Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1478~1485
e-ISSN: 2548-6861

Analysis of Factors Influencing Poverty in Indonesia Using the Bootstrap Aggregating Multivariate Adaptive Regression Spline

Lifeni Triara *, Devni Prima Sari [1]*
* Department of Mathematics, Universitas Negeri Padang
devniprimasari@fmipa.unp.ac.id

Article Info

Article history:

Received 2026-02-03
Revised 2026-03-17
Accepted 2026-04-10

Keyword:

Bootstrap Aggregating,
MARS,
Poverty,
Regression.

ABSTRACT

Poverty is a multidimensional problem faced by Indonesia. Poverty is caused by various interrelated factors, such as insufficient income, unequal access to education, health services, employment opportunities, and social participation. Therefore, this study aims to analyze the factors most influential on poverty at the district/city level in Indonesia using the Bagging MARS method. This applied study begins with an analysis of relevant theories and continues with data collection. The data used are secondary data obtained from the 2024 publication of the Central Statistics Agency (BPS), comprising 13 predictor variables covering indicators of education, employment, per capita expenditure, and housing facilities. The analysis method used was Bootstrap Aggregating Multivariate Adaptive Regression Spline (Bagging MARS). This method involved building a MARS model and subsequently applying bagging resampling to improve accuracy. The MARS model was trained with 39 basis functions (BF), a maximum interaction degree (MI) of 3, and a minimum observation between knots (MO) of 3, followed by 50 bagging resamples. The best Bagging MARS model was obtained on the 42nd iteration, with a minimum Generalized Cross Validation (GCV) value of 9.207091. Of the 13 variables analyzed, the variable that had the greatest influence on poverty levels was the percentage of school enrolment among poor residents aged 7 – 12 years (X_4), with a significance level of 100%.

This is an open access article under the CC–BY-SA license.

I. INTRODUCTION

Poverty is a multidimensional problem faced by Indonesia. Poverty here does not only refer to the lack of income among the people, but also various important aspects such as unequal access to education, health services, employment opportunities, and social and political participation [1]. These multidimensional aspects indicate that poverty is a complex phenomenon influenced by various socioeconomic factors that interact with each other.

According to the Central Statistics Agency (BPS), the number of poor people in Indonesia in September 2024 was 24.06 million. This figure is down by 1.84 million people compared to March 2023 [2]. The downward trend in poverty rates in recent years is a positive step in efforts to tackle poverty. Based on World Bank data for 2024, the mode that 60.3% of Indonesia's population, or around 171.8 million people, live below the poverty line for upper-middle-income countries [3]. This shows that even though there has been a downward trend in poverty rates in Indonesia, these rates are still relatively high compared to other countries, such as the Philippines at 50.6% and Vietnam at 18.2% [3]. These findings indicate that despite improvements, poverty reduction remains an important issue requiring further policy attention and effective analytical approaches.

Poverty is caused by a large population that is not matched by an increase in human resources. Educational inequality results in individuals lacking skills [4]. This makes it difficult for individuals to find decent work. Poverty has a broad and complex impact on people's lives. One of the most common effects is unemployment, a condition in which individuals have no income and therefore earn no money [5]. High unemployment rates place pressure on households, as low incomes can lead to poverty. In addition, low labor quality due to a lack of education and training can hinder an individual's ability to compete in the job market and earn an adequate income [6].

Furthermore, poverty not only affects economic access but also overall quality of life. individuals living in poverty often face barriers to adequate healthcare, education, and other essential services [7]. This condition has the potential to create a cycle of intergenerational poverty that is difficult to break. Therefore, to assist the government in its efforts to overcome poverty in Indonesia, it is necessary to identify the factors that influence poverty in Indonesia.

Poverty is influenced by multiple interrelated factors. To determine the relationship between these factors, regression analysis is performed. Regression analysis is a statistical method used to describe the relationship between response variables and predictor variables [8]. There are two ways to estimate the regression function [9]. Parametric regression is a statistical method used to determine the relationship pattern between response variables and predictor variables, assuming that the regression curve shape is known [10]. Non-parametric regression is an approach used to determine and estimate the relationship pattern between response variables and predictor variables when the curve shape is not clearly known [11]. If the parametric regression model approach and the non-parametric regression model approach are combined, it is called a semiparametric regression model [12].

In this study, a nonparametric regression approach was selected due to the limited prior knowledge regarding the functional form of the relationship between poverty and its predictor variables. The method used is Multivariate Adaptive Regression Spline (MARS), which is a flexible nonparametric regression technique capable of modelling nonlinear relationships and interactions between variables. MARS is particularly suitable for handling high-dimensional data involving multiple predictor variable namely, with a number of predictors between 3 and 20 variables and a sample size of 50 – 100 data points [13]. The MARS method constructs models using piecewise spline functions and automatically determines optimal knot locations and interaction effects through forward and backward stepwise procedures. This allows MARS to effectively capture complex patterns in socioeconomic data. The MARS approach combines Recursive Partitioning Regression (RPR) with the Spline approach. RPR is a non-parametric method considered to be a piecewise regression that excels in processing high-dimensional data and approximating unknown functions [13]. A spline is a piecewise polynomial function or a segmented or fragmented and continuous polynomial [14].

To further improve the stability and predictive accuracy of the MARS model, this study integrates Bootstrap Aggregating (Bagging), a resampling-based ensemble technique. Bagging works by generating multiple bootstrap samples from the original dataset and building separate models for each sample, which are then aggregated to produce a final model [15]. This approach reduces model variance, improves stability, and enhances predictive performancec

Therefore, this study aims to identify factors influencing poverty using the Bagging MARS approach. The results of this study are expected to provide valuable insights for policymakers in designing effective and targeted poverty alleviation strategies.

II. METHOD

This type of research is applied research. Applied research is research conducted systematically to apply or develop the results of previous research in order to solve real problems in society [17]. The study uses secondary data from the 2024 publication of the Central Statistics Agency (BPS). The dataset consists of 514 observations representing districts and cities across all provinces in Indonesia. These poverty data represent the percentage of people living in poverty in Indonesia for each district/city, and 13 influencing variables are described in Table 1. The selection of 13 predictor variables in this study was based on socioeconomic indicators commonly associated with poverty. Previous studies have shown that factors related to education. Employmenr, household expenditure, and housing facilities play a significant role in determining poverty levels. Therefore, these variables were selected to represent various socioeconomic dimensions that may influence poverty across districts and cities in Indonesia.

TABLE I
RESEARCH INDICATORS AND VARIABLES

Indicator	Variable
Education	X_1: Percentage of poor people aged 15 years and above who did not complete elementary school.
	X_2: Percentage of poor people aged 15 years and above with a final certificate of elementary school/junior high school.
	X_3: Percentage of poor people aged 15 years and above with a final certificate of senior high school or college.
	X_4: Percentage of school enrollment among poor people aged 7 – 12 years.
	X_5: Percentage of school enrollment among poor people aged 13 – 15 years.
Employment	X_6: Percentage of poor people aged 15 years and above who are unemployed.
	X_7: Percentage of poor people aged 15 years and above who work in the informal sector.
	X_8: Percentage of poor people aged 15 years and above who work in the formal sector.
	X_9: Percentage of poor people aged 15 years and above who work in the agricultural sector.
	X_{10}: Percentage of poor people aged 15 years and above who work outside the agricultural sector.
Per Capita Expenditure	X_{11}: Percentage of per capita expenditure on food for the poor.

Indicator	Variable
Housing Facilities	X_{12}: Percentage of poor households that use clean water.
	X_{13}: Percentage of poor households using their own or shared toilets.

The selection of 13 predictor variables was based on indicators commonly used in poverty analysis by the Central Statistics Agency (BPS) and previous socioeconomic studies. These variables represent several important dimensions of poverty, including education level, employment status, per capita expenditure, and housing facilities. These aspects are widely recognized as key factors that influence poverty conditions in developing countries.

The analysis steps can be carried out as follows:

- Data collection.
  Data were collected using secondary data obtained from the Central Statistics Agency (BPS), covering poverty indicators across districts/cities in Indonesia.
- Performing average imputation.
  Missing data in the dataset were handled using average imputation to ensure data completeness and avoid bias in the analysis.
- Descriptive statistical analysis.
  Descriptive analysis was conducted to provide an initial overview of the characteristics of the response variable and predictor variables.
- Analyzing the relationship between the response variables and predictor variables.
  This step was carried out using scatter plots to identify the pattern of relationships, whether linear or nonlinear.
- Divide the data into training data and testing data.
- Analyze the MARS model.
  Modelling was performed using Multivariate Adaptive Regression Spline (MARS) method to capture nonlinear relationship and interactions between variables. In the MARS modelling process, several parameter combinations were tested to obtain the optimal model. The number of basis functions (BF) controls the model's complexity, with a higher BF allowing the model to capture more flexible patterns in the data. The maximum interaction parameter (MI) determines the allowed level of interaction between predictor variables, while the minimum observations between knots (MO) controls the minimum number of observations required between two knots to avoid overfitting. Several combinations of BF, MI and MO are evaluated, and the best model is selected based on the smallest Ganeralized Cross_Validation (GCV) value.
- Determine the maximum possible number of basis functions (BF), which are 26, 39, and 52.
  The BF parameter controls the flexibility of the model in capturing complex patterns in the data.
- Determine the maximum number of interactions (MI), which are 1, 2, and 3.
- Determine the minimum number of observations between knots (MO), which are 0, 1, 2, and 3.
- Determine the best MARS model according to the smallest Generalized Cross Validation (GCV) value obtained from the combination of BF, MI, and MO.
- Test the significance of the MARS model.
  Significance testing was conducted to identify which predictor variables have a significant effect on the response variable.
- Perform Bagging modeling of the best MARS model with 50 replications.
  The Bootstrap Aggregating (Bagging) technique wa applied to improve model stability and prediction accuracy through repeated resampling.
- Determine the best Bagging MARS model.

Correlations between predictor variables can occur because several variables describe related social and economic conditions. However, the Multivariate Adaptive Regression Splines (MARS) method is relatively robust to multicollinearity because it automatically selects relevant variables through forward and backward procedures when constructing the basis fuction. Therefore, correlations between predictor variables do not significantly affect model stability.

III. RESULT AND DISCUSSION

A. *Descriptive Analysis*

This analysis was conducted to characterize poverty across districts/cities in Indonesia. This study used data on the percentage of poor people in 514 districts/cities in Indonesia in 2024, obtained from the Indonesian Central Statistics Agency (BPS).

During data processing, missing data were encountered. Missing data refers to unobserved values that would be meaningful for analysis if observed, or in other words, missing values conceal meaningful values [18]. Several data values are missing or unavailable in the Central Statistics Agency (BPS) publication. Therefore, several missing or unavailable data values were imputed using the specified method. One simple method commonly used is mean imputation. Imputation using the mean value method is done by filling in the unavailable data with the mean of all known values in a variable [19].

Based on Table II, the data not available in this study are in variable X_8 namely the percentage of poor people aged 15 years and above whowork in the formal sector, X_9 which is the percentage of poor people aged 15 years and above who work in the agricultural sector, X_{10} which is the percentage of poor people aged 15 years and above who work outside the agricultural sector, X_{12} which is the percentage of poor households that use proper water, and X_{13} which is the percentage of poor households that use their own or shared toilets.

Based on BPS data on poor people in Indonesia, the descriptive statistics are presented in Table II.

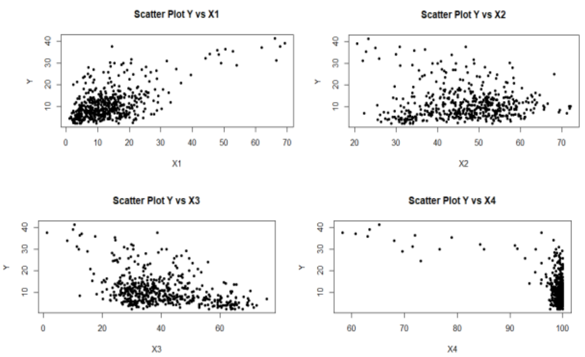
TABLE II
DESCRIPTIVE STATISTICS OF RESEARCH VARIABLES

Variable	Minimum	Maximum	Mean	Standard Deviation	Data Lost
Y	2.23	41.42	11.17	7.07	0
X_1	1.11	69.32	13.56	9.62	0
X_2	20.66	72.13	46.61	968	0
X_3	1.28	75.92	39.84	12.89	0
X_4	58.30	99.99	98.45	4.94	0
X_5	29.08	99.99	95.27	6.48	0
X_6	5.98	57.24	35.58	7.04	0
X_7	13.85	93.18	38.81	13.43	0
X_8	1.00	47.76	25.78	8.85	3
X_9	0.24	93.06	26.58	16.54	6
X_{10}	0.44	69.42	38.27	12.93	2
X_{11}	48.59	80.60	63.32	5.19	0
X_{12}	17.60	100.00	84.47	17.13	6
X_{13}	4.84	100.00	86.40	15.40	2

Based on Table II, it can be seen that there are two variables with a minimum percentage below 1% is variables X_9 and X_{10}. This means that in some districts and cities, only a small proportion of poor people aged 15 years and above work in the formal and agricultural sectors. The variable with the highest maximum value is the variable with a value of 100%. Two variables have a maximum percentage of 100%, namely variables X_{12} and X_{13}. This means that there are districts/cities where all poor households use proper water for variable X_{12}. For variable X_{13}, this means that there are districts/cities where all poor households use their own or shared toilets. The variable with the highest average value is variable X_4. This means that, among district- and city-level variables, the percentage of school participation among poor residents aged 7-12 years in Indonesia is relatively high.

Next, to determine the relationship pattern between the response variable and each predictor variable, a visual analysis was performed using a scatter plot diagram. Scatter plots are used to see the relationship pattern between the response variable and the predictor variable and to identify whether the relationship is linear or nonlinear.

The scatter plot between the response variable and the thirteen predictor variables used in this study can be seen in figure 1.



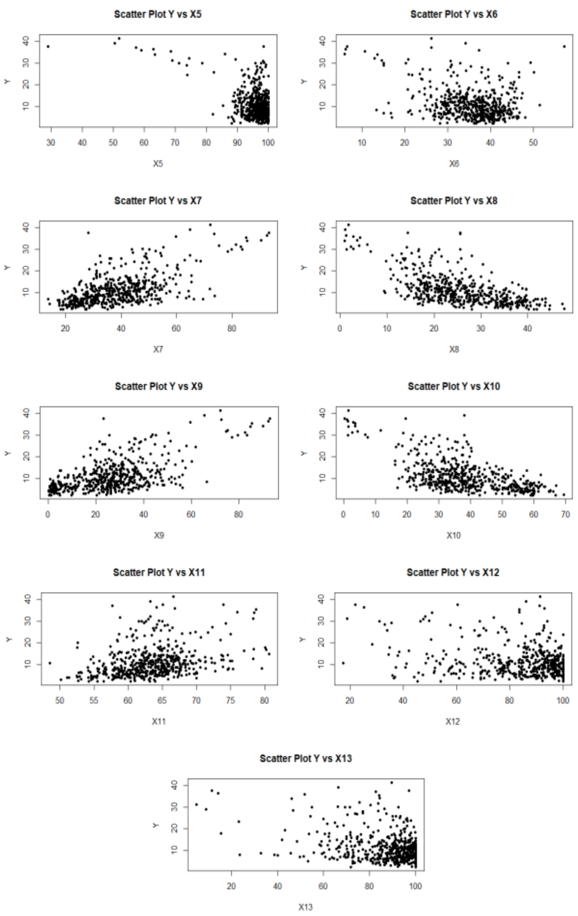


Figure 1. Scatter Plot Variable Y and Variable X_1 to X_{13}

Based on figure 1, it can be observed that some of the thirteen scatter plots do not show a clear pattern between the response variable and the predictor variables. In some cases, the data points tend to form pattern that change at certain intervals. This indicates that the relationship between the response variable and the predictor variables may not be entirely linear.

Therefore, due to the limited information regarding the functional for of the relationship between the response variables and the predictor variables, as well as the unclear patterns observed in the data, a nonparametric regression approach was chosen to model the data.

B. *Modeling of Poverty with MARS*

To analyze the MARS model, the first step is to split the data into training and test sets. The training data and testing data used in this study are 70% and 30%, respectively. Next, the model is constructed using the MARS method with the following equation [13]:

$$\hat{f}(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[S_{km}\left(x_{v(k,m)} - t_{km}\right) \right] \quad (1)$$

To determine the optimal MARS model, the Generalized Cross-Validation (GCV) criterion was used. The model with the lowest GCV is the best. The equation for Generalized Cross Validation (GCV) is as follows:

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

The minimum GCV value is determined from a combination of the maximum basis function (BF), maximum interaction (MI), and minimum observation (MO) using trial and error.

TABLE III
TRIAL AND ERROR ESTIMATION OF MARS MODELS

BF	**MI**	**MO**	**GCV**
26	1	0	19.16467
26	1	1	19.13147
26	1	2	19.12505
26	1	3	19.13936
26	2	0	20.36694
26	2	1	20.16542
26	2	2	20.15975
26	2	3	20.47708
26	2	0	20.03287
26	3	1	20.74002
26	3	2	20.73475
26	3	3	20.18621
39	1	0	19.21474
39	1	1	19.15535
39	1	2	19.18402
39	1	3	19.17413
39	2	0	19.13691
39	2	1	19.27595
39	2	2	19.27076
39	2	3	19.16356
39	3	0	19.21075
39	3	1	18.88742
39	3	2	18.88613
39	**3**	**3**	**18.17830**
52	1	0	19.44351
52	1	1	19.06154
52	1	2	19.29990
52	1	3	19.34973
52	2	0	19.13691
52	2	1	19.27595
52	2	2	19.27076
52	2	3	19.64202
52	3	0	19.47700
52	3	1	18.81919
52	3	2	19.04607
52	3	3	18.21164

Based on trial and error, the best MARS model with the smallest GCV value was obtained, namely a combination of BF = 39, MI = 3, MO = 3, and GCV of 18.1730 with the following model equation:

$$\begin{aligned}
\hat{Y} = {} & 21.5780875 + 0.1933702\mathrm{BF}_2 + 0.1359168\mathrm{BF}_3 \\
& +1.6256190\mathrm{BF}_4 - 0.0362631\mathrm{BF}_5 - 0.0369953\mathrm{BF}_6 \\
& +1.6717820\mathrm{BF}_7 - 0.0218438\mathrm{BF}_8 - 0.2659416\mathrm{BF}_9 \\
& +0.0332643\mathrm{BF}_{10} - 0.0780639\mathrm{BF}_{11} - 1.0427623\mathrm{BF}_{12} \\
& -0.2121136\mathrm{BF}_{13} - 0.0401120\mathrm{BF}_{14} + 0.310755\mathrm{BF}_{15} \\
& -0.0279433\mathrm{BF}_{16} + 0.0212066\mathrm{BF}_{17} - 0.5431099\mathrm{BF}_{18} \\
& -0.0278867\mathrm{BF}_{19} - 0.0027990\mathrm{BF}_{20}
\end{aligned}$$

with,

$$\begin{aligned}
BF_2 &= \max(0,(X_4 - 98.52))\max(0,(X_8 - 16.59)) \\
BF_3 &= \max(0,(98.52 - X_4))\max(0,(X_8 - 16.59)) \\
BF_4 &= \max(0,(X_6 - 44.37)) \\
BF_5 &= \max(0,(X_1 - 23.05))\max(0,(44.37 - X_6)) \\
BF_6 &= \max(0,\ (23.05 - X_1))\max(0,(44.37 - X_6)) \\
BF_7 &= \max(0,(X_1 - 19.37)) \\
BF_8 &= \max(0,(X_6 - 31.37))\,BF_2 \\
BF_9 &= \max(0,(48.28 - X_3)) \\
BF_{10} &= \max(0,(X_2 - 46.13))\max(0,(44.37 - X_6)) \\
BF_{11} &= \max(0,(X_8 - 16.66))\,BF_7 \\
BF_{12} &= \max(0,(X_9 - 65.61)) \\
BF_{13} &= \max(0,(65.61 - X_9)) \\
BF_{14} &= \max(0,(X_8 - 16.59))\max(0,(X_9 - 21.39)) \\
BF_{15} &= \max(0,(X_8 - 16.59))\max(0,(82.61 - X_{13})) \\
BF_{16} &= \max(0,(X_4 - 98.76))\,BF_{15} \\
BF_{17} &= \max(0,(X_5 - 88.68))\max(0,(55.11 - X_{10})) \\
BF_{18} &= \max(0,(X_4 - 98.93))\,BF_7 \\
BF_{19} &= \max(0,(52.49 - X_2))\,BF_9 \\
BF_{20} &= \max(0,(\ 30.13 - X_7))\,BF_{17}
\end{aligned}$$

In the MARS model above, not all variables are included. This means that of the 13 variables, there are two variables that have no effect on the poverty rate in Indonesia, namely the variables X_{11} and X_{12}. To determine the order of variables from highest to lowest effect on the model, the variables' importance can be assessed. The relative importance of the variables is shown in Table IV.

TABEL IV
LEVEL OF IMPORTANCE OF MARS MODEL PREDICTOR VARIABLES

Variable	**Level of Importance (%)**
X_1	100
X_8	65.7
X_6	57.2
X_9	57.2
X_4	41.8
X_3	37.1

Variable	Level of Importance (%)
X_2	37.1
X_5	29.1
X_{10}	29.1
X_{13}	24.5
X_7	0.6

Based on Table IV, it can be seen that the variable of poor people aged 15 years and above who did not complete elementary school (X_1) has a significance level of 100%. This means that the variable X_1 has the greatest influence on the poverty rate in Indonesia.

C. Significance Test of the MARS Model

After identifying the optimal MARS model, parameter testing will be conducted. Model parameter testing aims to evaluate the model's application, assess the significance of its parameters, and assess the model's suitability. Model significance testing is carried out by testing the regression coefficients simultaneously and partially [20].

1) *Simultaneous test*

Simultaneous testing is performed by evaluating the parameters in the MARS model. Let set $A = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$. The hypothesis used is:

$H_0 : \alpha_i = 0 \; ; \; \forall \, i \; \in A.$

H_1 : there exists $\alpha_i \neq 0$; for $i \in A$.

Using a significance level of 0.05, reject H_0 if the value of $F_{hitung} > F_\alpha(M, n - M - 1)$ or if the $p - values < 0.05$.

```
R-squared          : 0.749757
Adj R-squared      : 0.735731
Model DF (MDF)     : 19
Residual DF (NDF)  : 339
F-statistic        : 53.456955
p-value            : 8.69959181857784e-90
F-tabel (0.05)     : 1.6173

Kesimpulan         : Model SIGNIFIKAN secara simultan
```

Figure 2. Simultaneous MARS Model Test

Based on the results of the R Studio software in Figure I, $p - value$ of 8.6996×10^{-90} was obtained. This can be interpreted as rejecting H_0 because $p - value < 0.05$. Therefore, the MARS model is significant.

2) *Partial test*

Partial testing is conducted by evaluating each model parameter individually. Let say we have set $A = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$. The hypothesis used is:

$H_0 : \alpha_i = 0 \; ; \; \forall \, i \in A.$

H_1 : there exists $\alpha_i \neq 0$; for $i \in A$.

Using a significance level of 0.05, reject H_0 if the value of $|t_{hitung}| > t_{\frac{\alpha}{2}(n-M-1)}$ or if the $p - values < 0.05$.

UJI t (Parsial)

Parameter	Koefisien	Std.Error	t-value	p-value
(Intercept)	21.578088	1.917011	11.256108	0.000000
BF2	0.193370	0.046894	4.123526	0.000047
BF3	0.135917	0.021020	6.466159	0.000000
BF4	1.625619	0.247417	6.570353	0.000000
BF5	-0.036263	0.005573	-6.506832	0.000000
BF6	-0.036995	0.004543	-8.143876	0.000000
BF7	1.671782	0.185667	9.004187	0.000000
BF8	-0.021844	0.004996	-4.372433	0.000016
BF9	-0.265942	0.046205	-5.755665	0.000000
BF10	0.033264	0.005900	5.637576	0.000000
BF11	-0.078064	0.016876	-4.625709	0.000005
BF12	1.042762	0.202485	5.149819	0.000000
BF13	-0.212114	0.032117	-6.604406	0.000000
BF14	-0.040112	0.005110	-7.849701	0.000000
BF15	0.031075	0.006095	5.098812	0.000001
BF16	-0.027943	0.007836	-3.566150	0.000414
BF17	0.021207	0.003528	6.010587	0.000000
BF18	-0.543110	0.169668	-3.201009	0.001499
BF19	-0.027887	0.007150	-3.900438	0.000116
BF20	-0.002799	0.001193	-2.345949	0.019554

Figure 3. Partial Test of the MARS Model

Based on the MARS processing results using R Studio software in Figure 2, $p - value < 0.05$ was obtained, which means rejecting H_0. Thus, it can be concluded that all basis functions have a significant effect on the model.

D. Bagging MARS Approach

The Bagging approach on the best MARS model was replicated 50 times using the bootstrap method. On the new dataset $\{L^{(50)}\}$, the replication results were modelled using MARS with the same BF, MI, and MO criteria as those used to determine the best model. After performing Bagging MARS, the best model was obtained in the 42nd iteration with the smallest GCV value of 9.207091. A lower Generalized Cross-Validation (GCV) value indicates that the model has better predictive performance and is able to estimate the response variable more accurately. Therefore,the minimum GCV value obtained in this study indicates that the selected Bagging MARS model provides a good fit for explaining poverty levels across various regencies and cities in Indonesia. The resulting model equation is as follows:

$$\begin{aligned}
\hat{Y} &= 23{,}223329 - 2{,}2844860BF_2 + 1{,}5968114BF_3 \\
&+0{,}1236092BF_4 + 0{,}3513309BF_5 + 1{,}2238172BF_6 \\
&-0{,}0277388BF_7 + 0{,}0180894BF_8 + 2{,}4649357BF_9
\end{aligned}$$

$$\begin{aligned}
\hat{Y} &= 23.223329 - 2.2844860BF_2 + 1.5968114BF_3 \\
&+1.1236092BF_4 + 0.3513309BF_5 + 1.2238172BF_6 \\
&-0.0277388BF_7 + 0.0180894BF_8 + 2.4649357BF_9 \\
&-2.0738249BF_{10} - 0.0440250BF_{11} - 0.0837569BF_{12} \\
&-0.0053200BF_{13} - 0.0242118BF_{14} + 0.1947357BF_{15} \\
&-0.0958912BF_{16} - 0.0189449BF_{17} + 0.0306229BF_{18} \\
&+0.0242890BF_{19} + 0.0348320BF_{20} - 0.0032204BF_{21} \\
&-0.0038656BF_{22} + 0.0032281BF_{23}
\end{aligned}$$

with,

$BF_2 = \max(0, (X_4 - 91.39))$

$BF_3 = \max(0, (X_6 - 43.21))$

$BF_4 = \max(0, (43.21 - X_6))$

$BF_5 = \max(0, (92.15 - X_5))$

$BF_6 = \max(0, (X_1 - 18.43)) \max(0, (X_5 - 92.15))$

$BF_7 = X_{10} BF_6$

$BF_8 = \max(0,(X_5 - 92.15))\max(0,(X_{13} - 69.2))$
$BF_9 = \max(0,(X_5 - 92.15))\max(0,(69.2 - X_{13}))$
$BF_{10} = \max(0,(X_{12} - 98.89))$
$BF_{11} = \max(0,(98.89 - X_{12}))$
$BF_{12} = \max(0,(X_5 - 92.15))\max(0(31.21 - X_6))$
$BF_{13} = X_{12}BF_9$
$BF_{14} = X_2BF_9$
$BF_{15} = \max(0,(X_4 - 99.08))\max(0,(X_8 - 14.33))$
$BF_{16} = \max(0,(X_8 - 14.33))\max(0,(68.8 - X_{13}))$
$BF_{17} = X_7BF_9$
$BF_{18} = \max(0,(X_{13} - 81.23))BF_2$
$BF_{19} = \max(0,(81.23 - X_{13}))BF_2$
$BF_{20} = \max(0,(X_9 - 31.65))BF_2$
$BF_{21} = \max(0,(X_8 - 19.65))BF_{18}$
$BF_{22} = \max(0,(19.65 - X_8))BF_{18}$
$BF_{23} = \max(0,(X_1 - 4.89))\max(0,(X_{11} - 63.3))\ BF_2$

To assess the stability of the model, replication results between resamples were observed. Although there was slight variation between resamples, the order of the most influential variables was relatively consistent, indicating that the model was fairly stable and the prediction results were not overly sensitive to sample changes. This confirms that the Bagging MARS model can be relied upon to identify factors that influence poverty, so that policy interventions can be focused on the most influential variables.

In the Bagging MARS model, it can be seen that all variables are included in the model equation except for X_3. This indicates that, among the 13 variables, one has no effect on the poverty rate in Indonesia. To see the order of variables that have the highest to lowest effect on the variables, see Table V.

TABLE V
LEVEL OF IMPORTANCE OF BAGGING MARS MODEL PREDICTOR VARIABLES

Variable	**Level of Importance** (%)
X_4	100
X_8	56.4
X_{13}	51.1
X_5	45.6
X_6	45.6
X_1	44.5
X_{10}	37.7
X_{11}	32.6
X_{12}	29.8
X_2	24.1
X_7	16.8
X_9	12.8

Based on Table V, it can be seen that the variable percentage of school participation of poor people aged 7 – 12 years (X_4) has a significant effect on poverty levels in Indonesia. The results of the Bagging MARS model indicate that several predictor variables contribute differently to poverty rates across various districts and cities in Indonesia.

Among the analyzed variables, the percentage of school participation of poor people aged 7 – 12 years (X_4) exhibits the stongest influence on poverty rates.these findings suggest that access to basic education plays a crucial role in improving the well-being of poor haouseholds. Areas with higher participation school from poor families tend to have lower poverty rates because education can improve fitire employment opportunities and income potential.

Conventional approaches such as linear regression are commonly used to analyze determinants of poverty. However, linear regression assumes a linear relationship between variables and may not adequately capture nonlinear patterns in poverty data. Machine learning approaches such as Random Forest are also widely applied for predictive modelling. Nevertheless, the Multivariate Adaptive Regression Spline (MARS) method offers an advantage because it can model nonlinear relationships while maintaining interpretability through basis functions. Therefore, the Bagging MARS approach ised in this study provides a balance between predictive performance and model interpretability.

The findings of this study provide valuable information for policymakers in identifying factors that influence poverty levels across various districts and cities in Indonesia. Identifying the most influential variables can help policymakers design more targeted poverty alleviation programs.

This study has several limitations, namely:

- The analysis used aggregate data at the district/city level, which may not fully reflect poverty conditions at the household or individual level.
- Differences in socioeconomic characteristics between regions in Indonesia may cause regional bias in the analysis. Therefore, future studies are expected to include more detailed microdata or additional regional indicators to gain a more comprehensive understanding of the poverty factors.

IV. CONCLUSION

Based on the results of the research and discussion, the following conclusions can be drawn:

- The best Bagging MARS model for poverty data in Indonesia in 2024 obtained a minimum Generalized Cross Validation (GCV) value of 9.207091 in the 42nd resample with the following model equation:

$$\begin{aligned}\hat{Y} &= 23.223329 - 2.2844860BF_2 + 1.5968114BF_3\\ &+0{,}1236092BF_4 + 0.3513309BF\_5 + 1.2238172BF_6\\ &-0.0277388BF_7 + 0.0180894BF_8 + 2.4649357BF_9\\ &-2.0738249BF_{10} - 0.0440250BF_{11} - 0.0837569BF_{12}\\ &-0.0053200BF_{13} - 0.0242128BF_{14} + 0.2947357BF_{15}\\ &-0.0958912BF_{16} - 0.0189449BF_{17} + 0.0306229BF_{18}\\ &+0.0242890BF_{19} + 0.0348320BF_{20} - 0.0032204BF_{21}\\ &-0.0038656BF_{22} + 0.0032281BF_{23}\end{aligned}$$

with,

$BF_2 = \max(0,(X_4 - 91.39))$
$BF_3 = \max(0,(X_6 - 43.21))$

$BF_4 = \max(0, (43.21 - X_6))$
$BF_5 = \max(0, (92.15 - X_5))$
$BF_6 = \max(0, (X_1 - 18.43)) \max(0, (X_5 - 92.15))$
$BF_7 = X_{10} BF_6$
$BF_8 = \max(0, (X_5 - 92.15)) \max(0, (X_{13} - 69.2))$
$BF_9 = \max(0, (X_5 - 92.15)) \max(0, (69.2 - X_{13}))$
$BF_{10} = \max(0, (X_{12} - 98.89))$
$BF_{11} = \max(0, (98.89 - X_{12}))$
$BF_{12} = \max(0, (X_5 - 92.15)) \max(0(31.21 - X_6))$
$BF_{13} = X_{12} BF_9$
$BF_{14} = X_2 BF_9$
$BF_{15} = \max(0, (X_4 - 99.08)) \max(0, (X_8 - 14.33))$
$BF_{16} = \max(0, (X_8 - 14.33)) \max(0, (68.8 - X_{13}))$
$BF_{17} = X_7 BF_9$
$BF_{18} = \max(0, (X_{13} - 81.23)) BF_2$
$BF_{19} = \max(0, (81.23 - X_{13})) BF_2$
$BF_{20} = \max(0, (X_9 - 31.65)) BF_2$
$BF_{21} = \max(0, (X_8 - 19.65)) BF_{18}$
$BF_{22} = \max(0, (19.65 - X_8)) BF_{18}$
$BF_{23} = \max(0, (X_1 - 4.89)) \max(0, (X_{11} - 63.3))\ BF_2$

- Of the 13 predictor variables studied, based on the level of importance of the predictor variables, there is one variable that has a 100% level of importance, namely the variable of the percentage of school participation of poor people aged 7 – 12 years (X_4). This indicates that the variable percentage of school enrolment among poor children aged 7 – 12 (X_4) has the greatest influence on the poverty rate in Indonesia.

BIBLIOGRAPHY

[1] C. Suryawati, “Memahami Kemiskinan secara Multidimensioal,” *J. Manaj. Pelayanan Kesehat.*, vol. 08, no. 03, pp. 585–597, 2010.
[2] BPS, “Profil Kemiskinan di Indonesia Per Maret 2024,” Jakarta, 2024.
[3] World Bank, “Macro Poverty Outlook - East Asia and the Pacific,” Washington, DC, 2025.
[4] S. Arifin and F. Firmansyah, “Pengaruh Tingkat Pendidikan Dan Kesempatan Kerja Terhadap Pengangguran Di Provinsi Banten,” *J. Ekon.*, vol. 7, no. 2, 2017, doi: 10.35448/jequ.v7i2.4978.
[5] Norwan, *Kemiskinan di Negara Berkembang*. Pustaka Taman Ilmu, 2023. [Online]. Available: https://ipusnas2.perpusnas.go.id/read-book
[6] S. L. M. Muthoharoh and A. Wahyudi, “Pengelolaan Ketenagakerjaan Dan Pengangguran Dalam Islam: Sebab, Dampak dan Solusi,” *J. Masharif al-Syariah J. Ekon. dan Perbank. Syariah*, vol. 8, no. 3, pp. 276–301, 2023.
[7] D. Lazuardi, I. Gustina, P. Wahyuni, and M. Rinaldi, “Peningkatan Akses Layanan Dasar Untuk Mengurangi Kemiskinan : Pendekatan Berbasis Pemberdayaan Masyarakat Di Kota Medan,” *Lebah*, vol. 18, no. 2, pp. 69–76, 2025.
[8] M. F. Qudratullah, *Analisis Regresi Terapan: Teori, Contoh Kasus, dan Aplikasi dengan SPSS*. Penerbit ANDI, 2013.
[9] P. Bilski, “Analysis of the ensemble of regression algorithms for the analog circuit parametric identification,” *Meas. J. Int. Meas. Confed.*, vol. 160, p. 107829, 2020, doi: 10.1016/j.measurement.2020.107829.
[10] E. U. Dewi, “Model Regresi Semiparametrik Multivariabel Dengan Estimator Spline Parsial,” *J. Keperawatan Stikes William Booth*, vol. 3, no. 1, pp. 1–10, 2019.
[11] R. L. Eubank, *Nonparametric Regression and Spline Smoothing, Second Edition*. 1999.
[12] N. Chamidah and B. Lestari, “Analisis Regresi Nonparametrik dengan Perangkat Lunak R,” Airlangga University Press, 2022, pp. 16–19.
[13] J. H. Friedman, “Multivariate Adaptive Regression Splines,” *Ann. Stat.*, vol. 19, no. 1, pp. 1–67, 1991.
[14] I. N. Budiantara, *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. ITS Press, 2009.
[15] L. Breiman, “Bagging Predictors,” 1996.
[16] M. Kulekci, E. Eyduran, A. Yusuf Altm, and M. Masood Tariq, “Usefulness of MARS and Bagging MARS Algorithms in Prediction of Honey Production in Beekeeping Enterprises from Elazig Province of Turkey,” *Pak. J. Zool.*, 2022, doi: https://dx.doi.org/10.17582/journal.pjz/20200309160354.
[17] A. M. Yusuf, *Metode Penelitian: Kuantitatif, Kualitatif & Penelitian Gabungan*. Jaka: Kencana, 2014.
[18] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, Third Edition. Wiley Series in Probability and Statistics, 2020.
[19] R. S. V. Teegavarapu, *Imputation Methods for Missing Hydrometeorological Data Estimation*, vol. 108. 2024. [Online]. Available: https://doi.org/10.1007/978-3-031-60946-6
[20] W. Wicaksono, Y. Wilandari, and Suparti, “Pemodelan multivariat adaptive regression splines (MARS) pada faktor resiko angka kesakitan diare,” *J. Gaussian*, vol. 3, no. 2, pp. 253–262, 2014.