Modeling with Robust Kernel Nonparametric Regression on Childhood Stunting in Kalimantan

Samsul Arifin[1*], Selvi Annisa[1], and Siswanto[2]
[1]Department of Statistics, Lambung Mangkurat University, Banjarbaru, Indonesia
[2]Department of Statistics, Hasanuddin University, Makassar, Indonesia
*Corresponding author: samsularrr@ulm.ac.id

Received: 31 May 2026
Revised: 6 August 2026
Accepted: 8 August 2026

ABSTRACT – **This study models stunting prevalence across 56 regencies/cities in Kalimantan using robust kernel nonparametric regression. This approach addresses the nonlinear relationship between stunting and four predictors: access to improved sanitation, low birth weight, population density, and poverty rate. An examination of influential observations using DFFITS identified five regencies as outliers; thus, the robust MM-estimator approach was applied to mitigate the influence of these extreme observations on the estimation results. Optimal bandwidth selection was performed using the Cross-Validation (CV) method across several kernel functions, namely Epanechnikov, Gaussian, and Uniform. The results demonstrated that the Uniform kernel function yielded the smallest CV value with a bandwidth combination of $h_1 = 0.6$, $h_2 = 0.2$, $h_3 = 0.6$, and $h_4 = 0.2$. The Robust Uniform Kernel model delivered the best performance, with an MSE of 2.0556, RMSE of 1.4337, MAE of 0.6581, and R^2 of 0.9510. This study indicates that robust MM-estimator kernel nonparametric regression can produce stunting prevalence estimates that are more accurate, flexible, and stable in the presence of outliers.**

Keywords – **Stunting, Kernel Nonparametric Regression, Robust MM-Estimator, Uniform Kernel, Kalimantan.**

I. INTRODUCTION

Stunting remains a critical chronic nutritional problem in public health, reflecting impaired child growth due to long-term malnutrition [1]. Its impacts extend beyond physical deficits, such as low height-for-age [2], to impairments in cognitive development [3], weakened immune systems [4], and reduced adult productivity [5]. Stunting is influenced by a complex interplay of biological, social, economic, environmental, and healthcare access factors [6]. Specifically, factors such as low birth weight [7], lack of access to improved sanitation [8], poverty [9], maternal education [10], and child healthcare quality is frequently associated with increased stunting risk [11]. Consequently, stunting must be addressed not merely as an individual health issue but as a human resource development challenge that requires a comprehensive analytical approach.

In Indonesia, stunting remains a priority despite a recent decline in prevalence over the past few years [12]. According to the 2024 Indonesian Nutritional Status Survey (SSGI), the national stunting prevalence stood at 19.8%, down from 21.5% the previous year [13]. Although this decline demonstrates progress, a prevalence affecting nearly one in five children under five indicates that the issue still demands critical attention. This challenge is particularly evident in Kalimantan, where stunting rates are relatively high and vary considerably across provinces. The 2024 SSGI report noted that stunting prevalence reached 26.8% in West Kalimantan, 22.1% in Central Kalimantan, 22.9% in South Kalimantan, 22.3% in East Kalimantan, and 17.6% in North Kalimantan. These figures show that most provinces in Kalimantan still exceed the national average, with North Kalimantan the sole exception. Consequently, analyzing stunting in Kalimantan requires a flexible modeling approach that captures the complex relationships among variables.

Previous studies have predominantly used parametric regression approaches, such as linear [14], logistic [15], and panel regression [16], to model stunting in Indonesia. While parametric regression offers straightforward interpretation, it assumes a predetermined functional relationship (e.g., linear, quadratic, or cubic) between the response and predictor variables [17]. In reality, the relationships between stunting incidence and its driving factors are highly complex and rarely follow simple, rigid patterns. To address this limitation, a nonparametric regression approach offers a more flexible alternative, allowing the regression curve to be driven entirely by data patterns without imposing a specific functional form [18]. Commonly used nonparametric estimators include local polynomial [19], spline [20], and kernel estimators [21]. Among these, kernel nonparametric regression is widely favored because it assigns greater weight to observations close to the estimation point and less weight to distant ones, thereby capturing local patterns effectively [22]. Compared to other estimators, the kernel method is relatively simple to compute, eliminates the need to determine knot points as required by splines, and allows precise control over curve smoothness through bandwidth selection [23]. Therefore, kernel nonparametric regression is well-suited for modeling nonlinear stunting data.

In practice, standard kernel nonparametric regression remains highly sensitive to outliers or extreme values. These outliers can severely distort estimation, pulling the regression curve away from the overall data trend [24]. This distortion leads to unstable estimates, inflated error estimates, and inaccurate interpretations of relationships among variables. To mitigate this issue, a robust approach can be integrated into the kernel nonparametric regression framework. Robust methods aim to neutralize the distorting effects of outliers without dropping data points from the dataset [25]. Common robust estimation methods include M-estimation [26], S-estimation [27], and MM-estimation [28]. Among these, the MM-estimator is superior because it combines high breakdown resistance against outliers with excellent statistical efficiency.

It begins with a highly robust preliminary estimate, followed by a more efficient estimation process to yield stable final parameters. Integrating a robust MM-estimator into kernel nonparametric regression is thus expected to produce a model that is both resistant to outliers and flexible enough to capture the nonlinear dynamics of stunting data.

Driven by these challenges, this study develops a robust kernel-based nonparametric regression model to map the incidence of child stunting in Kalimantan. The framework is expected to yield a flexible modeling tool that accommodates nonlinear trends and is highly robust to outliers. Ultimately, this research aims to advance nonparametric regression methodologies and provide highly accurate empirical insights to support evidence-based stunting intervention policies in Indonesia.

II. LITERATURE REVIEW

A. Kernel Nonparametric Regression

Kernel regression is a nonparametric statistical method used to estimate the conditional expectation of a response variable given predictor variables by applying a kernel function. It is widely utilized due to its flexibility, relatively straightforward estimation procedure, and ability to adapt to data relationship patterns without predefining a specific functional form [29]. Unlike parametric regression, which assumes a predetermined relationship, such as linear or quadratic, kernel nonparametric regression allows the shape of the regression curve to be driven entirely by the data. Consequently, this approach is highly appropriate when the relationship between the response and predictor variables is non-linear or mathematically unknown. The kernel nonparametric regression model is given by (1) [30].

$$\begin{aligned} Y_i &= m(X_i) + \varepsilon_i, \qquad i = 1,2,\dots,n \\ m(x) &= E(Y \mid X = x) \end{aligned} \tag{1}$$

where Y_i is the response variable at the i-th observation, x_i is the value of the predictor variable at the i-th observation, $m(x_i)$ is the regression function with an unknown curve shape, ε_i is the random error, and n is the number of observations. The function $m(x)$ represents the conditional expectation of the response variable Y given that the predictor variable X equals x.

One commonly used estimator in kernel regression is the Nadaraya-Watson estimator. This estimator works by assigning a larger weight to observations close to the estimation point x, and a smaller weight to observations that are far from x [31]. The Nadaraya-Watson estimator is given by (2).

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K_h\,(x - X_i) Y_i}{\sum_{i=1}^{n} K_h\,(x - X_i)} \tag{2}$$

where

$$K_h(x - X_i) = K\left(\frac{x - X_i}{h}\right)$$

or it can be expressed as shown in (3)

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)} \tag{3}$$

where K is the kernel function and h is the bandwidth. The bandwidth is a crucial parameter in kernel regression as it determines the smoothness level of the regression curve.

Several kernel functions commonly used in nonparametric regression include the Gaussian, Epanechnikov, and Uniform kernels. These kernel function forms can be expressed as shown in Equations 4–6 [32].

$$\text{Gaussian: } K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) \tag{4}$$

$$\text{Epanechnikov: } K(u) = \frac{3}{4}(1 - u^2) I(\mid u \mid \leq 1 \tag{5}$$

$$\text{Uniform: } K(u) = \frac{1}{2} I(\mid u \mid \leq 1) \tag{6}$$

where $u = \frac{x - X_i}{h}$ and $I(\mid u \mid \leq 1)$ is an indicator function that takes the value of 1 if $\mid u \mid \leq 1$ and 0 otherwise. Selecting the appropriate kernel function and bandwidth is crucial in kernel regression, as both significantly influence the estimated regression curve.

B. Bandwidth Selection

Bandwidth is a crucial parameter in kernel nonparametric regression as it determines the smoothness level of the estimated curve. A bandwidth value that is too small yields a curve that overfits the data patterns, leading to an unstable model; conversely, an excessively large bandwidth produces an overly smooth curve, causing local data patterns to be poorly captured [33]. Therefore, selecting the optimal bandwidth is a critical step in kernel regression. One commonly used method for determining the optimal bandwidth is Cross-Validation (CV). The CV method is widely used in kernel regression for evaluating the model's performance on data not directly involved in the estimation process. In kernel nonparametric regression, bandwidth selection via CV is performed using the leave-one-out approach, in which each observation is alternately removed from the estimation process, and its response is predicted from the remaining data. Mathematically, the CV function is given by (7).

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{m}_{-i}(X_i))^2 \tag{7}$$

where $\widehat{m}_{-i}(X_i)$ is the kernel regression estimate at point X_i computed by omitting the i-th observation. This estimator is given by (8).

$$\widehat{m}_{-i}(X_i) = \frac{\sum_{j\neq i}^{n} K\left(\frac{X_i - X_j}{h}\right) Y_j}{\sum_{j\neq i}^{n} K\left(\frac{X_i - X_j}{h}\right)} \tag{8}$$

The optimal bandwidth value is obtained by selecting the h value that minimizes the CV function, as shown in (9).

$$h_{opt} = \arg\min_{h} CV(h) \tag{9}$$

Consequently, the optimal bandwidth is the one that yields the smallest prediction error. Bandwidth selection is performed by defining a lower bound, an upper bound, and a bandwidth increment value. Subsequently, each candidate's bandwidth is evaluated using the $CV(h)$ value, and the bandwidth with the minimum CV value is chosen as the optimal bandwidth.

C. Robust Regression

Robust regression is an estimation approach developed to mitigate the influence of outliers on modeling results. In classical regression, particularly the Ordinary Least Squares (OLS) method, parameters are obtained by minimizing the sum of squared residuals, as shown in (10).

$$\hat{\beta}_{OLS} = \arg\min_{\beta}\sum_{i=1}^{n} e_i^2 \tag{10}$$

where $e_i = y_i - \hat{y}_i$ represents the residual of the i-th observation. The primary weakness of OLS is its reliance on the squared residual; consequently, observations with large residuals exert a disproportionate influence on the estimates. Therefore, when data contain outliers, OLS estimates can become unstable and fail to represent the general data pattern adequately.

One widely used robust approach is the M-estimator. The M-estimator is a generalization of maximum likelihood estimation that replaces the squared residual function with a robust objective function, ρ. In general, the M-estimator is obtained by solving (11).

$$\hat{\beta}_{M} = \arg\min_{\beta}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) \tag{11}$$

where s is a robust residual scale measure, while $\rho(\cdot)$ is the robust objective function. The function ρ is selected such that small residuals continue to contribute to the estimation, whereas the influence of large residuals is bounded. Its estimation equation is given by (12).

$$\sum_{i=1}^{n} \psi\left(\frac{e_i}{s}\right) x_i = 0 \tag{12}$$

where $\psi(u) = \frac{\partial\rho(u)}{\partial u}$ represents the influence function. In terms of weighting, the M-estimator can be written as shown in (13).

$$\sum_{i=1}^{n} w_i\, e_i x_i = 0 \tag{13}$$

where the robust weight is $w_i = \frac{\psi(u_i)}{u_i}$ and $u_i = \frac{e_i}{s}$.

In addition to the M-estimator, the S-estimator was developed to obtain estimates with high breakdown resistance to outliers. The S-estimator operates by minimizing a robust residual scale rather than directly minimizing the sum of residuals. In general, the S-estimator can be expressed as shown in (14).

$$\hat{\beta}_{S} = \arg\min_{\beta} s(e_1, e_2, \dots, e_n) \tag{14}$$

where the residual scale s satisfies (15).

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = b \tag{15}$$

The value b is a constant determined based on the specific function ρ utilized. The primary advantage of the S-estimator is its high breakdown point, making it more robust to outliers. However, the S-estimator tends to be less efficient than classical estimators when the data contain few outliers.

To combine the advantages of both the M-estimator and the S-estimator, the MM-estimator is employed. The MM-estimator was introduced as a robust estimator that possesses two critical properties: a high breakdown point and high statistical efficiency [34]. The procedure in MM-estimation is executed in three stages. First, an initial estimate is computed using a robust estimator with a high breakdown point, such as the S-estimator shown in (16).

$$\hat{\beta}_0 = \hat{\beta}_S \tag{16}$$

Second, a robust residual scale is computed based on the initial estimate, as shown in (17).

$$s = s(e_1, e_2, \dots, e_n) \tag{17}$$

Third, the final estimate is determined using the M-estimator principle with the previously obtained residual scale, as shown in (18).

$$\hat{\beta}_{MM} = \arg\min_{\beta} \sum_{i=1}^{n} \rho_1 \left(\frac{y_i - x_i^T \beta}{s} \right) \tag{18}$$

where ρ_1 is a robust objective function selected to yield high estimation efficiency. Consequently, the MM-estimator uses the S-estimator to obtain initial robustness to outliers, and subsequently employs the M-estimator to improve the efficiency of the estimates.

D. Robust Kernel Nonparametric Regression

Robust kernel nonparametric regression is an extension of kernel nonparametric regression that integrates kernel weights with robust weights based on the MM-estimator. This approach is employed to overcome the limitations of conventional kernel regression, which is highly sensitive to outliers or extreme values. In standard kernel regression, the estimated regression function is obtained via local weighted averages based on the proximity between observations. However, if observations with extreme response values are present, they can still distort the shape of the estimated curve, particularly if they are located close to the estimation point. Therefore, robust weighting is required to mitigate the influence of extreme observations. In general, the Nadaraya-Watson estimator for kernel nonparametric regression is given by (19).

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K_h\,(x - X_i) Y_i}{\sum_{i=1}^{n} K_h\,(x - X_i)} \tag{19}$$

This estimator considers only the proximity of observations to the estimation point, without accounting for whether any observation is an outlier.

In robust MM-estimator kernel nonparametric regression, kernel weights are combined with robust weights obtained from the model residuals. The form of the robust MM kernel nonparametric regression estimator is given in (20).

$$\widehat{m}_{MM}(x) = \frac{\sum_{i=1}^{n} K_h\,(x - X_i) w_i^{MM} Y_i}{\sum_{i=1}^{n} K_h\,(x - X_i) w_i^{MM}} \tag{20}$$

where w_i^{MM} represents the robust MM weight for the i-th obeservation. This weight is assigned based on the magnitude of the residual. Observations with small residuals will receive relatively large weights, whereas observations with large residuals will receive smaller weights. Consequently, outliers are not removed from the data, but their influence on the estimation results is mitigated.

The initial residual in kernel regression is given by (21).

$$e_i = Y_i - \widehat{m}(X_i) \tag{21}$$

Subsequently, the residuals are standardized using a robust scale s, yielding (22).

$$u_i = \frac{e_i}{s} \tag{22}$$

The robust MM weights can then be expressed in a general form as shown in (23).

$$w_i^{MM} = \frac{\psi(u_i)}{u_i} \tag{23}$$

where $\psi(u_i)$ represents the robust influence function. One commonly used weighting function in robust estimation is the Tukey Bisquare function, as it can assign small weights or even weights approaching zero to observations with exceptionally large residuals. In general, the Tukey Bisquare weight can be expressed as shown in (24).

$$w_i^{MM} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |\, u_i \,| \leq c \\ 0, & |\, u_i \,| > c \end{cases} \tag{24}$$

where c is a tuning constant that regulates the resilience and efficiency of the estimator. If the standardized residual value $|\, u_i \,|$ lies within the bound c, the observation is still utilized with a specific weight. Conversely, if $|\, u_i \,| > c$, the observation is considered highly extreme, rendering its influence on the estimation negligible or zero.

E. Model Goodness-of-Fit Evaluation

Model goodness-of-fit evaluation assesses the model's ability to produce estimates close to the actual values. In this study, the performance of the robust MM kernel nonparametric regression model is evaluated using several prediction error metrics, namely Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These four metrics are utilized to compare model performance between conventional kernel regression and robust MM kernel regression [35].

1) Mean Squared Error (MSE) is used to measure the average squared difference between the actual and estimated values. MSE can be expressed as shown in (25).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{25}$$

2) Root Mean Squared Error (RMSE) is the square root of MSE, meaning it shares the same unit of measurement as the response variable. RMSE is expressed as shown in (26).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{26}$$

3) Mean Absolute Error (MAE) is used to compute the average of the absolute differences between the actual and estimated values. MAE is expressed as shown in (27).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i| \tag{27}$$

4) In addition to the prediction error metrics, this study also utilizes the coefficient of determination (R^2) to examine the proportion of variance in the response variable that the model can explain. The R^2 value is expressed as shown in (28).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{28}$$

where Y_i is the actual value, $\hat{Y}_i$ is the model's estimated value, and $\bar{Y}$ is the mean of the actual values. A robust model is characterized by low MSE, RMSE, and MAE values, alongside a high R^2 value.

F. Stunting

Stunting is a condition characterized by impaired growth and development in children, driven by chronic malnutrition, recurrent infections, and long-term environmental factors. Anthropometrically, a child is classified as stunted if their height-for-age z-score falls below minus two standard deviations from the median of the WHO child growth standards. Consequently, stunting does not merely indicate that a child is shorter relative to their age standard; it also serves as a critical indicator of chronic growth disruption.

The impacts of stunting are long-term, as it can adversely affect physical development, cognitive abilities, educational attainment, health, and productivity in adulthood. It is explained in [1]. that stunting is associated with long-term consequences, including low cognitive capacity, suboptimal educational performance, reduced income in adulthood, and lost productivity. Therefore, stunting is not only a nutritional and child health issue but is also intricately linked to the quality of human resources and the socioeconomic development of a region.

The determinants of stunting are multidimensional. It has been demonstrated in [6] that the determinants of stunting encompass diverse aspects, including sanitation, parenting practices, socioeconomic status, maternal and child health, and access to healthcare services. In this study, variables such as access to proper sanitation, low birth weight (LBW), population density, and poverty rates are used as factors hypothesized to be associated with variations in stunting prevalence across regencies/cities in Kalimantan.

III. METHODOLOGY

This study uses secondary data, with regencies/cities across Kalimantan as the units of observation, covering the provinces of West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, and North Kalimantan. The response variable used is stunting prevalence ($\boldsymbol{Y}$), while the predictor variables include access to improved sanitation ($\boldsymbol{X_1}$), low birth weight ($\boldsymbol{X_2}$), population density ($\boldsymbol{X_3}$), and the percentage of the poor population ($\boldsymbol{X_4}$). Stunting prevalence data were sourced from the Indonesian Nutritional Status Survey (SSGI) Report published by the Ministry of Health, whereas the data for the predictor variables were obtained from the Central Bureau of Statistics (BPS) publications, specifically the Province in Figures for each respective province in Kalimantan for the year 2025.

The method applied in this study is robust kernel nonparametric regression with MM-estimation. Kernel nonparametric regression is employed because it can model the relationship between stunting prevalence and its influencing factors without predefined functional forms. However, conventional kernel regression remains sensitive to outliers or extreme values. Therefore, the robust MM approach is integrated to mitigate the influence of outliers on the estimation results, yielding a more stable and representative model of the overall data pattern. The analytical steps in this study are structured as follows:

1) Defining the research variables: stunting prevalence as the response variable ($\boldsymbol{Y}$), and access to improved sanitation ($\boldsymbol{X_1}$), low birth weight ($\boldsymbol{X_2}$), population density ($\boldsymbol{X_3}$), and the poverty rate ($\boldsymbol{X_4}$) as the predictor variables.
2) Performing a descriptive analysis to outline the distribution of stunting prevalence and the characteristics of each predictor variable across regencies/cities in Kalimantan.
3) Constructing scatterplots between the response variable and each predictor variable to observe initial relationship patterns. This stage aims to determine whether the inter-variable relationships are linear, nonlinear, or lack a distinct pattern.
4) Conducting outlier diagnosis or influential observation checks using DFFITS values. An observation is categorized as an outlier if its absolute DFFITS value exceeds the threshold:

$$\mathbf{2}\sqrt{\frac{\boldsymbol{p}+\mathbf{1}}{\boldsymbol{n}}} = \mathbf{2}\sqrt{\frac{\mathbf{4}+\mathbf{1}}{\mathbf{56}}} = \mathbf{0.598}$$

where $\boldsymbol{p}$ represents the number of predictor variables.

5) Estimating the conventional kernel nonparametric regression model using the Nadaraya-Watson estimator. This stage involves selecting the kernel function, computing kernel weights, and choosing the optimal bandwidth via cross-validation.
6) Estimating both the conventional kernel nonparametric regression model and the robust MM kernel nonparametric regression model based on the optimal bandwidth.
7) Evaluating model performance using MSE, RMSE, MAE, and $\boldsymbol{R}^{\mathbf{2}}$.

IV. RESULTS AND DISCUSSIONS

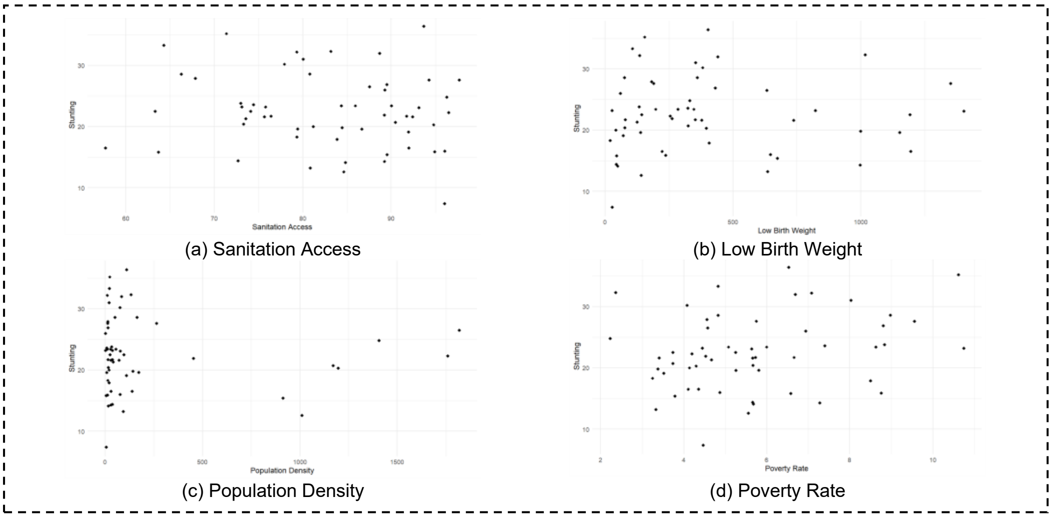
A. Descriptive Statistical Analysis

A descriptive statistical analysis was conducted to provide a general overview of the characteristics of the research data, as presented in Table 1. The summary statistics used include the number of observations (N), minimum (Min), median, maximum (Max), mean, and standard error (Std. Error).

Table 1 Descriptive Statistics of Research Variables

Variable	N	Min	Median	Max	Mean	Std. Error
Stunting (Y)	56	7.40	22.10	36.40	22.48	0.82
Access to improved sanitation (X_1)	56	57.71	84.41	97.70	82.83	1.34
Low birth weight (X_2)	56	20.00	304.50	1404.00	396.89	49.94
Population density (X_3)	56	2.00	37.00	1819.42	218.68	59.36
Poverty rate (X_4)	56	2.23	5.41	10.75	5.70	0.27

Based on Table 1, the number of observations across all variables is 56, indicating that the dataset encompasses 56 regencies/cities in the Kalimantan region. The stunting variable has a minimum value of 7.40 and a maximum value of 36.40, with a mean of 22.48. This suggests that the prevalence of stunting varies considerably across the regencies/cities in Kalimantan. For the access to improved sanitation variable, the mean value of 82.83 indicates that, in general, access to sanitation across the regencies/cities in the Kalimantan region is relatively high. However, the minimum value of 57.71 indicates that there are still areas with relatively low access to sanitation. The low birth weight variable exhibits a substantial range, spanning 20.00 to 1404.00, indicating considerable disparities in the number or incidence of low birth weight across regions. The population density variable exhibits very high variation, with a minimum of 2.00 and a maximum of 1819.42. This illustrates a sharp contrast in regional characteristics between sparsely populated areas and highly dense urban areas. Meanwhile, the poverty rate has a mean of 5.70, with a minimum of 2.23 and a maximum of 10.75. Subsequently, scatterplots were constructed to examine the initial relationships between stunting prevalence and each predictor variable, as shown in Figure 1.



(a) Sanitation Access

(b) Low Birth Weight

(c) Population Density

(d) Poverty Rate

Figure 1 Relationship Patterns between Stunting Prevalence and Research Variables

Based on Figure 1, all scatterplots between the response variable (stunting prevalence) and each predictor variable exhibit a dispersion of data points that does not form a distinct pattern. The relationships between stunting prevalence and access to improved sanitation, low birth weight, population density, and the poverty rate do not demonstrate a clear linear trend. Therefore, this study employs a kernel nonparametric regression approach, which offers greater flexibility in fitting the underlying data pattern.

B. Outlier Detection

An assessment of outliers or influential observations was conducted using DFFITS values to identify observations that could affect the stability of the model's estimated results. The presence of extreme observations must be carefully considered, as they can cause estimates to become unstable and fail to adequately represent the overall data pattern. An observation is categorized as an outlier or an influential observation if its absolute DFFITS value exceeds the predetermined threshold. The results of the outlier detection are presented in Table 2.

Table 2 Outlier Detection Results Based on DFFITS Values

Observation	Regency/City	DFFITS	Status
10	Melawi Regency	0.6087	Outlier
18	Barito Selatan Regency	0.6249	Outlier
31	Banjar Regency	0.7584	Outlier
32	Barito Kuala Regency	0.9215	Outlier
54	Tana Tidung Regency	0.7913	Outlier

Based on Table 2, five regencies are identified as outliers, namely Melawi Regency, South Barito Regency, Banjar Regency, Barito Kuala Regency, and Tana Tidung Regency. These five regions exhibit DFFITS values exceeding the outlier threshold; hence, their presence could influence the model's estimation results. This condition indicates that utilizing conventional kernel regression may produce unstable estimates if the influence of outliers is not controlled. Consequently, this study employs a robust MM kernel nonparametric regression approach to mitigate the impact of extreme observations without excluding them from the analysis.

C. Optimal Bandwidth Selection

Optimal bandwidth selection is conducted to determine the smoothness level of the kernel nonparametric regression curve. A bandwidth that is too small may cause the model to overfit the data, whereas an excessively large bandwidth may fail to capture local patterns adequately. The optimal bandwidth was selected using an exhaustive four-dimensional grid search. Ten candidate bandwidth values, ranging from 0.2 to 3.8 in increments of 0.4, were evaluated for each predictor. Therefore, 104 (10000) bandwidth combinations were assessed for each kernel weighting function, namely Epanechnikov, Gaussian, and Uniform. The optimal combination was determined based on the minimum CV value. Table 3 presents the three bandwidth combinations with the lowest CV values for each kernel weighting function.

Table 3 Optimal Bandwidth Selection Based on Kernel Weighting Functions

Kernel Weighting Function	h_1	h_2	h_3	h_4	CV
	0.6	0.2	0.2	0.6	0.3457
Epanechnikov	0.6	0.2	1.0	1.0	0.3462
	0.6	0.2	1.4	0.6	0.3473
	0.2	0.2	0.6	0.2	0.3492
Gaussian	0.2	0.2	2.6	0.2	0.3512
	0.6	0.2	2.1	0.2	3578
	0.6	**0.2**	**0.6**	**0.2**	**0.3453**
Uniform	0.6	0.2	0.2	1.4	0.3457
	1.0	0.2	0.2	0.6	0.3459

Based on Table 3, the minimum CV value of 0.3453 is obtained using the Uniform weighting function combined with the bandwidths $h_1 = 0.6$, $h_2 = 0.2$, $h_3 = 0.6$, and $h_4 = 0.2$. This indicates that this specific combination constitutes the optimal bandwidth for the kernel nonparametric regression model. Consequently, the Uniform weighting function is employed in the subsequent modeling stage because it yields the lowest validation error among the Epanechnikov and Gaussian weighting functions.

D. Robust Kernel Nonparametric Regression

The robust MM kernel nonparametric regression model in this study, in accordance with (1), can be expressed as follows.

$$Y_i = m(X_{1i}, X_{2i}, X_{3i}, X_{4i}) + \varepsilon_i, \qquad i = 1,2,\dots,56$$

where Y_i represents the stunting prevalence in the i-th regency/city, X_{1i} is the access to improved sanitation, X_{2i} is the low birth weight, X_{3i} is the population density, and X_{4i} is the poverty rate. The function $m(\cdot)$ is an unknown regression function estimated using the robust MM kernel approach. In accordance with (20), the robust MM kernel nonparametric regression estimator can be expressed as follows.

$$\hat{m}_{MM}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_H\,(\mathbf{x} - \mathbf{X}_i) w_i^{MM} Y_i}{\sum_{i=1}^{n} K_H\,(\mathbf{x} - \mathbf{X}_i) w_i^{MM}}$$

where

$$K_H(\mathbf{x} - \mathbf{X}_i) = \prod_{j=1}^{4} K\left(\frac{x_j - X_{ji}}{h_j}\right)$$

The best-performing kernel weighting function utilized is the Uniform Kernel with an optimal bandwidth combination of $h_1 = 0.6, h_2 = 0.2, h_3 = 0.6$, and $h_4 = 0.2$, enabling the estimated model for stunting prevalence to be expressed as follows.

$$\hat{Y}_i = \frac{\sum_{k=1}^{56} K_H\,(\mathbf{X}_i - \mathbf{X}_k) w_k^{MM} Y_k}{\sum_{k=1}^{56} K_H\,(\mathbf{X}_i - \mathbf{X}_k) w_k^{MM}}$$

where w_k^{MM} represents the robust MM weights assigned based on the magnitude of the residuals. Observations with large residuals will receive smaller weights, whereas observations that follow the general pattern of the data will receive larger weights. Consequently, the resulting model remains capable of capturing local data patterns via the kernel weights while demonstrating enhanced resilience to outliers via the robust MM weights. The estimation results of the robust MM kernel nonparametric regression model are displayed in Figure 2.

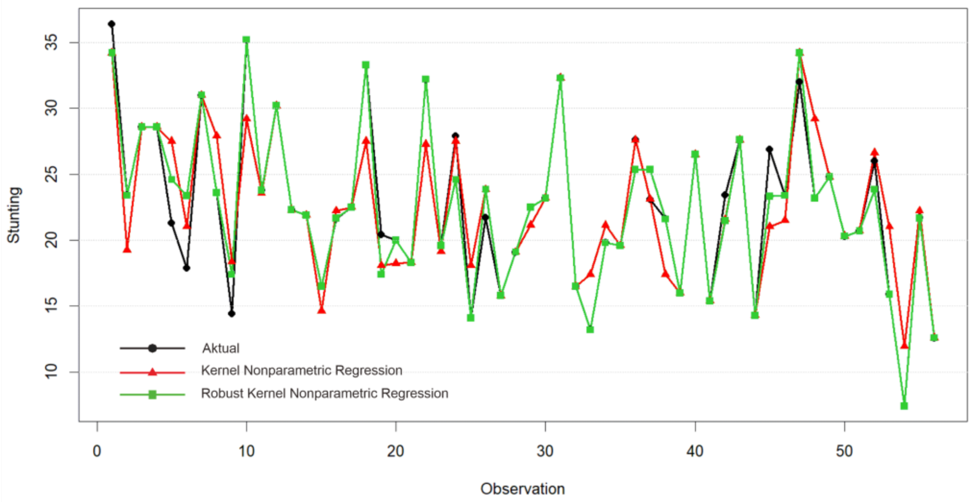


Figure 2 Comparison of Kernel and Robust Kernel Regression Estimates

Based on Figure 2, the estimation results for both the nonparametric kernel regression and the robust MM kernel regression generally track the fluctuating patterns of the actual stunting prevalence across the observed regencies/cities in Kalimantan. The estimation curve of the robust MM kernel model lies relatively close to the actual values, indicating that the model flexibly captures variation in the stunting data. Furthermore, integrating the robust MM approach enables the model to mitigate the influence of extreme observations previously identified by DFFITS values. Consequently, the robust MM kernel regression provides more stable estimates for depicting stunting prevalence patterns than conventional nonparametric kernel regression, particularly when dealing with data containing outliers or influential observations.

E. Model Evaluation

Model evaluation was conducted to assess the model's ability to estimate stunting prevalence in Kalimantan. The evaluation metrics employed include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Lower MSE, RMSE, and MAE values signify a lower estimation error rate, whereas a higher R^2 value indicates a superior capability of the model in explaining the variance in stunting prevalence.

Table 4 Model Evaluation of Kernel and Robust Kernel

Goodness of Fit	Uniform Kernel	Robust Uniform Kernel
MSE	7.1515	2.0556
RMSE	2.6742	1.4337
MAE	1.6816	0.6581
R^2	0.8061	0.9510

Based on Table 4, the Robust Uniform Kernel model outperforms the Uniform Kernel model. This is evident from its lower MSE, RMSE, and MAE values. The Uniform Kernel model yields an MSE of 7.1515, an RMSE of 2.6742, and an MAE of 1.6816, whereas the Robust Uniform Kernel model produces an MSE of 2.0556, an RMSE of 1.4337, and an MAE of 0.6581. Furthermore, the R^2 value increases from 0.8061 for the Uniform Kernel to 0.9510 for the Robust Uniform Kernel. These results indicate that incorporating the robust approach improves the in-sample fit of stunting prevalence estimates and yields a superior model for explaining variance in the data, particularly in the presence of outliers or influential observations. The model's predictive performance was not evaluated on an independent test sample due to the limited number of observations. Therefore, the reported performance measures represent in-sample model fit and may be optimistic. Future studies should assess the model's generalizability using larger datasets and out-of-sample validation procedures.

V. CONCLUSIONS AND SUGGESTIONS

This study demonstrates that robust MM kernel nonparametric regression can be effectively utilized to model stunting prevalence across the Kalimantan region. The bandwidth selection process reveals that the Uniform Kernel weighting function yields the minimum CV value, leading to its adoption in the modeling stage. Based on the model evaluation, the Robust Uniform Kernel outperforms the conventional Uniform Kernel, achieving an MSE of 2.0556, an RMSE of 1.4337, an MAE of 0.6581, and an $\boldsymbol{R^2}$ of 0.9510. The lower estimation errors alongside the higher $\boldsymbol{R^2}$ value confirms that integrating the robust MM method successfully improves model accuracy in estimating stunting prevalence. Consequently, robust MM kernel nonparametric regression provides a flexible and stable alternative for modeling stunting, especially when dealing with nonlinear relationship patterns and data containing outliers.

Future research is recommended to incorporate additional relevant predictor variables, such as maternal education, exclusive breastfeeding coverage, access to healthcare services, immunization, food security, and environmental factors. Furthermore, model development could be extended by incorporating spatial aspects, given that regency/city-level stunting data potentially exhibit interregional dependencies. Comparisons with alternative methods, such as robust splines, robust local polynomials, or machine learning approaches, could also be explored to develop a more comprehensive model for explaining the variance in stunting in Kalimantan.

ACKNOWLEDGEMET

The authors express their gratitude to the Ministry of Health of the Republic of Indonesia and the Central Bureau of Statistics (BPS) for providing the secondary data utilized in this study. Appreciation is also extended to all parties who provided support and valuable feedback throughout the drafting of this article.

REFERENCES

[1] A. Soliman *et al.*, “Early and Long-term Consequences of Nutritional Stunting: From Childhood to Adulthood,” *Acta Bio Medica Atenei Parm.*, vol. 92, no. 4, hal. 1–12, 2021, doi: 10.23750/abm.v92i1.11346.

[2] R. R. Akbar, W. Kartika, dan M. Khairunnisa, “The Effect of Stunting on Child Growth and Development,” *Sci. J.*, vol. 2, no. 4, hal. 153–160, 2023.

[3] E. Lestari, A. Siregar, A. K. Hidayat, dan A. A. Yusuf, “Stunting and its association with education and cognitive outcomes in adulthood : A longitudinal study in Indonesia,” *PLoS One*, vol. 19, no. 5, hal. 1–18, 2024, doi: 10.1371/journal.pone.0295380.

[4] C. Santos, O. Vasconcelos, S. Pereira, dan R. Garganta, “Stunting and Physical Fitness. The Peruvian Health and Optimist Growth Study,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 10, hal. 1–15, 2020.

[5] S. Wijeakumar *et al.*, “Stunting in infancy is associated with atypical activation of working memory and attention networks,” *Nat. Hum. Behav.*, vol. 7, no. December, hal. 2199–2211, 2023, doi: 10.1038/s41562-023-01725-3.

[6] T. Vaivada, N. Akseer, S. Akseer, A. Somaskandan, M. Stefopulos, dan Z. A. Bhutta, “Stunting in childhood: an overview of global burden, trends, determinants, and drivers of decline,” *Am. J. Clin. Nutr.*, vol. 112, hal. 777S-791S, 2020, doi: 10.1093/ajcn/nqaa159.

[7] A. N. Sartika, M. Khoirunnisa, E. M. Id, dan E. Ermayani, “Prenatal and postnatal determinants of stunting at age 0-11 months: A cross-sectional study in Indonesia,” *PLoS One*, vol. 16, no. 7, hal. 1–14, 2021, doi: 10.1371/journal.pone.0254662.

[8] J. R. M. Silva *et al.*, “Water, sanitation, and hygiene vulnerability in child stunting in developing countries: a systematic review with meta-analysis,” *Public Health*, vol. 219, hal. 117–123, 2023, doi: 10.1016/j.puhe.2023.03.024.

[9] A. Kustanto, “The Prevalence Of Stunting, Poverty, And Economic Growth in Indonesia: A Dynamic Panel Data Causality Analysis,” *JDE (Journal Dev. Econ.*, vol. 6, no. 2, hal. 150–173, 2021, doi: 10.20473/jde.v6i2.22358.

[10] A. D. Laksono, R. D. Wulandari, N. Amaliah, dan R. W. Wisnuwardani, “Stunting among children under two years in Indonesia: Does maternal education matter?,” *PLoS One*, hal. 1–11, 2022, doi: 10.1371/journal.pone.0271509.

[11] Ramadhaniyati, L. Lestari, R. Saito, dan A. Tsuda, “Exploring Mothers’ Experiences in Accessing Primary,” *J. Rural Community Nurs. Pract.*, vol. 3, no. 1, hal. 129–149, 2025.

[12] I. S. Anzarkusuma, S. Fikawati, dan A. T. Rahmi, “Determinan Stunting pada Anak Usia 24-59 Bulan di Provinsi Nusa Tenggara Barat: Analisis Data Survei Status Gizi Indonesia (SSGI) Tahun 2022,” *Gizi Indones.*, vol. 48, no. 1, hal. 55–68, 2025, doi: 10.36457/gizindo.v48i1.1094.

[13] R. Kemenkes, *Survei Status Gizi Indonesia (SSGI) 2024 dalam Angka*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2025. [Daring]. Tersedia pada: https://www.badankebijakan.kemkes.go.id/survei-status-gizi-indonesia-ssgi-2024/

[14] A. Adityaningrum, N. Arsad, dan H. Jusuf, “Faktor Penyebab Stunting di Indonesia: Analisis Data Sekunder Data SSGI Tahun 2021,” *Jambura J. Epidemiol.*, vol. 3, no. 1, hal. 1–10, 2023.

[15] S. Ningsih, M. R. Madonsa, S. L. Mahmud, I. Djakaria, dan S. K. Nasib, “Implementasi Regresi Logistik Biner Stratifikasi Pada Pemodelan Stunting Untuk Anak Balita Di Kabupaten Gorontalo,” *Jambura J. Probab. Stat.*, vol. 5, no. 1, hal. 19–23, 2024.

[16] A. F. Dewi dan H. Husain, “Analisis Regresi Data Panel dalam Memodelkan Faktor yang Mempengaruhi Stunting di Indonesia,” *J. Math. Comput. Stat.*, vol. 7, no. 2, hal. 313–322, 2024.

[17] S. Arifin, A. Islamiyati, dan R. Raupong, "Kemampuan Estimator Spline Linear dalam Analisis Komponen Utama," *ESTIMASI J. Stat. Its Appl.*, vol. 1, no. 1, hal. 40, 2020, doi: 10.20956/ejsa.v1i1.9262.

[18] S. Arifin, "Pemodelan Regresi Logistik Ordinal dengan Estimator Spline Truncated Ordinal Logistic Regression Modeling with Truncated Spline Estimator," 2023, *Universitas Hasanuddin*.

[19] Z. Pei *et al.*, "Local Polynomial Order in Regression Discontinuity Designs," *J. Bus. Econ. Stat.*, vol. 0, no. 0, hal. 1–18, 2022, doi: 10.1080/07350015.2021.1920961.

[20] S. Arifin, D. Anggraini, N. Salam, S. Arifin, D. Anggraini, dan N. Salam, "A Comparative Study of Linear and Quadratic Spline Regression Models for Predicting HbA1c Levels in Patients with Diabetes Mellitus," *Jambura J. Math.*, vol. 7, no. 2, hal. 183–188, 2025, doi: 10.37905/jjom.v7i2.33292.

[21] B. Lestari, N. Chamidah, dan T. Saifudin, "Estimasi Fungsi Regresi Dalam Model Regresi Nonparametrik Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel," *JMSK J. Mat. Stat. dan Komputasi*, no. 2, hal. 20–24, 2019.

[22] T. H. Ali, "Modification of the adaptive Nadaraya-Watson kernel method for nonparametric regression," *Commun. Stat. - Simul. Comput.*, vol. 51, no. 2, hal. 391–403, Feb 2022, doi: 10.1080/03610918.2019.1652319.

[23] C. R. Oktarina, S. Nugroho, I. Sriliana, dan P. Novianti, "Estimation of Stunting and Wasting in Sumatra 2022 with Nadaraya-Watson Kernel and Penalized Spline," *Inferensi*, vol. 8, no. 3, hal. 197–208, 2025.

[24] A. W. Omer dan T. H. Ali, "Dealing with the Outlier Problem in Multivariate Linear Regression Analysis Using the Hampel Filter," *Kurdistan J. Appl. Res.*, vol. 10, no. 1, hal. 1–15, 2025, doi: 10.24017/science.2025.1.

[25] A. Fitrianto dan S. H. Xin, "Comparisons Between Robust Regression Approaches in The Presence Of Outliers and High Leverage Points," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 1, hal. 243–252, 2022.

[26] A. Prahutama, A. Rusgiyono, D. Ispriyanti, dan T. W. Utami, "Pemodelan Regresi Robust M-Estimator dalam Menangani Pencilan (Studi Kasus Pemodelan Jumlah Kematian Ibu Nifas di Jawa Tengah," *J. Stat. Univ. Muhammadiyah Semarang*, vol. 9, no. 1, hal. 35–39, 2021.

[27] K. K. Gasul, A. Atti, dan M. A. Kleden, "Model Regresi Robust dengan Metode Estimasi M , Estimasi S dan Estimasi MM untuk Produksi Beras di Nusa Tenggara Timur," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, hal. 20–32, 2023, doi: 10.20956/ejsa.vi.24879.

[28] I. D. Nugrahani, Y. Susanti, dan N. Qona, "Modeling of Rice Production in Indonesia Using Robust Regression with The Method of Moments (MM) Estimation," *Basic Appl. Sci. Conf.*, vol. 2021, hal. 79–87, 2021.

[29] I. F. Yuliati dan P. R. Sihombing, "Pemodelan Fertilitas di Indonesia Tahun 201 7 Menggunakan Pendekatan Regresi Nonparametrik Kernel dan Spline," *J. Stat. dan Apl.*, vol. 4, no. 1, hal. 48–60, 2020.

[30] U. N. Safitri, I. Sriliana, R. Adelisa, M. Hafiz, dan P. Novianti, "Perbandingan regresi nonparametrik spline truncated dan kernel gaussian dalam menganalisis faktor-faktor penentu indeks pembangunan manusia (ipm) di indonesia," *J. Gaussian*, vol. 14, no. 2, hal. 554–564, 2025, doi: 10.14710/j.gauss.14.2.554-564.

[31] S. A. Putri, A. R. Aristya, N. A. Janad, dan Y. N. Tadale, "Estimator Nadaraya-Watson dengan Fungsi Kernel Normal dan Fungsi Kernel Kuadratik," *J. Syst. Comput. Eng.*, vol. 3, no. 1, hal. 206–219, 2022.

[32] N. Y. Adrianingsih, A. T. R. Dani, I. N. Budiantara, D. L. Ull, dan R. A. Kosasih, "Computatioal Analysis of Kernel-Based Nonparametric Regression Applied to Poverty Data," *Mandalika Math. Educ. J.*, vol. 7, no. September, hal. 1336–1347, 2025.

[33] Y. Farida, I. Purwanti, dan N. Ulinnuha, "Comparing Gaussian and Epanechnikov Kernel of Nonparametric Regression in Forecasting Indonesia Sharia Stock Index (ISSI)," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 1, hal. 323–332, 2022.

[34] A. M. Djalumang, R. Raupong, dan S. Siswanto, "Mixed Geographically Weighted Regression Modeling Using the MM-Estimator Method on Data of Poverty," *J. Varian*, vol. 8, no. 3, hal. 293–306, 2025.

[35] F. M. A. Soliman, A. M. Mohamed, dan M. R. Abonazel, "New Robust Estimators for the Nonparametric Regression Model: Application and Simulation Study," *Int. J. Anal. Appl.*, vol. 23, no. 1, hal. 1–18, 2025.