Conducting descriptive statistical analysis and creating scatter plots between the dependent variable and each independent variable to observe data patterns.
2. Standardizing the data.
3. Performing MARS (Multivariate Adaptive Regression Splines) modeling:

   a. Determining the maximum number of Basis Functions (BF), which is set to 2–4 times the number of independent variables.
   b. Determining the maximum degree of interaction (MI), namely 1, 2, and 3 interactions. Interactions greater than 3 may lead to highly complex model interpretation.
   c. Determining the minimum number of observations (MO), namely 0, 1, 2, and 3.
4. Selecting the MARS model with the smallest Generalized Cross Validation (GCV) value obtained from the combination of Basis Functions (BF), maximum interactions (MI), and minimum observations (MO).
5. Conducting significance tests for the MARS model
   a. Performing a simultaneous test using the F-test to examine whether the MARS model is significant overall.
   b. Performing a partial test using the t-test to examine whether each basis function (BF) or variable contributes significantly to the model.
6. Obtaining the best MARS model and identifying the level of contribution of independent variables influencing the model.
7. Calculating the coefficient of determination (R^2).
8. Interpreting the best MARS model and the influential variables within the model.

4. Results and Discussion

4.1. Description of Poverty in Indonesia in 2024

Before conducting further analysis, all variables used in this study are described descriptively through their mean, minimum, maximum, and standard deviation values. The detailed description is presented as follows.

Table 1. Descriptive Statistics of Each Variable

Variable	Mean	Standard Deviation	Minimum	Maximum
y	10,67	6,37	3,8	29,66
x_1	4,38	1,41	1,32	6,75
x_2	70,7	4,31	65,1	88,2
x_3	40,5	11,9	4,24	68,45
x_4	81,14	15,07	12,61	96,83
x_5	8,84	1,28	4,21	11,49
x_6	5,45	3,17	0,77	20,8
x_7	0,34	0,05	0,235	0,431

Table 1 shows that the average percentage of poor population (y) in Indonesia is 10,67%, with a standard deviation of 6,37%. The lowest percentage is observed in Bali (3,8%), while the highest is found in Highland Papua (29,66%). The data visualization is presented in Figure 1.

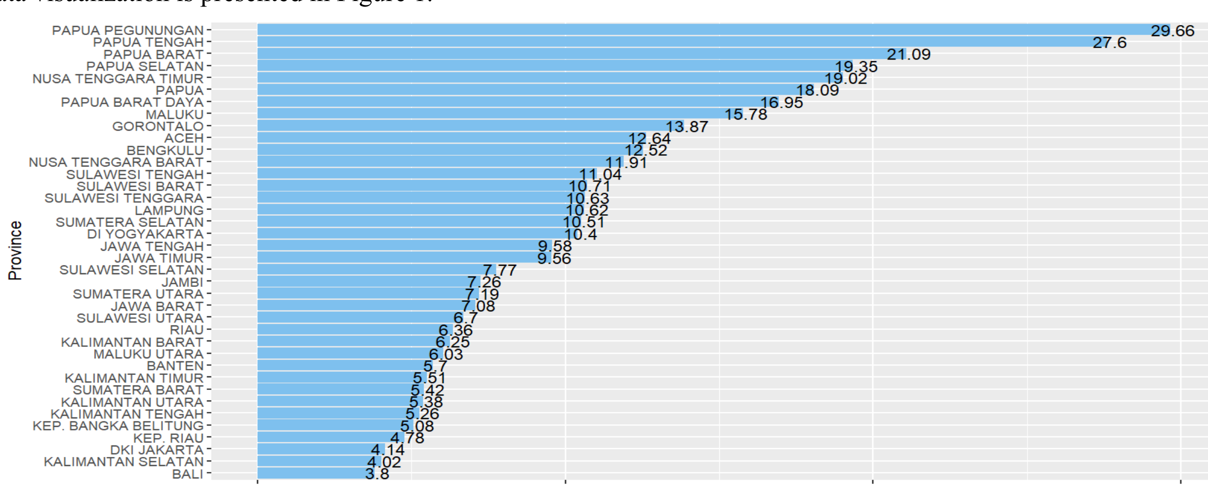

Figure 1. Percentage of Poor Population in Indonesia

4.2. The Relationship Pattern Between the Percentage of Poor Population and the Independent Variables

The following describes the relationship pattern between the percentage of poor population and the seven independent variables used in this study.

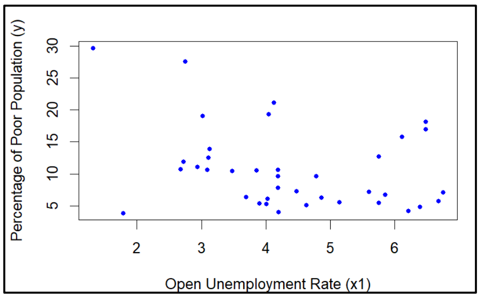

Figure 2. Scatter plot between P_0 (y) and the Open Unemployment Rate (x_1)

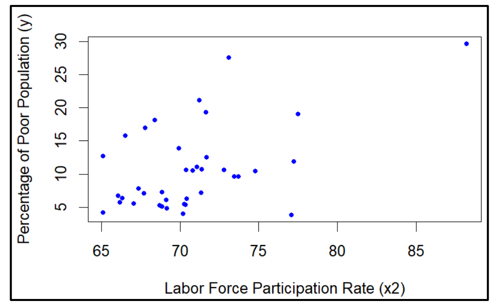

Figure 3. Scatter plot between P_0 (y) and the Labor Force Participation Rate (x_2)

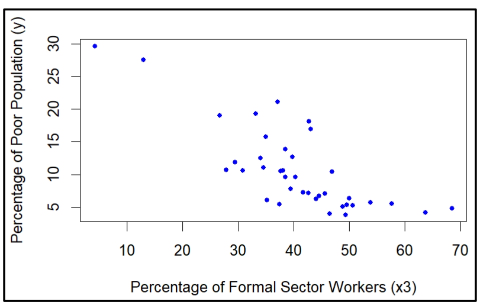

Figure 4. Scatter plot between P_0 (y) and the Percentage of Formal Sector Workers (x_3)

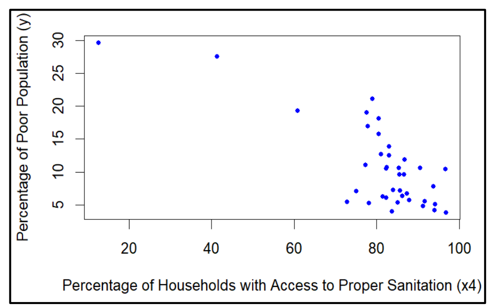

Figure 5. Scatter plot between P_0 (y) and the Percentage of Households with Access to Proper Sanitation (x_4)

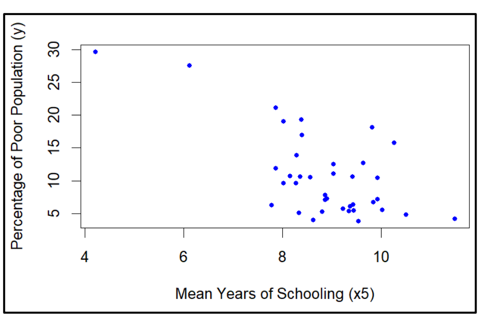

Figure 6. Scatter plot between P_0 (y) and Mean Years of Schooling (x_5)

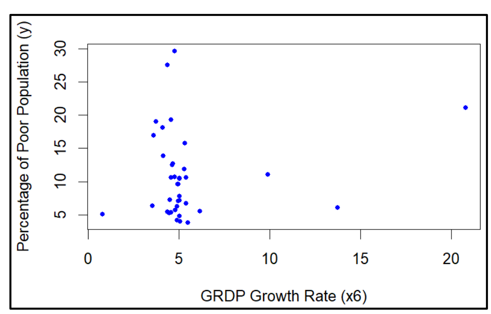

Figure 7. Scatter plot between P_0 (y) and the GRDP Growth Rate (x_6)

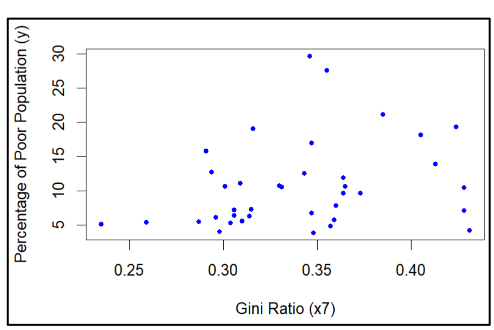

Figure 8. Scatter plot between P_0 (y) and the Gini Ratio (x_7)

Based on Figures 2 to 8, the generated scatter plots indicate that there is no clear pattern formed between the Percentage of Poor Population and each of the independent variables used. Therefore, a nonparametric regression approach can be applied to model the data using the Multivariate Adaptive Regression Splines (MARS) method (Hutabarat & Sitepu, 2024).

4.3. Estimation of MARS Modeling

The first step in constructing the MARS model is to determine the combination of the number of Basis Functions (BF), Maximum Interaction (MI), and Minimum Observation (MO). Since there are 7 independent variables used, the number of BF considered is 14, 21, and 28. The MI values used are 1, 2, and 3, while the MO values are 0, 1, 2, and 3. After determining these parameters, the model is developed using forward stepwise and backward stepwise algorithms to obtain the best model from the combination of BF, MI, and MO (Mattalunru et al., 2022).

From the combinations of BF, MI, and MO that have been processed through forward stepwise and backward stepwise algorithms, a total of 36 models are obtained. The best model is selected based on the minimum Generalized Cross Validation (GCV) value. The results of the MARS modeling are presented in Table

Table 2. Results of MARS Modeling with Combinations of BF, MI, and MO

Model	BF	MI	MO	GCV
1	14	1	0	0,3108188
2	14	1	1	0,3497928
3	14	1	2	0,3518857
4	14	1	3	0,3472869
5	14	2	0	0,3985602
6	14	2	1	0,3567274
7	14	2	2	0,3699154
8	14	2	3	0,3699154
9	14	3	0	0,3985602
10	14	3	1	0,3567274
11	14	3	2	0,3699154
12	14	3	3	0,3699154
13	21	1	0	0,336207
14	21	1	1	0,3218668
15	21	1	2	0,3738936
16	**21**	**1**	**3**	**0,3102717**
17	21	2	0	0,3743491
18	21	2	1	0,3589859
19	21	2	2	0,3333798
20	21	2	3	0,3357603
21	21	3	0	0,3743491
22	21	3	1	0,3278361
23	21	3	2	0,3333798
24	21	3	3	0,3357603
25	28	1	0	0,336207
26	28	1	1	0,3218668
27	28	1	2	0,3738936
28	**28**	**1**	**3**	**0,3102717**
29	28	2	0	0,3743491
30	28	2	1	0,3589859
31	28	2	2	0,3333798
32	28	2	3	0,3357603
33	28	3	0	0,3743491
34	28	3	1	0,3150168

35	28	3	2	0,3333798
36	28	3	3	0,3357603

After performing MARS modeling with combinations of BF, MI, and MO, Table 2. presents the modeling results showing that the minimum GCV value is obtained for the combinations BF = 21, MI = 1, MO = 3 and BF = 28, MI = 1, MO = 3, both yielding the same GCV value of 0,3102717. Therefore, the best MARS model is determined to be the combination BF = 21, MI = 1, MO = 3. This model is selected based on the principle of parsimony, as it has lower complexity while providing results equivalent to a more complex model (Putra et al., 2021).

4.4. MARS Model Significance Test

a. Simultaneous Testing

The results of the simultaneous significance test are presented in Table 3. as follows:

Table 3. Results of Simultaneous Significance Test

Source	**df**	**SS**	**MS**	$F_{calculated}$	$p-value$
Regression	4	30,1332	7,5333	36,203	$1{,}232e-11$
Residual	33	6,8668	0,2081		
Total	37	37			

Based on Table 3. the $p-value$ $(1{,}232e-11) < 0{,}05$ thus the decision is to reject H_0. It can be concluded that the independent variables have a significant effect on the dependent variable simultaneously, indicating that the model is appropriate for modeling the percentage of poor population.

b. Partial Testing

The results of the partial significance test are presented in Table 4. as follows:

Table 4. Results of Partial Significance Test

Parameter	**Estimate**	$p-value$	**Remark**
Intercept	−0,29119	0,12912	Not Significant
BF_1	0,87420	$3{,}11e-09$	Significant
BF_2	−5,75608	0,00255	Significant
BF_3	5,547074	0,00786	Significant
BF_4	0,30867	0,00255	Significant

Based on Table 4. all BF parameters have $p-value < 0{,}05$ leading to the rejection of H_0. This indicates that the parameters BF_1, BF_2, BF_3, and BF_4 individually have a significant effect on the dependent variable.

4.5. Best MARS Model

The best MARS model is obtained with BF = 21, MI = 1, MO = 3, with a GCV value of 0,3102717. This model satisfies both simultaneous and partial significance criteria. The model equation based on the parameter estimates in Table 4 is as follows:

$$y_i = -0{,}2911927 + 0{,}8742022\ BF_1 - 5{,}75608\ BF_2 + 5{,}547074\ BF_3 + 0{,}3086743\ BF_4$$

Where:

BF_1 = max(0, $0{,}426164 - Zx_3$)
BF_2 = max(0, $Zx_4 - (-0{,}0381354)$)
BF_3 = max(0, $Zx_4 - 0{,}779819$)
BF_4 = max(0, $Zx_7 - (-0{,}662306)$)

Based on the form of the model equation, there are four basis functions involved in constructing the model. The variable selection results indicate that the model selects 3 out of 7 predictor variables that contribute to explaining the response variable, namely the Percentage of Formal Sector Workers (x_3), the Percentage of Households with Access to Proper Sanitation (x_4), and the Gini Ratio (x_7).

The contribution level of each variable in the MARS model is presented in Table 5.

Table 5. Contribution of Significant Variables in the MARS Model

Variable	Contribution Level
Percentage of Formal Sector Workers (x_3)	100%
Gini Ratio (x_7)	46,8%
Percentage of Households with Access to Proper Sanitation (x_4)	35,8%

Table 5. shows that the Percentage of Formal Sector Workers (x_3) has a contribution of 100% in influencing the percentage of poor population, indicating that this variable provides the largest contribution to poverty in Indonesia. Meanwhile, the Gini Ratio (x_7) has a contribution level of 46.8%, and the Percentage of Households with Access to Proper Sanitation (x_4) has a contribution level of 35.8%.

4.7. Coefficient of Determination (R^2)

The R^2 value of the best model is obtained as follows:

$$R^2 = \frac{SS_{regression}}{SS_{total}} \times 100\%$$
$$= \frac{30{,}1332}{37} \times 100\%$$
$$= 81{,}44\%$$

The result shows that 81.44% of the variation in the percentage of poor population is explained by the Percentage of Formal Sector Workers (x_3), the Percentage of Households with Access to Proper Sanitation (x_4), and the Gini Ratio (x_7), while the remaining 18.56% is influenced by other factors.

4.8. Interpretation of Each Basis Function in the MARS Model

The interpretation of each basis function coefficient in the MARS model is as follows:

1. $BF_1 = \max(0, 0{,}426164 - Zx_3)$

$$= \begin{cases} (0{,}426164 - Zx_3), if\ Zx_3 < 0{,}426164 \\ 0, if\ Zx_3 \geq 0{,}426164 \end{cases}$$

The coefficient of BF_1 is 0,8742022. This means that for every one-unit increase in BF_1 the percentage of poor population in Indonesia increases by 0,8742022 assuming other basis functions remain constant. BF_1 takes the value $(0{,}426164 - Zx_3)$ when the Percentage of Formal Sector Workers (x_3) is less than 0,426164. Otherwise, when x_3 is greater than or equal to 0,426164, BF_1 equals zero and has no effect. This indicates that regions with a lower proportion of formal sector workers tend to experience higher levels of poverty, as opportunities for obtaining decent jobs are still limited. Furthermore, a higher number of workers employed below standard working hours may increase the proportion of working poor due to insufficient income to meet living needs (Pangestu, 2019).

2. $BF_2 = \max(0,\ Zx_4 - (-0{,}0381354))$

$$= \begin{cases} (Zx_4 - (-0{,}0381354)), if\ Zx_4 > -0{,}0381354 \\ 0, if\ Zx_4 \leq -0{,}0381354 \end{cases}$$

BF_2 has a coefficient of $-5{,}75608$. This implies that for every one-unit increase in BF_2 the percentage of poor population in Indonesia decreases by $-5{,}75608$ assuming other basis functions remain constant. BF_2 takes the value $(Zx_4 - (-0{,}0381354))$ when the Percentage of Households with Access to Proper Sanitation (x_4) lebih is greater than $-0{,}0381354$. Otherwise, when x_4 lis less than or equal to $-0{,}0381354$, BF_2 equals zero and has no effect. This indicates that the availability of proper sanitation facilities in a region contributes positively to public health and labor productivity, thereby improving economic opportunities. Conversely, inadequate sanitation facilities may exacerbate poverty by reducing productivity and increasing healthcare costs (WHO, 2023).

3. $BF_3 = \max(0,\ Zx_4 - 0{,}779819)$

$$= \begin{cases} (Zx_4 - 0{,}779819), if\ Zx_4 > 0{,}779819 \\ 0, if\ Zx_4 \leq 0{,}779819 \end{cases}$$

BF_3 has a coefficient of $5{,}547074$. This means that for every one-unit increase in BF_3 the percentage of poor population in Indonesia increases by $5{,}547074$ assuming other basis functions remain constant. BF_3 takes the value $(Zx_4 - 0{,}779819)$ when the Percentage of Households with Access to Proper Sanitation (x_4) is greater than $0{,}779819$. Otherwise, when x_4 is less than or equal to $0{,}779819$, BF_3 equals zero and has no effect. This result may appear counterintuitive, as an increase in access to proper sanitation is generally expected to reduce poverty levels. However, this condition does not always lead to poverty reduction if it is not accompanied by proper environmental management. In particular, if household waste disposal systems are not properly managed, this may lead to soil contamination and reduced land quality. As a consequence, communities that rely on agriculture or plantations may lose their source of income, potentially increasing poverty levels (Sitorus & Simamora, 2023).

4. $BF_4 = \max(0,\ Zx_7 - (-0{,}662306))$

$$= \begin{cases} (Zx_7 - (-0{,}662306)), if\ Zx_7 > -0{,}662306 \\ 0, if\ Zx_7 \leq -0{,}662306 \end{cases}$$

BF_4 has a coefficient of $0{,}3086743$. his indicates that for every one-unit increase in BF_4 the percentage of poor population in Indonesia increases by $0{,}3086743$ assuming other basis functions remain constant. BF_4 takes the value $(Zx_7 - (-0{,}662306))$ when the Gini Ratio (x_7) is greater than $-0{,}662306$. Otherwise, when x_7 is less than or equal to $-0{,}662306$, BF_4 equals zero and has no effect. Hal Ini his finding indicates that a higher Gini Ratio reflects greater income inequality within society, which in turn can influence the level of poverty (Fadila et al., 2023).

5. Conclusion and Suggestions

5.1 Conclusion

- The best Multivariate Adaptive Regression Splines (MARS) model is obtained with BF = 21, MI = 1, MO = 3, and a GCV value of 0,3102717. The resulting model equation is as follows:

$$y_i = -0{,}2911927 + 0{,}8742022\ BF_1 - 5{,}75608\ BF_2 + 5{,}547074\ BF_3 + 0{,}3086743\ BF_4$$

Where:

$BF_1 = \max(0,\ 0{,}426164 - Zx_3)$
$BF_2 = \max(0,\ Zx_4 - (-0{,}0381354))$
$BF_3 = \max(0,\ Zx_4 - 0{,}779819)$
$BF_4 = \max(0,\ Zx_7 - (-0{,}662306))$

- The independent variables that significantly influence the percentage of poor population in Indonesia are the Percentage of Formal Sector Workers (x_3), the Percentage of Households with Access to Proper Sanitation (x_4), and the Gini Ratio (x_7) with a coefficient of determination (R^2) of 81,44%.

5.2 Suggestions

Future research is encouraged to include additional variables to improve model performance. In addition, the government is expected to increase the number of formal sector workers, expand access to proper sanitation, and reduce income inequality (Gini Ratio). Policies focused on employment sectors, basic infrastructure, and equitable development are expected to be more effective in reducing poverty levels.

References

Anam, S., Sugiman, & Sunarmi. (2017). Ketepatan Klasifikasi dengan Menggunakan Metode Multivariate Adaptive Regression Spline (MARS) pada Data Kelompok Rumah Tangga Kabupaten Cilacap. *Unnes Journal of Mathematics*, *6*(1), 92–101. http://journal.unnes.ac.id/sju/index.php/ujmUJM6

Astiti, D. A. W., Sumarjaya, I. W., & Susilawati, M. (2016). Analisis Regresi Nonparametrik Spline Multivariat Untuk Pemodelan Indikator Kemiskinan Di Indonesia. *E-Jurnal Matematika*, *5*(3), 111–116. https://doi.org/10.24843/mtk.2016.v05.i03.p129

Aswi & Sukarna. (2006). Analisis Data Deret Waktu : Teori dan Aplikasi (1st ed.). Andira Publisher.

BPS. (2023). Profil kemiskinan di Indonesia 2023. https://www.bps.go.id

Budiantara, I. N., Suryadi, F., Otok, B. W., & Guritno, S. (2006). Pemodelan B-Spline Dan Mars Pada Nilai Ujian Masuk Terhadap Ipk Mahasiswa Jurusan Disain Komunikasi Visual Uk. Petra Surabaya. *Jurnal Teknik Industri*, *8*(1), 1–13. https://doi.org/10.9744/jti.8.1.1-13

Darma, I. G. P. S., Ratna, M., & Budiantara, I. N. (2019). Pemodelan Faktor yang Memengaruhi Angka Kasus Tuberculosis di Kota Surabaya Menggunakan Regresi Nonparametrik Spline Truncated. *Jurnal Sains Dan Seni ITS*, *8*(2), 216–222. https://doi.org/10.12962/j23373520.v8i2.45404

Eubank, R. L. (1999). Nonparametric Regression and Spline Smoothing, 2nd ed. New York: Marcel Dekker.

Fadila, F., Ulhaq, N., Khoir, R., & Fatih, S. (2023). Analisis Data Panel Dalam Memprediksi Faktor determinan Kemiskinan di Provinsi Kalimantan Selatan. *JIEP: Jurnal Ilmu Ekonomi Dan Pembangunan*, *6*(1), 303–314.

Fajriyyah, N., & Budiantara, I. N. (2015). Pemodelan indeks pembangunan gender dengan pendekatan regresi nonparametrik spline di Indonesia. *Jurnal Sains Dan Seni ITS*, *4*(2), 217–222.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, *19*(1), 1–141.

Hutabarat, J. A. R., & Sitepu, S. (2024). Model Multivariate Adaptive Regression Spline (MARS) Pada Persamaan Regresi Nonparametrik (Studi Kasus : Pertumbuhan Ekonomi di Sumatera Utara) Suryati Sitepu menyelesaikan suatu masalah dalam berbagai bidang baik dalam bidang sosial , bahkan Salah sa. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam (JURRIMIPA)*, *3*(1), 186–196.

Iskandar, A., & Subekan, A. (2016). Analisis Determinan Kemiskinan Di Sulawesi Selatan (Determinant Analysis of Poverty in South Sulawesi). *Jurnal Tata Kelola Dan Akuntabilitas Keuangan Negara*, *2*(1), 1–25. https://doi.org/10.28986/jtaken.v2i1.36

Mahmud, F., Olilingo, F. Z., & Akib, F. H. Y. (2020). Faktor-Faktor Yang Mempengaruhi Kemiskinan di Pulau Sulawesi. *Oikos Nomos: Jurnal Kajian Ekonomi Dan Bisnis*, *13*(2), 130–147. https://doi.org/10.37479/jkeb.v13i2.11872

Mar'ah, Z., Ruliana, & Septiana, M. (2024). Pemodelan Multivariate Adaptive Regression Spline (Mars) Pada Indeks Harga Saham Gabungan (IHSG) Tahun 2018 – 2023. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research Vol.*, *6*(1), 1–10. https://doi.org/10.35580/variansiunm151

Mattalunru, M. R., Annas, S., & Aidid, M. K. (2022). Aplikasi Multivariate Adaptive Regression Splines (Mars) Untuk Mengetahui Faktor Yang Mempengaruhi Curah Hujan Di Kota Makassar. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(1), 9–19. https://doi.org/10.35580/variansiunm2

Ni'matuzzahroh, L., & Dani, A. T. R. (2022). Pemodelan Rata-Rata Lama Sekolah Di Provinsi Nusa Tenggara Timur (Ntt) Menggunakan Pendekatan Regresi Nonparametrik Spline …. *Prosiding Seminar Nasional …*, 289–301.

Oktora, S. I. (2015). Analisis Multivariate Adaptive Regression Splines (Mars) Pada Prediksi Ketertinggalan Kabupaten Tahun 2014. *Jurnal Aplikasi Statistika & Komputasi Statistik*, *7*(2), 1–7.

Otok, B. W. (2010). Pendekatan Multivariate Adaptive Regression Spline (MARS) pada Pengelompokkan Zona Musim Suatu Wilayah. *Statistika*, *10*(2), 107–120.

Otok, B. W., Guritno, S., & Subanar. (2012). Asimtotik Model Multivariate Adaptive Regression Spline. *Jurnal Natur Indonesia*, *10*(2), 112–119. https://doi.org/10.31258/jnat.10.2.112-119

Pangestu, I. M. (2019). Fenomena Pekerja Miskin dan Faktor-Faktor yang Memengaruhi Persentase Pekerja Miskin di Indonesia Tahun 2015-2018. *Jurnal Ekonomi Dan Kebijakan Pembangunan*, *8*(1), 62–74. https://doi.org/10.29244/jekp.v8i1.26691

Pintowati, W., & Otok, B. W. (2012). Pemodelan Kemiskinan di Propinsi Jawa Timur dengan Pendekatan Multivariate Adaptive. *Jurnal Sains Dan Seni ITS*, *1*(1), 283–288. http://www.ejurnal.its.ac.id/index.php/sains_seni/article/view/2072

Putra, R. Y., Roza, A., & Putri, H. M. (2021). Pendekatan Multivariate Adaptive Regression Splines Untuk Memodelkan Tingkat Kemiskinan Di Provinsi Sumatera Barat. *MAp (Mathematics and Applications) Journal*, *3*(2), 92–98. https://doi.org/10.15548/map.v3i2.3143

Rositawati, A. F. D., & Budiantara, I. N. (2020). Pemodelan Indeks Kebahagiaan Provinsi di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated. *Jurnal Sains Dan Seni ITS*, *8*(2), 113–120. https://doi.org/10.12962/j23373520.v8i2.45160

Septa, A., Fairuzani, A., Yasmin, C., Hutagaol, G., & Hutabarat, V. (2025). Analisis Variabel Kemiskinan di Indonesia dengan Model Linear Regresi dan Algoritma K-Means pada Tahun 2020-2023 Gambar 1 . Perkembangan Kemiskinan Nasional dan Ekstrem di Indonesia seseorang untuk memenuhi kebutuhan pokok hidupnya . Garis kemiskinan ini. *Profit: Jurnal Manajemen, Bisnis Dan Akuntansi*, *4*(2), 28–52.

Sitorus, A. L. T., & Simamora, E. (2023). Metode Geographically Weighted Panel Regression (GWPR) Untuk Menganalisis Faktor Yang Mempengaruhi Kemiskinan Di Provinsi Sumatera Utara. *Ranah Research : Journal of Multidisciplinary Research and Development*, *6*(1), 155–167. https://doi.org/10.38035/rrj.v6i1.808

Tiro, M. A. (2008). Analisis Regresi dengan Data Kategori (Edisi Ketiga). Andira Publisher.

Ulfitasari, Yonita, A., Siswanto, & Kalondeng, A. (2025). Pemodelan Indeks Kebahagiaan Negara Dengan Metode Multivariate Adaptive Regression Spline. *Jurnal Ilmiah Matematika*, *13*(1), 209–216. https://media.neliti.com/media/publications/249234-model-infeksi-hiv-dengan-pengaruh-percob-b7e3cd43.pdf

WHO. (2023). *Tracking universal health coverage 2023 global monitoring report*. https://doi.org/https://doi.org/10.1596/40348

Wicaksono, W., Wilandari, Y., Suparti, & Pengajar Jurusan Statistika, S. (2014). Pemodelan Multivariate Adaptive Regression Splines (Mars) Pada Faktor-Faktor Resiko Angka Kesakitan Diare (Studi Kasus : Angka Kesakitan Diare Di Jawa Tengah, Jawa Timur Dan Daerah Istimewa Yogyakarta Tahun 2011). *Jurnal Gaussian*, *3*(2), 253–262. http://ejournal-s1.undip.ac.id/index.php/gaussian

Wulandari, G., Febriyanti, N. A., Anwar, K., & Nohe, D. A. (2022). Pemodelan Persentase Penduduk Miskin Di Indonesia Menggunakan Regresi Probit Dan Regresi Logistik. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya Terbitan II*, *2*, 354–368.

Zhang. H. & Singer. B. H. (2010). Recursive Partitioning and Applications 2^{nd} Edition. New York: Springer