Multiple Nonparametric Regression Approach for Regional Mapping by Considering the Significance of the Independent Variables

Rahmat Hidayat, Muhammad Ilyas, Yuliani

***Abstract*— This article presents a novel perspective on classification, utilizing the Spline nonparametric model as its foundation. In the realm of classification, this Spline model serves as a viable alternative to geographically weighted regression, logistic regression, and other classification models, especially when data patterns do not adhere to a specific distribution. Model parameter estimation is performed through the Ordinary Least Squares (OLS) approach, and the data poverty severity index is employed to assess the model's performance. The results indicate that the Spline model with three knots outperforms other models, thus establishing itself as the cornerstone for classification. Additionally, we provide an overview of the distribution map based on the derived mathematical model.**

***Index Terms*— classification, OLS, map, nonparametric, Spline**

I. Introduction

The Millennium Development Goals (MDGs), the outcome of the Millennium Summit attended by 189 United Nations member states in 2000, encompass a range of objectives. These goals encompass addressing poverty and hunger, ensuring universal access to basic education, promoting gender equality and women's empowerment, enhancing maternal and child health, combating HIV/AIDS, preserving the environment, and fostering global collaboration for development. It is crucial to emphasize that the central aim of the MDGs is poverty eradication [1]. Following the conclusion of the MDGs in 2015, the international community adopted the Sustainable Development Goals (SDGs), which are in effect from 2015 to 2030. The issue of poverty remains a pressing concern for many developing nations, including Indonesia. An individual is deemed impoverished if they cannot fulfill fundamental needs such as clothing, food, shelter, healthcare, and education.

Manuscript received August 10, 2023; revised January 16, 2024.

This work was supported in part by Dir. Riset, Teknologi, dan Pengabdian Kepada Masyarakat, Dirjen Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia for funding this research (grant number: 161/E5/PG.02.00.PT/2022).

R. Hidayat is a lecturer in Mathematics Department, Faculty of Science, Universitas Cokroaminoto Palopo, Indonesia (corresponding author to provide e-mail: dayatmath@gmail.com).

M. Ilyas is a professor in Master of Mathematics Education Program, Universitas Cokroaminoto Palopo, Indonesia. (e-mail: muhammadilyas@uncp.ac.id).

Yuliani is a lecturer in Mathematics Department, Faculty of Science, Universitas Cokroaminoto Palopo, Indonesia. (e-mail: yulimath2507@gmail.com).

Addressing poverty in Indonesia has emerged as a top priority within the government's strategic agenda, as outlined in the National Long-Term Development Plan. This overarching plan is further segmented into four periods, known as the National Medium-Term Development Plans, with the third period spanning from 2015 to 2019. The poverty rate reached its peak in 2015 and steadily declined through 2019 [2]. Nevertheless, this decline in the poverty rate fell short of the government's 2019 target of 8%.

Data released by the Central Bureau of Statistics reveals the significant economic impact of the Covid-19 pandemic, which has subsequently affected people's income. The proportion of individuals living in poverty rose from 9.22% in September 2019 to 9.78% in March 2020 [3]. In March 2020, both urban and rural areas saw an increase in the poverty depth index, with the index climbing from 1.50 points in September to 1.61 points. Furthermore, the poverty severity index also saw an uptick, rising from 0.36 points to 0.38 points. Analyzing the factors contributing to poverty is a critical step in alleviating it. One valuable metric for assessing poverty within a region is the Poverty Severity Index (PSI).

Regression analysis is a valuable method for scrutinizing the factors that impact the Poverty Severity Index (PSI). Notably, it is important to acknowledge that the increase in one of the factors affecting PSI does not necessarily guarantee a corresponding increase in PSI itself. Hence, opting for the Spline Nonparametric Regression method is a prudent choice.

Spline Nonparametric Regression offers several advantages, primarily owing to its segmented polynomial nature. This segmentation property enhances its flexibility compared to standard polynomials, enabling it to adapt more effectively to the local nuances of a function or dataset [4]. Furthermore, Spline possesses the added benefit of autonomously estimating data points, allowing it to accurately capture the shifting data patterns. In the context of our study, the data plots depicting the relationships between response variables and each predictor variable lack a specific pattern. This observation reinforces the use of PSI research as an economic growth indicator in Indonesia, particularly when employing the Nonparametric Regression approach.

Accurate area mapping is of paramount importance to support economic growth in each province of Indonesia.

Social mapping serves the crucial purpose of establishing a foundation or reference for identifying potential program targets, ensuring that development planning aligns with the area's inherent potential. Mathematical modeling has been extensively employed in mapping areas, with various approaches documented in the literature. For instance, we can utilize MARS [5], employ geographically weighted regression models [6][7][8][9], and apply logistic regression [10][11]. In addition to that, we can also utilize Artificial Neural Networks [12][13][14], explore Neuro-Fuzzy techniques [15], implement decision trees [16], and incorporate frequency ratios [17]. Nevertheless, given that not all factors used as the basis for mapping adhere to a specific distribution, such as linearity, the development of an area mapping system utilizing a nonparametric regression model becomes essential.

The absence of distinct patterns in the data plots between the response variables and each predictor necessitates an analysis employing the Spline Nonparametric Regression approach to discern the factors affecting PSI in Indonesia, as outlined in the preceding background description. Furthermore, the study aims to develop a classification or regional mapping utilizing the obtained Spline model. The results of this research are expected to serve as a valuable tool for modeling PSI in Indonesia and provide essential insights for government policymakers to consider when formulating strategies to enhance economic growth in various Indonesian regions.

II. Method

A. Spline

Spline is a method used in non-parametric regression known for its flexibility in tracing the shape of data patterns [4]. The general form of a nonparametric regression model is represented as:

$$y_i = f(x_i) + \varepsilon_i \tag{1}$$

where y_i is the i-th response variable and $f(x_i)$ is the regression function for which the shape of the regression curve is unknown.

In nonparametric regression, the Spline method stands out as a technique capable of independently discovering data patterns without adhering to predefined patterns. As elucidated in [18,19], a Spline is a polynomial where distinct polynomial segments are connected at knots denoted as k_1, k_2, …, k_r. These segments maintain continuity, resulting in increased flexibility compared to conventional polynomials. Knot points represent the junctures on the Spline where the curve's behavior changes. The Spline function $f(x_i)$ with degree of p can be expressed as [20]:

$$f(x_i) = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{m=1}^{r} \beta_{p+m} (x_i - k_m)_+^p \tag{2}$$

if equation (2) is substituted into equation (1) then the nonparametric spline regression model is obtained as follows:

$$y_i = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{m=1}^{r} \beta_{p+m} (x_i - k_m)_+^p + \varepsilon_i \tag{3}$$

Function $(x_i - k_m)_+^p$ is a chunk function given by:

$$(x_i - k_m)_+^p = \begin{cases} (x_i - k_m)^p, & x_i \ge k_m \\ 0, & x_i < k_m \end{cases} \tag{4}$$

where from these functions β_j is the parameter of the polynomial model where j = 1,2, …, m. x_i is an independent variable with i = 1,2, …, n. $\beta m{+}k$ is a truncated component parameter with k = 1,2, …, r. r is the number of knots and k_m is the points of the knots.

B. Optimal Knot Point

The identification of optimal knot points serves as a critical indicator of the Spline model's accuracy. These knot points signify the junctures where the data pattern undergoes a significant change, whether it is an upward or downward transition [21][22][23]. This study employs the Generalized Cross Validation (GCV) method for the selection of these optimal knots. The GCV method can be formulated as follows [4]:

$$GCV(k) = \frac{MSE(k)}{\left[n^{-1} tr(I - A)\right]^2} \tag{5}$$

where I is the identity matrix, n is the number of observations, $k = (k_1, k_2, ..., k_r)$ is the knot points.

$$MSE(k) = n^{-1} \sum_{i=1}^{n} \left(y_i - \hat{f}(x_i)\right)^2 \tag{6}$$

III. Result and Discussion

A. Parameter Estimation

Parameter estimation in multiple non-parametric regression was performed using the Ordinary Least Square (OLS) method. The OLS method aims to minimize the total squared error for parameter estimation. The matrix representation of the Spline non-parametric regression model with K nodes and univariable predictors can be expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{7}$$

where:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, x = \begin{pmatrix} 1 & x_1 & (x_1 - k_1)_+^1 & \cdots & (x_1 - k_k)_+^1 \\ 1 & x_2 & (x_2 - k_1)_+^1 & \cdots & (x_2 - k_k)_+^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - k_1)_+^1 & \cdots & (x_n - k_k)_+^1 \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

From equation (7), the residual equation can be expressed in writing as follows.

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{8}$$

The residual sum of squares matrix can be expressed as follows:

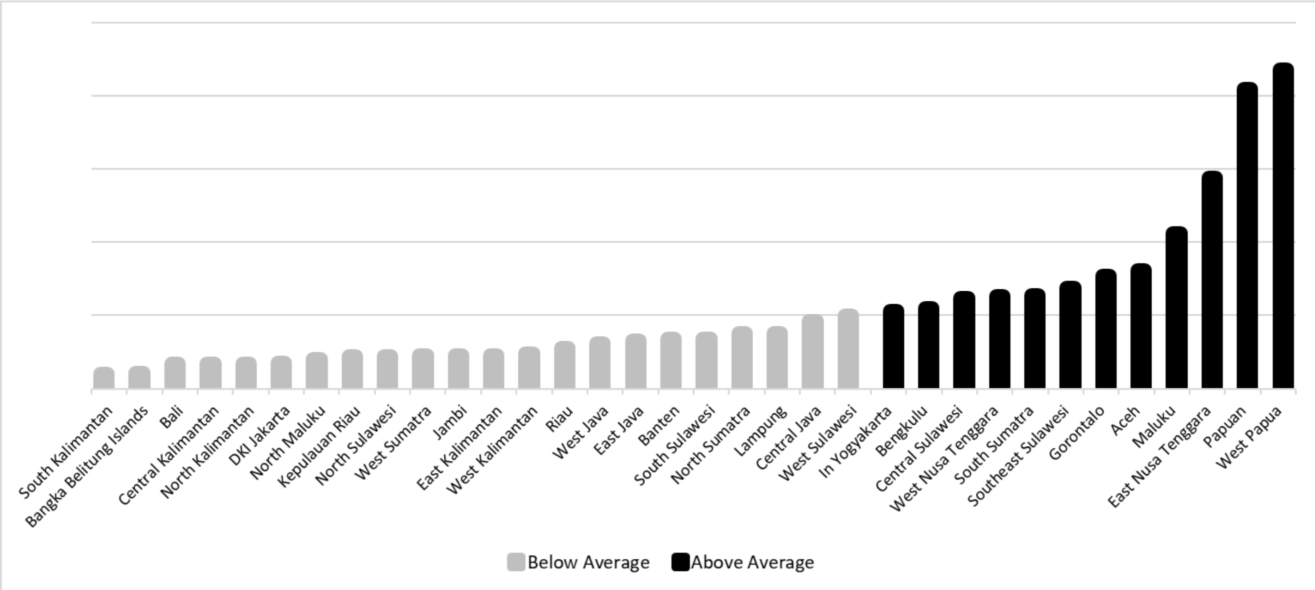

Fig. 1. Poverty Severity Index in Indonesia

$$\sum_{i=1}^{n}\varepsilon^2=\varepsilon'\varepsilon$$
$$=(\mathbf{y}-\mathbf{X\beta})'(\mathbf{y}-\mathbf{X\beta}) \tag{9}$$
$$=\mathbf{y'y}-\mathbf{y'X\beta}-\mathbf{X'\beta'y}+\mathbf{X\beta'X\beta}$$
$$=\mathbf{y'y}-2\mathbf{X'\beta'}+\mathbf{X'\beta'X\beta}$$

In order to minimize the total squared sum, the first derivative with respect to β must equal zero.

So the following equation is obtained:

$$-2\mathbf{X'y}+2\mathbf{X'X}\hat{\boldsymbol{\beta}}=\mathbf{0}$$
$$\mathbf{X'X}\hat{\boldsymbol{\beta}}=\mathbf{X'y} \tag{10}$$
$$\hat{\boldsymbol{\beta}}=(\mathbf{X'X})^{-1}\mathbf{X'y}$$

The estimate of $\hat{\mathbf{y}}$ can be written as follows

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$$
$$=\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'y} \tag{11}$$
$$=\mathbf{A}(\mathbf{k})\mathbf{y}$$

with $\mathbf{A}(\mathbf{k})$ being the matrix used for calculations on the Generalized Cross Validation (GCV) formula in selecting optimal knot points.

B. Characteristics of the Poverty Severity Index and Factors Allegedly Influential

Several factors are believed to influence poverty in Indonesia, including Gross Regional Domestic Income (x_1), the gross labor force participation rate (x_2), the Human Development Index (x_3), and the open unemployment rate (x_4). This section will present the characteristics of the Poverty Severity Index in Indonesia and the factors considered to have an impact, providing data such as mean, variance, minimum value, and maximum value. The following table illustrates the characteristics of the Poverty Severity Index and the associated influencing factors.

TABLE 1
CHARACTERISTICS OF POVERTY SEVERITY AND FACTORS ALLEGEDLY INFLUENCING

No.	Variables	Mean	Variance	Minimum	Maximum
1	y	0,52	0,24	0,10	2,18
2	x_1	45278,05	1122452503	13298,85	182908,69
3	x_2	68,64	12,86	63,08	77,75
4	x_3	71,99	14,79	61,39	81,65
5	x_4	68,64	12,86	63,08	77,75

Referring to the Table 1, the variable in question is the Poverty Severity Index for Indonesia in the year 2022. It exhibits an average value of 4.23, a variation of 0.52, and ranges between 0.10 and 2.18. A corresponding graph (Fig. 1) indicates that West Papua province holds the highest Poverty Severity Index value, while South Kalimantan province registers the lowest value.

Figure 1 clearly illustrates that West Papua Province has the highest percentage of the Poverty Severity Index, thus leading to the conclusion that it is the province in Indonesia with the highest poverty severity index, whereas South Kalimantan Province has the lowest poverty severity index. Among the regions in Indonesia, 22 fall below the provincial average for the poverty severity index, while 12 other provinces surpass this average. It's worth noting that the Covid-19 pandemic has intensified poverty levels, with the percentage rising from 9.22% in September 2019 to 9.78% in March 2020, resulting in an increased number of people classified as poor.

The x_1 variable represents Gross Regional Domestic Product (GRDP). The GRDP has an average value of 45,278.05, a variance of 11,224,503, a minimum value of 1,329.85, and a maximum value of 182,908.69. The substantial variance value highlights the uneven distribution of GRDP across the Indonesian provinces, with DKI Jakarta Province boasting the highest GRDP, while East Nusa Tenggara Province records the lowest GRDP.

The x_2 variable represents the Labor Force Participation Rate (LFPR). It has an average value of 68.64, a variance of

12.86, a minimum value of 63.0, and a maximum value of 77.75. Papua province exhibits the highest LFPR among the provinces, while DKI Jakarta has the lowest LFPR.

The x_3 variable represents the Human Development Index (HDI). The average HDI value is 2.94, with a variance of 14.79. The range extends from a minimum value of 61.39 to a maximum value of 81.65. DKI Jakarta boasts the highest HDI among the provinces, while Papua records the lowest HDI.

The x_4 variable represents the Open Unemployment Rate (UR). The range for this variable span from a minimum value of 63.0 to a maximum of 77.75. The average UR is 68.64, with a variance of 12.86. Papua Province records the highest UR among all provinces, while DKI Jakarta exhibits the lowest UR across Indonesia.

C. Scatter plot of Poverty Severity Index and Factors Allegedly Influencing

Scatter plots serve as effective tools for visualizing and comprehending the relationships between variables. When scatter plots reveal distinct patterns, such as linearity or quadratic shapes, a parametric approach is appropriate. However, when the scatter plot fails to exhibit a discernible pattern, the regression model is best developed using a nonparametric approach.

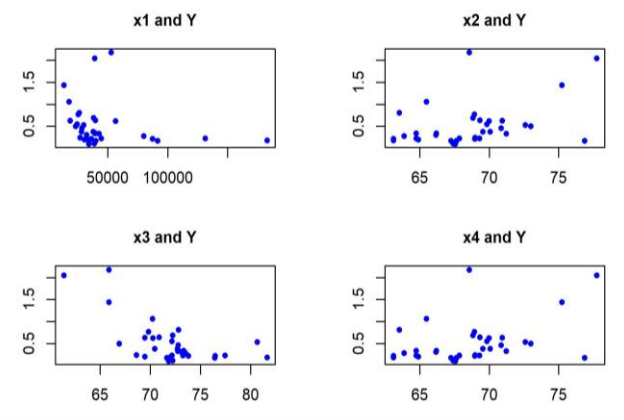

Fig. 2. Relationship between variables

In Figure 2 above, the relationship between the Poverty Severity Index and the variables believed to influence it is depicted. The four scatter plots reveal that these relationships do not conform to specific patterns. The absence of distinctive patterns in the relationships and the variations in the behavior of regression curves across multiple research variables underscore the rationale for adopting a nonparametric regression approach to model the data patterns. Additionally, the results of the Ramsey Reset linearity test, with a p-value tolerance exceeding 0.10 as presented in Table 2, indicate that the relationship between predictors and responses is nonlinear. This finding provides a compelling rationale for employing nonparametric regression analysis in this research.

TABLE 2
RAMSEY RESET NON-LINEARITY TEST

Relation	p-value	Decision
Relationship of x_1, x_2, x_3, x_4 to y	0,0249	not linear

D. Poverty Severity Index Modeling Using Spline Truncated Nonparametric Regression Method

The selection of knot points, specifically knot points 1, 2, and 3, was carried out using the Generalized Cross Validation method, commonly known as the GCV method. The choice of the best knot point is determined by the minimum GCV value. The following results display the minimum GCV values for 1, 2, and 3 knot points.

Spline with One Knot Point

In nonparametric spline truncated regression, parameter estimation with a single knot point involves 58 iterations for each GCV. The following represents the iteration that yields the minimum GCV value.

TABLE 3
MINIMUM GCV WITH ONE KNOT POINT

x_1	x_2	x_3	x_4	GCV
21923,08	63,83	62,42	63,83	0.1634

In Table 3 above, the 3rd iteration stands out with the lowest GCV value, measuring 0.163452, and it corresponds to the knot point. The associated values are as follows: GRDP variable (x_1) at 21,923.08, LFPR variable (x_2) at 63.83, HDI variable (x_3) at 62.42, and OUR variable (x_4) at 63.83. These minimum GCV results will be compared with other GCV outcomes to determine the optimal spline model.

Spline with Two Knots Point

Parameter estimation involving two knot points in the truncated spline nonparametric regression necessitated 1225 iterations for each GCV. The following iteration represents the one with the minimum GCV value.

TABLE 4
MINIMUM GCV WITH TWO KNOTS POINT

x_1	x_2	x_3	x_4	GCV
23683,13	63,98	62,63	63,98	0,105462
75604,51	68,47	68,83	68,47	

Table 4 reveals that the minimum GCV value is achieved during the 49th iteration, with GCV values associated with two knot points totaling 0.105462. Specifically, the values for knots 1 and 2 are as follows: for the GRDP variable x_1, 23,683.13 and 75,604.51; for the LFPR variable x_2, 63.98 and 68.47; for the HDI variable x_3, 62.63 and 68.83; and for the OUR variable x_4, 63.98 and 68.47.

Spline with Three Knot Points

Parameter estimation with three knot points in nonparametric spline truncated regression entails 17,296 iterations for each GCV. The following represents the iteration featuring the minimum GCV value.

TABLE 4
MINIMUM GCV WITH THREE KNOT POINT

x_1	x_2	x_3	x_4	GCV
34067,4	64,88	63,87	64,88	
72143,08	68,17	68,41	68,17	0,0848
75604,51	68,47	68,83	68,47	

In Table 4 above, the minimum GCV value is identified in the 10,084th iteration, with a GCV value of 0.08488212, observed at three knot points. These values are as follows: for the GRDP variable x_1, the knot values are 34,067.4; 72,143.08; 75,604.51. For the LFPR variable x_2, the knot values are 64.88; 68.17; 68.47. As for the HDI variable x_3, the knot values are 63.87; 68.41; 68.83. Similarly, for the OUR variable x_4, the knot values are 64.88; 68.17; 68.47.

E. Model Selection

Following the implementation of nonparametric spline truncated regression modeling with 1, 2, and 3 knot points, the final step involves comparing the minimum GCV values for each knot to select the most effective model. The comparison of minimum GCV values for 1, 2, and 3 knot points is presented as follows:

TABLE 5
GCV VALUE OF EVERY KNOT

Number of Knots	Minimum GCV
1	0,1634
2	0,1054
3	0,0849*

Upon analyzing the GCV values in Table 5 above, it becomes evident that the minimum GCV value is identified at the 3-knot point. Consequently, it can be concluded that the three-knot point model represents the optimal choice for truncated Spline nonparametric regression analysis on the Poverty Severity Index. This model encompasses 17 parameters, including parameter β_0. The following presents a Spline nonparametric regression model derived from three knot points.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_{11}x_1 + \hat{\beta}_{12}(x_1 - K_1)_+^1 + \hat{\beta}_{13}(x_1 - K_2)_+^1 + \hat{\beta}_{14}(x_1 - K_3)_+^1 + \hat{\beta}_{21}x_2 + \hat{\beta}_{22}(x_2 - K_4)_+^1 + \hat{\beta}_{23}(x_2 - K_5)_+^1 + \hat{\beta}_{24}(x_2 - K_6)_+^1 + \hat{\beta}_{31}x_3 + \hat{\beta}_{32}(x_3 - K_7)_+^1 + \hat{\beta}_{33}(x_3 - K_8)_+^1 + \hat{\beta}_{34}(x_3 - K_9)_+^1 + \hat{\beta}_{41}x_4 + \hat{\beta}_{42}(x_4 - K_{10})_+^1 + \hat{\beta}_{43}(x_4 - K_{11})_+^1 + \hat{\beta}_{44}(x_4 - K_{12})_+^1$$

Upon careful evaluation of the minimum GCV value, the parameter estimators for the best model are determined. In accordance with the GCV value calculations, the optimal spline model utilizes three knot points, resulting in the subsequent parameter estimation for the truncated Spline nonparametric regression model with three knot points:

$$\hat{y} = 3{,}11 + 4{,}15x_1 + 7{,}49(x_1 - 34067{,}4)_+^1 - 4{,}14(x_1 - 72143{,}08)_+^1 + 3{,}78(x_1 - 75064{,}51)_+^1 + 1{,}08x_2 - 1{,}23(x_2 - 64{,}88)_+^1 - 1{,}28(x_2 - 68{,}17)_+^1 + 1{,}02(x_2 - 68{,}47)_+^1 - 1{,}65x_3 - 2{,}22(x_3 - 63{,}87)_+^1 + 1{,}96(x_3 - 68{,}42)_+^1 + 2{,}40(x_3 - 68{,}83)_+^1 + 1{,}08x_4 - 1{,}23(x_4 - 64{,}88)_+^1 - 1{,}27(x_4 - 68{,}17)_+^1 + 1{,}02(x_4 - 68{,}47)_+^1$$

F. Parameter Significance Test of Spline Truncated Nonparametric Regression Model

The subsequent step involves conducting a significance test on the model parameters to ascertain their impact on the open unemployment rate, once the parameter estimates of the truncated Spline nonparametric regression model are obtained. This parameter test is carried out in two stages—partial and simultaneous testing. The partial test is conducted if the results of the simultaneous test suggest that only a single parameter holds significance.

Simultaneous Testing

To determine whether the independent variables, when used simultaneously, exert an influence on the model, a simultaneous test is employed. This test will assess the following hypothesis simultaneously.

$H_0 : \beta_1 = \beta_2 = ... = \beta_{16} = 0$
H_1 : at least there is one $\beta_0 \neq 0; l = 1, 2, ..., q + r$

The following is a table of simultaneous test results using the *F* test described in the ANOVA table.

TABLE 6
SIMULTANEOUS PARAMETER TESTING RESULTS

Source of Variation	df	Sum of Squares	Middle Square	F
Regression	16	6,5859	0,4116	4,485
Error	17	1,5603	0,0917	
Total	33	8,1463		

Based on Table 6, the F-value test statistic registers at 4.485, with a corresponding p-value of 0.00185. With a p-value < α (0.05), the decision is to reject the null hypothesis. This implies that at least one parameter or independent variable exerts a substantial impact on the model, signifying that the Indonesian poverty severity index is significantly influenced by at least one independent variable.

Partial Testing

The partial test will proceed as the results of the simultaneous test indicate the presence of one significant variable. A partial test is conducted with the aim of assessing the individual impact of each variable on the regression model. The hypothesis for the partial test is as follows:

$H_0 : \beta_j = 0$

H_1 : at least there is $\beta_j \neq 0$

TABLE 7
PARTIAL PARAMETER TESTING RESULTS

Variables	Parameter	Estimation	p-value	Decision
Constant	β_0	3,11	8,21	Significant
x_1	β_1	-4,151	0,0062	Significant
	β_2	7,49	0,0043	Significant
	β_3	-4,14	0,014	Significant
	β_4	3,78	0,015	Significant
x_2	β_5	1,08	0,0001	Significant
	β_6	-1,33	0,0003	Significant
	β_7	-1,27	0,45	not significant
	β_8	1,02	0,96	not significant
x_3	β_9	-1,65	4,59	Significant
	β_{10}	-2,22	0,00044	Significant
	β_{11}	1,96	0,00034	Significant
	β_{12}	2,40	0,0001	Significant
x_4	β_{13}	1,08	0,0001	Significant
	β_{14}	-1,23	0,0003	Significant
	β_{15}	-1,28	0,45	not significant
	β_{16}	1,02	0,96	not significant

As observed in Table 7, a p-value < α (0.05) leads to the rejection of the null hypothesis. This conclusion indicates that only 12 parameters hold significance in the regression model, while the remaining 4 parameters do not. Notably, even though there are 4 parameters within the variables x_2 and x_3 that lack significance, these variables can still be retained in the analysis. This decision is informed by the partial test, where at least one variable exhibits a significant parameter. Thus, it can be inferred that independent factors do indeed exert a substantial influence on the dependent variable.

IV. DISCUSSION

The most effective regression model developed for modeling the open unemployment rate in Indonesian provinces incorporates three knot points, resulting in a GCV value of 0.08488212. Additionally, the model exhibits R^2 value of 80.85%. The R^2 value, representing the coefficient of determination, serves as an indicator of the variables' collective influence. In this case, the R^2 value of 80.85% explains the impact on the open unemployment rate, while the remaining 19.15% can be attributed to other variables not included in the model.

To ensure the validity of the model we developed, comparative studies with other models were conducted using the same dataset and modeled through multiple linear regression. The outcomes revealed R^2 value of 38.01%. When contrasted with the previously obtained R^2 value, it becomes evident that our proposed model significantly outperforms the alternative model.

The following is the best model using three selected knot points from nonparametric regression analysis using a truncated spline:

$$\hat{y} = 3,11 - 4,15x_1 + 7,49(x_1 - 34067,4)_+^1 - 4,14(x_1 - 72143,08)_+^1 + 3,78(x_1 - 75064,51)_+^1 + 1,08x_2 - 1,23(x_2 - 64,88)_+^1 - 1,28x_2(x_2 - 68,17)_+^1 + 1,02(x_2 - 68,47)_+^1 - 1,65x_3 - 2,22(x_3 - 63,87)_+^1 + 1,96(x_3 - 68,42)_+^1 + 2,40(x_3 - 68,83)_+^1 + 1,08x_4 - 1,23(x_4 - 64,88)_+^1 - 1,27(x_4 - 68,17)_+^1 + 1,02(x_4 - 68,47)_+^1$$

The regression model demonstrates that each of the independent variables employed exerts a substantial impact on the poverty severity index. The interpretation of these variable effects is as follows.

1. The effect of the relationship between the Gross Regional Domestic Income based on Price (GRDP) variable x_1 on the Poverty Severity Index y in Indonesia in 2022.

$$\hat{y} = -4,15x_1 + 7,49(x_1 - 34067,4)_+^1 - 4,14(x_1 - 72143,08)_+^1 + 3,78(x_1 - 75064,51)_+^1$$

$$= \begin{cases} -4,15x_1 & ; x_1 < 34067,4 \\ 3,34x_1 - 225164,8 & ; 34067,4 \le x_1 \le 72143,08 \\ -0,8x_1 - 73507,55 & ; 72143,08 \le x_1 \le 75064,51 \\ 2,59x_1 - 357251,4 & ; x_1 \ge 75064,51 \end{cases}$$

The model illustrates that province with a GRDP (x_1) below 34,067.4% experience a 4.15% reduction in the poverty severity index for each unit increase in GRDP. This category includes 14 provinces: East Nusa Tenggara, Maluku, West Nusa Tenggara, West Sulawesi, Bengkulu, Gorontalo, Aceh, West Kalimantan, Lampung, Central Java, DI Yogyakarta, North Maluku, West Java, and West Sumatra.

Provinces with a GRDP (x_1) falling in the range of 34,067.4% to 72,143.08% witness a 3.34% increase in the poverty severity index for every unit increase in GRDP. This category encompasses 15 provinces in Indonesia: South Kalimantan, Bali, North Sulawesi, North Sumatra, Southeast Sulawesi, Bangka Belitung, South Sulawesi, Papua, Banten, South Sumatra, Central Kalimantan, East Java, Jambi, West Papua, and Central Sulawesi.

Provinces with a GRDP (x_1) in the range of 72,143.08% to 75,064.51% experience a 0.8% reduction in the poverty severity index for each unit increase in GRDP. There are no provinces in Indonesia falling within this category.

On the other hand, provinces with a GRDP (x_1) exceeding 75,064.51% see a 2.59% increase in the poverty severity index for every unit increase in GRDP. This category comprises 5 provinces: Riau, Kep. Riau, North Kalimantan, East Kalimantan, and DKI Jakarta.

The classification of provinces in Indonesia based on their GRDP in 2022 can be visually represented in Figure 3 below, assuming other variables remain constant.

Figure 3 reveals that the majority of provinces in

Indonesia have a GRDP below the average. This aligns with the findings of the descriptive statistics analysis, which indicates an average GRDP value of 452,780.5.

2. The effect of the relationship between the Labor Force Participation Rate (LFPR) variable (x_2) on the Poverty Severity Index y in Indonesia in 2022.

$$\hat{y} = 1,08x_2 - 1,23(x_2 - 64,88)_+^1 - 1,28x_2(x_2 - 68,17)_+^1 + 1,02(x_2 - 68,47)_+^1$$

$$= \begin{cases} 1,08x_2 & ; x_2 < 64,88 \\ -0,15x_2 + 79,802 & ; 64,88 \le x_2 \le 68,17 \\ -2,33x_2 + 167,05 & ; 68,17 \le x_2 \le 68,47 \\ -1,31x_2 + 97,210 & ; x_2 \ge 68,47 \end{cases}$$

The model indicates that provinces with LFPR (x_2) below 64.88% experience a 1.08% increase in the poverty severity index for every 1-unit rise in LFPR. Six provinces fall into this category: DKI Jakarta, North Sulawesi, Aceh, Riau, Banten, and East Kalimantan.

Provinces with LFPR (x_2) ranging from 64.88% to 68.17% witness a 0.15% decrease in the poverty severity index for every 1-unit increase in LFPR. Nine provinces are included in this group: North Maluku, Maluku, West Java, South Sulawesi, Central Kalimantan, Bangka Belitung, South Kalimantan, North Kalimantan, and Jambi.

Provinces with LFPR (x_2) between 68.17% and 68.47% experience a 2.33% decrease in the poverty severity index for every 1-unit increase in LFPR. There are no provinces in Indonesia falling within this category.

Conversely, provinces with LFPR (x_2) above 68.47% see a 1.31% reduction in the poverty severity index for every 1-unit increase in LFPR. This category comprises 19 provinces: West Papua, Southeast Sulawesi, Gorontalo, Riau, West Kalimantan, West Sumatra, South Sumatra, North Sumatra, Bengkulu, Central Sulawesi, Lampung, Central Java, West Nusa Tenggara, East Java, DI Yogyakarta, West Sulawesi, East Nusa Tenggara, Bali, and Papua.

The classification of provinces in Indonesia based on their LFPR in 2022 can be visually represented in Figure 4, assuming other variables remain constant.

Figure 4 reveals that the majority of provinces in Indonesia already have LFPR levels exceeding the provincial average. This aligns with the findings of the descriptive statistics analysis, which indicates an average LFPR value of 68.64.

3. Assuming all other factors remain constant, the impact of the Human Development Index (HDI) variable on the Poverty Severity Index in Indonesia in 2022 is as follows.

$$\hat{y} = -1,65x_3 - 2,22(x_3 - 63,87)_+^1 + 1,96(x_3 - 68,42)_+^1 + 2,40(x_3 - 68,83)_+^1$$

$$= \begin{cases} -1,65x_3 & ; x_3 < 63,87 \\ -3,87x_3 - 141,79 & ; 63,87 \le x_3 \le 68,42 \\ -1,91x_3 - 275,89 & ; 68,42 \le x_3 \le 68,83 \\ 0,49x_3 - 441,08 & ; x_3 \ge 68,83 \end{cases}$$

The model indicates that provinces with HDI (x_3) below 63.87% experience a 1.65% reduction in the poverty severity index for every 1-unit increase in HDI. Only one province falls into this category, namely Papua Province.

Provinces with HDI (x_3) ranging from 63.87% to 68.42% see a 3.87% decrease in the poverty severity index for every 1-unit increase in HDI. Three provinces fall into this category: West Papua, East Nusa Tenggara, and West Sulawesi.

Provinces with HDI (x_3) between 68.43% and 68.83% experience a 1.91% decrease in the poverty severity index for every 1-unit increase in HDI. There is one province in this category, namely West Kalimantan Province.

Conversely, provinces with HDI (x_3) above 68.83% witness a 0.49% increase in the poverty severity index for every 1-unit increase in HDI. This category comprises 29 provinces: West Nusa Tenggara, North Maluku, Gorontalo, Maluku, Central Sulawesi, Lampung, South Sumatra, Central Kalimantan, North Kalimantan, South Kalimantan, Jambi, Bengkulu, Southeast Sulawesi, Bangka Belitung, North Sumatra, East Java, Central Java, Aceh, South Sulawesi, West Java, West Sumatra, Banten, Riau, North Sulawesi, Bali, Riau, East Kalimantan, DI Yogyakarta, and DKI Jakarta.

The classification of provinces in Indonesia based on their HDI in 2022 can be visually represented in Figure 5, assuming other variables remain constant.

Figure 5 illustrates that the majority of regions in Indonesia already boast an HDI surpassing the average, as depicted in the image above. This aligns with the findings of the descriptive statistics analysis, which indicates an average HDI of 71.99.

4. The result of the connection between the Open Unemployment Rate (UR) variable (x_4) on the Poverty Severity Index y in Indonesia in 2022 assuming other variables are held constant is as follows.

$$\hat{y} = 1,08x_4 - 1,23(x_4 - 64,88)_+^1 - 1,27(x_4 - 68,17)_+^1 + 1,02(x_4 - 68,47)_+^1$$

$$= \begin{cases} 1,08x_4 & ; x_4 < 64,88 \\ -0,15x_4 + 79,802 & ; 64,88 \le x_4 \le 68,17 \\ -2,33x_4 + 167,05 & ; 68,17 \le x_4 \le 68,47 \\ -1,31x_4 + 97,210 & ; x_4 \ge 68,47 \end{cases}$$

The model reveals that provinces with UR (x_4) below 64.88% lead to an increase of 1.08% in the poverty severity index for every 1-unit increase in UR. Six provinces fall into this category, including DKI Jakarta, North Sulawesi, Aceh, Riau, Banten, and East Kalimantan.

Provinces with UR (x_4) within the range of 64.88% to

68.17% result in a decrease of 0.15% in the poverty severity index for every 1-unit increase in UR. This category includes nine provinces: North Maluku, Maluku, West Java, South Sulawesi, Central Kalimantan, Bangka Belitung, South Kalimantan, North Kalimantan, and Jambi.

Provinces with UR (x_4) in the interval of 68.17% to 68.47% experience a decrease of 2.33% in the poverty severity index for every 1-unit increase in UR. Currently, no provinces in Indonesia fall into this category.

Conversely, provinces with UR (x_4) above 68.47% lead to a decrease of 1.31% in the poverty severity index for every 1-unit increase in OUR. This category includes 19 provinces: West Papua, Southeast Sulawesi, Gorontalo, Riau, West Kalimantan, West Sumatra, South Sumatra, North Sumatra, Bengkulu, Central Sulawesi, Lampung, Central Java, West Nusa Tenggara, East Java, DI Yogyakarta, West Sulawesi, East Nusa Tenggara, Bali, and Papua.

The visual representation of provincial classification by UR in Indonesia in 2022, assuming other variables remain constant, can be seen in Figure 6.

The Figure 6 illustrates that the majority of provinces in Indonesia have an UR exceeding the provincial average. This observation aligns with the results of the analysis based on descriptive statistics, which indicate that the average value of UR is 68.64.

V. Conclusion

In this paper, our primary objective is to introduce a multiple Spline regression model as a classification tool. We selected the Spline model due to its capacity to estimate the regression curve by adapting to changes in data patterns within specific distribution intervals. Its ability to mirror the data pattern's shape serves as the foundation for creating a classification tool. The results demonstrate that the Spline model effectively models the data, as evidenced by the relatively high R^2 value. The resulting mathematical model can also serve as a classification tool, as indicated by the creation of a map. This model is expected to offer an alternative statistical tool for classification.

These discoveries carry important implications for estimating regression curves. While the study concentrated on Splines, the techniques explored can be extended to other methods for nonparametric regression. However, this research's focus is narrow as it doesn't involve hypothesis or confidence interval testing. Recognizing the significance of robust statistical methods, there's a plan to perform thorough model adequacy tests. Consequently, future research will aim to incorporate hypothesis and confidence interval testing to enhance the findings.

Fig. 3. Distribution of Provinces Based on GRDP

Fig. 4. Distribution of Provinces by LFPR

Fig. 5. Distribution of Provinces Based on HDI

Fig. 6. Distribution of Provinces by UR

REFERENCES

[1] Fehling, Maya, D. Brett, Nelson, and S. Venkatapuram, "Limitations of the Millennium Development Goals: a literature review," *Global public health*, vol. 8, no. 10, pp. 1109-1122, 2013.

[2] BPS—Statistics Indonesia. Indeks Pembangunan Manusia 2018; BPS—Statistics Indonesia: Jakarta, Indonesia, 2019.

[3] Ningrum, P. Anggriati, A. Hukom, and S. Adiwijaya, "The Potential of poverty in the City of Palangka Raya: Study SMIs Affected Pandemic Covid 19," *Budapest Int. Res. and Critics Institute-Journal (BIRCI-Journal)*, vol. 3, pp. 1626-1634, 2020.

[4] R. L. Eubank, "Nonparametric regression and spline smoothing, 2nd ed., Marcel Dekker, Inc.: New York, NY, USA, 1999.

[5] J. H. Friedman, "Multivariate adaptive regression splines. *The ann. of stat*, vol. 19, no. 1, pp. 1-67, 1991.

[6] J. Mennis, "Mapping the results of geographically weighted regression," *The cartographic journal*, vol. 43, no. 2, pp. 171-179, 2006.

[7] Y. Hanchen, A. S. Fotheringham, Z. Li, T. Oshan, W. Kang, and L. J. Wolf, "Inference in multiscale geographically weighted regression," *Geographical Anal.*, vol. 52, no. 1, pp. 87-106, 2020.

[8] L. Ziqi and A. S. Fotheringham, "Computational improvements to multi-scale geographically weighted regression," *Int. J. of Geographical Information Science*, vol. 34, no. 7, pp. 1378-1397, 2020.

[9] G. Yuejing, J. Zhao, and L. Han, "Exploring the spatial heterogeneity of urban heat island effect and its relationship to block morphology with the geographically weighted regression model," *Sustainable Cities and Society*, vol. 76, pp. 103-431, 2022.

[10] M. E. A. Budimir, P. M. Atkinson, and H. G. Lewis, "A systematic review of landslide probability mapping using logistic regression," *Landslides*, vol. 12, pp. 419-436, 2015.

[11] R. Greco, S. Valvo, and E. Catalano, "Logistic regression analysis in the evaluation of mass movements susceptibility: The Aspromonte Case Study, Calabria, Italy. *Eng. Geol.*, vol. 89, pp. 47–66, pp. 2007.

[12] S. Lee, J. H. Ryu, K. Min, J. S. Won, "Landslide susceptibility analysis using GIS and artificial neural network," *Earth Surface Processes and Landforms*, vol. 28, pp. 1361– 1376, 2003.

[13] L. Ermini, F. Catani, N. Casagli, "Artificial neural networks applied to landslide susceptibility assessment," *Geomorphology*, vol. 66, pp. 327–343, 2005.

[14] W. Yi, Z. Fang, and H. Hong, "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China," *Science of the total environment*, vol. 666, pp. 975-993, 2005.

[15] B. Pradham, E. A. Sezer, C. Gokceoglu, and M. F. Buchroithner, Landslide susceptibility mapping by neuro-fuzzy approach in a landslide prone area (Cameron Highland, Malaysia), *IEEE Trans Geosci. Rem. Sens.*, vol. 48, pp. 4164–4177, 2010.

[16] Y. K. Yeon, J. G. Han, and K. H. Ryu, "Landslide susceptibility mapping in Injae, Korea, using a decision tree," *Eng. Geol.*, vol. 116, pp. 274–283, 2010.

[17] I. Yilmaz, "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Katlandslides (Tokat–Turkey)," *Com. Geosci.*, vol. 35, pp. 1125–1138, 2009.

[18] R. Hidayat, M. Ilyas, and Y. Yuliani, "Spline model in the case of cervical cancer patient resilience," *Trends in Sciences*, vol. 20, no. 8, pp. 6170-6180, 202. https://doi.org/10.48048/tis.2023.6170.

[19] R. Hidayat, "Integral estimator with kernel approach for estimating nonparametric regression functions," *Journal of Physics: Conference Series*, IOP Publishing, vol. 2123, no. 1, pp. 12-22, 2021.

[20] W. Yun, Y. Li, R. Zou, M. Aoife, Foley, D. A. Kez, D. Song, Q. Hu, and D. Srinivasan, "Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 191-201, 2020.

[21] Adams, S. Olorunfemi, D. A. Obaromi, and A. A. Irinews, "Goodness of fit test of an autocorrelated time series cubic smoothing Spline model," *Journal of the Nigerian Society of Physical Sciences*, pp. 191-200, 2021.

[22] Edwards, R. Jeffrey, and M. E. Parry, "On the use of spline regression in the study of congruence in organizational research," *Organizational Research Methods*, vol. 21, no. 1, pp. 68-110, 2018.

[23] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "A reproducing kernel Hilbert space approach and smoothing parameters

selection in spline-kernel regression," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 2, pp. 465-475, 2019.

Rahmat Hidayat born on March 12, 1990, is a lecturer and researcher at Mathematics Department, Faculty of Science, Universitas Cokroaminoto Palopo, Indonesia. He earned a bachelor's degree in mathematics (S.Si.) from State University of Makassar, Indonesia. He obtained a M.Sc. degree at the Department of applied mathematics, form IPB University, and doctoral (Dr.) degree in statistics from Institut Teknologi Sepuluh Nopmeber, Surabaya. Research interest in nonparametric regression, survival analysis, and computational statistics.

Muhammad Ilyas born on July 21, 1965, is a professor at Master of Mathematics Education Program, Universitas Cokroaminoto Palopo, Indonesia. He earned a bachelor's degree in mathematics education (S.Pd.) from State University of Makassar, Indonesia. He obtained a M.Ed. degree from State University of Makassar, Indonesia, and doctoral (Dr.) degree from Jakarta State University. Research interest in Education Management and Mathematics Education.

Yuliani born on July 25, 1990, is a lecturer at Mathematics Department, Faculty of Science, Universitas Cokroaminoto Palopo, Indonesia. He earned a bachelor's degree in mathematics (S.Si.) from Hasanuddin University, Indonesia. He obtained a M.Sc. degree at the Department of applied mathematics, form Hasanuddin University. Research interest in applied mathemtics.