International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 3 March - 2026, Pages: 9-16

Modeling Toddler Weight Growth in Banyuwangi Based on Z-Score Curves Using Local Linear Regression for Health Assessment

Nur Chamidah*, Ardi Kurniawan, Toha Saifudin, Naufal Ramadhan Al Akhwal Siregar, Vanisia Suci Ananda, Ratna Anugrahaningtyas, Amelia Fatihah, Nararya Raissa Tiara Putri

Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia
*Corresponding author: nur-c@fst.unair.ac.id

Abstract: *Children represent an important generation that will influence the direction of national development; therefore, their growth must be monitored carefully. Stunting remains a significant public health concern in Indonesia, including Banyuwangi Regency. Although the prevalence decreased from 8.64% in 2021 to 2.44% in 2024, more than 2,200 toddlers are still projected to experience stunting by 2025. This study models toddler weight growth in Banyuwangi Regency using local linear nonparametric regression based on Z-scores. The analysis is based on primary data from 1,627 toddlers aged 0–60 months across nine villages. The estimation is performed using the Gaussian Kernel (GaussK) method with bandwidth selected through Cross-Validation (CV). The results show low prediction errors across all Z-score levels. For male toddlers, MAPE values range from 1.28% to 3.71%, with* R^2 *between 0.983 and 0.996. For female toddlers, MAPE ranges from 1.51% to 2.50%, with* R^2 *between 0.976 and 0.995. Based on the MAPE criterion (≤ 10%), the estimates fall within the highly accurate prediction category. The smallest errors are observed around the median Z-score (0 SD), while relatively higher variability appears at extreme Z-score levels. These findings indicate that the Gaussian Kernel local linear regression with CV bandwidth selection provides consistent and reliable estimates of toddler weight growth. The results may serve as a statistical reference for evaluating growth patterns and supporting evidence-based nutritional interventions in Banyuwangi Regency.*

Keywords— Local linear regression; Gaussian kernel; Z-score; Toddler growth; Health Assessment.

1. INTRODUCTION

Children have a strategic role in determining the direction of future development as the golden generation that will succeed the nation. The population of children in Indonesia aged 0-14 years in 2025 will reach around 66.7 million, or nearly a quarter of Indonesia's total population of 284.44 million [1]. According to CIA data, Indonesia has an estimated birth rate of 1.96 in East and Southeast Asia, ranking 8th [2]. This means that children have the right to grow, develop, and have their needs met. In accordance with the 1945 Constitution, this must be a top priority in the national development agenda.

To achieve success in realizing key priorities, there are certainly challenges related to children's physical growth. One of these challenges is stunting, which occurs when a child's weight is below average for their age due to long-term malnutrition. This challenge reflects a serious obstacle in efforts to achieve sustainable nutrition, which will have an impact on children's health, development, and quality of life in the future. In Banyuwangi, although the prevalence of stunting has decreased significantly from 8.64% in 2021 to around 2.44% in 2024, there are still 2,269 children who will experience stunting until 2025 [3]. Stunting has multidimensional impacts, namely hindering physical growth, affecting brain development, work ability, and the quality of life of children when they reach adulthood [4].

The issue of stunting in Banyuwangi has attracted serious attention from various parties, including the East Java National Committee for Stunting Prevention (KNPS). As a form of commitment to addressing this issue, the local government has allocated a budget of Rp10 billion to support various nutritional intervention and stunting prevention programs [5]. This also supports the global commitment to the Sustainable Development Goals (SDGs), particularly the second SDG, which is to end hunger and ensure access to nutritious food, and the third goal, which is to ensure healthy lives and promote well-being for all ages.

The importance of studying weight is to understand children's growth patterns during their development. Nonparametric approaches, particularly splines, local linear regression, and kernels, provide flexibility in modeling without having to rely on linearity assumptions. These methods are capable of capturing complex growth variations while producing a more accurate picture of the factors that influence toddler height. Previous studies have demonstrated the effectiveness of similar methods in describing toddler weight growth patterns, such as research in Buleleng using spline regression [6]. Other research in Jember applied local linear regression to analyze factors that influence children's nutritional status[7].

There are other studies that confirm the relevance of nonparametric methods in the study of stunting and toddler growth. A study in Banyuwangi used local linear regression to model toddler height and found differences with the WHO-2007 standard [8]. Another study also applied nonparametric spline regression to cases of stunting in Indonesia to capture complex data patterns [9]. There is also a study that designed a Z-score standard curve based on height-for-age using a local linear estimator, which produced a more representative growth chart for determining stunting status [10]. Meanwhile, in Probolinggo, there was a study using parametric and nonparametric ordinal logistic regression to identify the determinants of stunting [11]. These four studies reinforce the evidence that nonparametric methods have high flexibility in describing growth variations and stunting determinants, making them an important basis for this study.

Thus, this study aims to present a comprehensive analysis of the weight gain patterns of toddlers in Banyuwangi. The results obtained are expected to serve as a reference for local governments, non-governmental organizations, and other stakeholders in designing more targeted nutrition and health interventions. In addition, this study is expected to support efforts to reduce stunting rates and contribute to the realization of a healthy, productive, and competitive generation as the foundation for national development.

2. RESEARCH METHODS

2.1 Data Source

The data used in this study is a primary data of toddlers and children obtained from Banyuwangi Regency, Indonesia in 2025. Each toddlers and children aged 0-60 months old is observed for their gender, age and weight. The samples were selected using a non-probability purposive sampling technique. The study obtained cross-sectional data with a total of 1627 observations, consisting of 809 observations for male toddlers and 818 observations for female toddlers.

2.2 Analysis Method

The overall analysis procedure is outlined as follows.

1. The weight variable is examined as the dependent variable, while age is treated as the independent variable.
2. The z-score (ZSD) was calculated under the assumption that the observed data follow a normal distribution. This study employed the LMS method, which incorporates three age-specific parameters, namely the Box–Cox power (L), the median (M), and the generalized coefficient of variation (S) [12]. The z-score at age t for the i-th standard deviation is defined as.

$$ZSD_i(t) = M(t)[1 + L(t) \times i]^{1/L(t)} \qquad (1)$$

where $i = -3, -2, -1, 0, 1, 2, 3$ and $t = 0, 1, \dots, 60$

The estimation of the parameters L, M, and Sfollowed the procedure described by Cole [13]. Initially, the geometric mean and geometric coefficient of variation were derived by calculating the mean and standard deviation of the natural logarithm of the observed data. The geometric mean $M_g(t)$and geometric coefficient of variation $S_g(t)$were computed as.

$$M_g(t) = \exp\left(\frac{\sum_{i=1}^{n_t} \ln x_i(t)}{n_t}\right) \qquad (2)$$

$$S_g(t) = \sqrt{\frac{\sum_{i=1}^{n_t}\left(\ln x_i(t) - \frac{\sum_{i=1}^{n_t} \ln x_i(t)}{n_t}\right)^2}{n_t - 1}} \qquad (3)$$

Subsequently, the arithmetic mean and arithmetic coefficient of variation were calculated from the original data. The arithmetic mean $M_a(t)$ and arithmetic coefficient of variation $S_a(t)$ are expressed as.

$$M_a(t) = \frac{\sum_{i=1}^{n_t} x_i(t)}{n_t} \qquad (4)$$

$$S_a(t) = \sqrt{\frac{\sum_{i=1}^{n_t}\left(x_i(t) - \frac{\sum_{i=1}^{n_t} \ln x_i(t)}{n_t}\right)^2}{n_t - 1} \times \frac{1}{M_g(t)}} \qquad (5)$$

In addition, the harmonic mean and harmonic coefficient of variation were obtained by computing the mean and standard deviation of the inverse of the data. The harmonic mean $M_h(t)$ and harmonic coefficient of variation $S_h(t)$ are defined as.

$$M_h(t) = \frac{n_t}{\sum_{i=1}^{n_t} \frac{1}{x_i(t)}} \qquad (6)$$

$$S_h(t) = \sqrt{\frac{\sum_{i=1}^{n_t}\left(\frac{1}{x_i(t)} - \frac{\sum_{i=1}^{n_t} \ln x_i(t)}{n_t}\right)^2}{n_t - 1} \times M_h(t)} \qquad (7)$$

The arithmetic, geometric, and harmonic coefficients of variation were then used to compute the intermediate parameters $A(t)$ and $B(t)$ according to.

$$A(t) = log\left(\frac{S_a(t)}{S_h(t)}\right) \qquad (8)$$

$$B(t) = \log\left(\frac{S_a(t)S_h(t)}{S_g(t)^2}\right) \qquad (9)$$

The Box–Cox power parameter $L(t)$ was estimated as:

$$L(t) = \frac{-A(t)}{2B(t)} \qquad (10)$$

Based on the estimated value of $L(t)$, the generalized coefficient of variation $S(t)$was calculated using.

$$S(t) = S_g(t)\exp\left(\frac{A(t)L(t)}{4}\right) \qquad (11)$$

Finally, the generalized median $M(t)$was obtained by combining the arithmetic, geometric, and harmonic means as follows.

$$M(t) = M_g(t) + \frac{\big(M_a(t) - M_h(t)\big)L(t)}{2} + \frac{(M_a(t) - 2M_g(t) + M_h(t))L(t)^2}{2} \qquad (12)$$

This LMS-based formulation allows for flexible modeling of skewed growth data and enables the transformation of raw measurements into standardized z-scores across age groups.

3. The association between age and weight is estimated and modeled for the entire dataset using a nonparametric regression framework.

Nonparametric regression refers to a set of statistical estimation approaches that do not assume a specific parametric form for the regression function *f(x)* and instead derive the relationship between predictor and response variables directly from the data, providing flexibility when the functional form is unknown [14].

Nonparametric regression identifies the functional form of the relationship pattern by using information derived from the observed data, where the regression function is assumed to be smooth within a certain functional space, resulting in a highly flexible model that can adapt to the data without assuming a predetermined form [15]. Nonparametric regression can be defined as follows:

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \dots, n \qquad (13)$$

where y_i denotes the response variable, x_i represents the predictor variable, $f(x_i)$ is an unknown regression function, and ε_i is the error term associated with the i-th observation, which is assumed to follow a normal distribution with mean zero and variance σ_ε^2, that is $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

The steps involved in the local linear estimation procedure are described as follows:

a. Verify that the Kernsmooth and locpol packages are properly installed and loaded in the RStudio environment.

b. Load the dataset into RStudio using the readxl package, adjusted to the specific format of the data file.

c. Determine the optimal bandwidth parameter (h) for the full dataset, as well as for in-sample and out-of-sample data, by applying the Plug-in selection method with a Gaussian kernel. In general, a kernel function $K$with bandwidth $h$can be expressed as follows.

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), -\infty < x < \infty, h > \qquad (14)$$

In kernel-based nonparametric regression, the variable $x$denotes the specific point at which the kernel weighting function is evaluated, determining the influence of each observation on the estimated regression curve according to its proximity to x[16]. The Gaussian kernel, a common choice due to its smooth and infinitely differentiable weighting behavior, assigns weights that decay exponentially with the squared distance between $x$and the observed data points, thus yielding a flexible local approximation of the underlying relationship. Mathematically, the Gaussian kernel function used in this context is given by

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}x^2\right) \qquad (15)$$

where this form reflects the symmetric, continuous weighting around the target point xand supports nonparametric estimation without assuming a predefined functional form for the regression relationship[16]

d. The regression model is constructed using a local linear estimator with an optimal bandwidth h. In local linear regression, the unknown regression function $g(x)$in the nonparametric regression model is approximated locally around a target point $x_0$using a first-order Taylor expansion, expressed as.

$$g(x) = \sum_{j=0}^{1} \frac{(x - x_0)^j}{j!} g^{(j)}(x_0) \qquad (16)$$

$$= \sum_{j=0}^{1} (x - x_0)^j \beta_j(x_0)$$

$$x \,\epsilon\, (x_0 - h)(x_0 + h)$$

Equation (4) represents a local linear approximation of the regression function, where $\beta_0(x_0)$and $\beta_1(x_0)$ denote the local intercept and slope at point x_0, respectively. The local linear model in Equation (4) can be written in matrix form as.

$$g(x) = \boldsymbol{x}(x_0)\boldsymbol{\beta}(x_0) \qquad (17)$$

Where $\boldsymbol{x}(x_0) = [1 \; x - x_0], x \,\epsilon\, \epsilon\, (x_0 - h)(x_0 + h)$ and $\boldsymbol{\beta}(x_0) = \begin{pmatrix} \beta_0(x_0) \\ \beta_1(x_0) \end{pmatrix}$. The regression function $g(x)$ in Equation (5) is estimated using a local linear estimator, yielding.

$$\hat{g}(x) = \boldsymbol{x}(x_0)\hat{\boldsymbol{\beta}}(x_0) \qquad (18)$$

Based on Equation (6), the local linear nonparametric regression model with one predictor variable and one response variable can be expressed in matrix notation by the following equation.

The estimation of the parameter vector $\boldsymbol{\beta}(x_0)$ in Equation (7) leads to the following system of equations as follow.

$$\boldsymbol{y_1} = \beta_0 x_0 + \beta_1 x_0 (x_1 - x_0) + \varepsilon_1 \qquad (19)$$

$$\boldsymbol{y_2} = \beta_0 x_0 + \beta_1 x_0 (x_2 - x_0) + \varepsilon_2$$

$$\boldsymbol{y_n} = \beta_0 x_0 + \beta_1 x_0 (x_n - x_0) + \varepsilon_n$$

Equation (8) can be compactly written in mathematical notation as follows.

$$\boldsymbol{y} = \boldsymbol{X}(x_0)\boldsymbol{\beta}(x_0) + \boldsymbol{\varepsilon} \qquad (20)$$

Where $\boldsymbol{X}(x_0) = \begin{pmatrix} 1 & x_1 - x_0 \\ 2 & x_2 - x_0 \\ & \vdots \\ n & x_n - x_0 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, and $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

The estimate of $\boldsymbol{\beta}(x_0)$ is obtained from n paired data samples with a weight function is $\boldsymbol{K}_h(x_i - x_0)$.

The weight function, referred to as the kernel function, assigns weights to observations according to their proximity to the target point, with the weight magnitude controlled by the optimal bandwidth h. Parameter estimation is carried out using the Weighted Least Squares (WLS) approach through the minimization of the following criterion function.

$$Q(x_0) = [\boldsymbol{y} - \boldsymbol{X}(x_0)\boldsymbol{\beta}(x_0)]^T \boldsymbol{K}_h(x_0) [\boldsymbol{y} - \boldsymbol{X}(x_0)\boldsymbol{\beta}(x_0)] \tag{21}$$

where $\mathbf{K}_h(x_0)$denotes a diagonal weighting, matrix defined as follows.

$$\boldsymbol{K}_h(x_0) = \begin{bmatrix} \boldsymbol{K}_h(x_1 - x_0) & 0 & \cdots & 0 \\ 0 & \boldsymbol{K}_h(x_2 - x_0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{K}_h(x_n - x_0) \end{bmatrix}$$

The estimated coefficient vector $\boldsymbol{\beta}(x_0)$ at a target point $x_0$is obtained by minimizing the localized weighted least squares criterion, leading to.

$$\widehat{\boldsymbol{\beta}}(x_0) = [\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{X}(x_0)]^{-1}\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{y}, \tag{22}$$

where $\mathbf{K}_h(x_0)$is the diagonal matrix of kernel weights centered at $x_0$and $\mathbf{X}(x_0)$contains the local design matrix at x_0. Based on this, the local linear estimator for the regression function $g(x)$at $x_0$can be expressed as.

$$\hat{g}(x_0) = \boldsymbol{x}(x_0)[\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{X}(x_0)]^{-1}\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{y}, \tag{23}$$

which directly estimates the smooth regression surface by fitting locally weighted linear approximations around each evaluation point. Furthermore, using the unit vector $e_1 = (1 \quad 0)$ isolates the intercept component to yield the point estimate of the nonparametric regression function at x_0 such as.

$$\hat{g}(x_0) = (\boldsymbol{e}_1)[\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{X}(x_0)]^{-1}\mathbf{X}^\top(x_0)\mathbf{K}_h(x_0)\mathbf{y}, \tag{24}$$

This local linear smoothing framework is widely adopted due to its favorable boundary behavior and bias properties, and optimal selection of the smoothing parameter (bandwidth) plays an important role in balancing bias and variance in the estimator [17].

The choice of the bandwidth parameter is crucial for achieving an optimal local linear estimator, as it directly influences the degree of smoothness in nonparametric regression: a very small bandwidth tends to produce an estimator that overfits the data with excessive variability, while a very large bandwidth oversmooths the function, potentially obscuring important structure in the relationship between variables. Therefore, an optimal bandwidth balances the tradeoff between bias and variance to minimize estimation error and improve performance of the regression estimator [18].

The selection of an appropriate smoothing parameter (bandwidth) is fundamental in nonparametric regression because it governs how smooth or wiggly the estimated regression curve will be. Specifically, the bandwidth controls the trade-off between variance and bias: a very small bandwidth tends to overfit the data by producing a highly jagged function, whereas a very large bandwidth oversmooths important local features of the data. A frequently used data-driven strategy for identifying the optimal bandwidth is cross-validation (CV), which evaluates the performance of candidate bandwidth values by minimizing an estimate of the prediction error derived from the data itself, thus providing an objective criterion for balancing smoothness and fit [19].

In cross-validation for kernel regression, suppose that the observed dataset is given by $D_n = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. For a given bandwidth h, the leave-one-out cross-validation (LOOCV) approach evaluates the predictive performance of the estimator by sequentially excluding each observation from the fitting process. The optimal bandwidth is selected by minimizing the cross-validation score function, with $\hat{g}_{-i,h}(x_i)$ denoting the regression estimate obtained by omitting the i-th observation. The CV criterion is defined as follows.

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}_{-i,h}(x_i)\right)^2. \tag{25}$$

The smoothing parameter, denoted as h_{opt}, is determined by identifying the bandwidth value that minimizes the cross-validation (CV) criterion, as defined by the following expression.

$$h_{\text{opt}} = \arg\min_h \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \hat{g}_{n,i}^{(h)}(x_i)\right]^2 \tag{26}$$

with

$$\hat{g}_{n,i}{}^{(h)}(x_i) = x_i(x_o)[X^{(-i)}(x_0)\mathrm{K}_h{}^{(-i)}(x_0)X^{(-i)}(x_0)]^{-1} X^{(-i)T}(x_0)\,\mathrm{K}_h{}^{(-i)}(x_0)y^{(-i)}, \text{ and } X^{(-i)}(x_0) \tag{27}$$

where $\boldsymbol{X}^{(-i)}(x_0)$ and $\mathbf{K}_h^{(-i)}(x_0)$ denote the design matrix and kernel weight matrix with the i-th observation removed, and $\boldsymbol{y}^{(-i)}$ represents the response vector excluding the i-th observation. Based on this formulation, the optimal bandwidth is defined as the value of $h$that yields the minimum CV score, ensuring an appropriate balance between bias and variance in the local linear estimator [20].

After computing the estimated regression function, it is important to perform a thorough evaluation to assess how well the model fits the data and to verify its validity for interpretation and decision-making. This evaluation step ensures that the estimated function appropriately captures the underlying relationship among the variables and produces reliable predictive results. In this study, the goodness-of-fit is assessed using the coefficient of determination (R^2) and the mean squared error (MSE) as evaluation metrics, which are commonly applied in regression contexts to quantify the explanatory power and the average squared deviation of predicted values from the observed values, respectively [21].

The coefficient of determination (R^2) measures the proportion of the total variance in the response variable that is accounted for by the regression model. A higher R^2 indicates that the model explains a larger share of the variability in the data, which generally signifies a better fit. Formally, R^2 is expressed as.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \qquad (28)$$

where y_i is the observed value for the i-th observation, $\hat{y}_i$ is the predicted value from the model, and $\bar{y}$ is the average of all observed values. This ratio reflects the proportion of explained variance relative to the total variance in the dataset [22].

Mean Squared Error (MSE), which quantifies the average of the squared differences between observed values and those predicted by a model. MSE serves as an estimate of the residual variance and is widely used as a performance metric in regression analysis: the smaller the MSE, the closer the model's predictions are to the actual observed values, indicating better predictive performance. To identify the best model, the smoothing parameter can be selected by minimizing the MSE, where the model with the lowest MSE value is considered the most accurate approximation to the real data [23]. The MSE is formally computed as.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (29)$$

where n denotes the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value from the model. Because MSE reflects the average squared residuals, its value ranges from 0 to ∞, with lower values indicating better model fitting and smaller prediction error [24].

Mean Absolute Percentage Error (MAPE) is a measure used to evaluate the accuracy level of a forecasting model by comparing the absolute difference between the actual values and the predicted values in percentage form. The smaller the MAPE value, the higher the accuracy level of the forecasting model [25]. The MAPE formula is given as follows.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \times 100\% \qquad (30)$$

where
A_i : Actual data at the i-th observation
F_i : Predicted value at the i-th observation
n : Number of data points

The resulting MAPE value can be interpreted as follows.

Tabel 1. Kriteria Nilai MAPE (%)

MAPE (%)	**Interpretation**
$\text{MAPE} \leq 10\%$	Highly accurate prediction
$10 < \text{MAPE} \leq 20\%$	Good prediction
$20 < \text{MAPE} \leq 50\%$	Reasonable prediction
$\text{MAPE} > 50\%$	Inaccurate prediction

3. RESULT AND DISCUSSION

This section presents the results and discussion of the study on estimating toddler weight growth in Banyuwangi using Local Linear Regression based on Z-score curves. The reported results include the optimal bandwidth, MSE, RMSE, MAPE, and R^2 for both male and female toddlers. Prior to modeling, descriptive analysis is presented through scatter plots to illustrate the distribution and growth pattern of weight by age. This visualization provides an initial overview of the growth trend that underlies the subsequent estimation process.

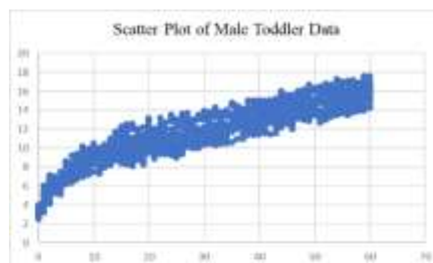

Figure 1. Scatter plot of Male toddler weight by age in Banyuwangi.

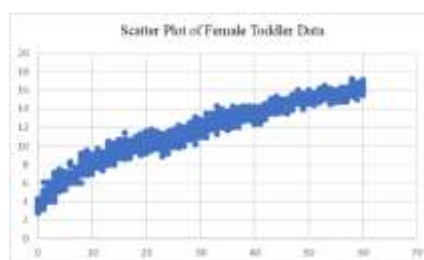

Figure 2. Scatter plot of Female toddler weight by age in Banyuwangi.

Based on the observed growth pattern in the descriptive analysis, the nonparametric estimation is implemented using the Gaussian Kernel (GaussK) weighting function. The optimal bandwidth parameter is determined through Cross-Validation (CV) to minimize prediction error. Table 4 presents

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 3 March - 2026, Pages: 9-16

the selected bandwidth values along with the corresponding MSE, MAPE, and R^2 for each Z-score curve of male toddlers.

Table 2. GaussK weight function for male toddler

Method	Z-Scores	h	MSE	MAPE	R^2
CV	(-3 SD)	1.93	0.12	3.71%	0.984
	(-2 SD)	1.76	0.33	2.36%	0.991
	(-1 SD)	1.80	0.17	1.84%	0.994
	(0 SD)	1.40	0.13	1.57%	0.996
	(1 SD)	1.00	0.03	1.28%	0.996
	(2 SD)	3.12	0.11	2.35%	0.988
	(3 SD)	3.53	0.18	2.78%	0.983

Table 2 indicates that the Gaussian Kernel model with bandwidth selected using Cross-Validation produces MAPE values ranging from 1.28% to 3.71% across all Z-score curves. Based on the MAPE criteria (MAPE ≤ 10%), these results fall within the category of highly accurate prediction. Additionally, the R^2 values range from 0.983 to 0.996, indicating that more than 98% of the variability in toddler weight is explained by the model. The MSE values are also relatively small across all curves, reflecting low estimation error. These results demonstrate that the model provides a statistically consistent fit across all Z-score levels. Based on the evaluation metrics discussed above, the estimated growth curves are visualized in Figure 3 to demonstrate the overall pattern and smoothness of the fitted model.

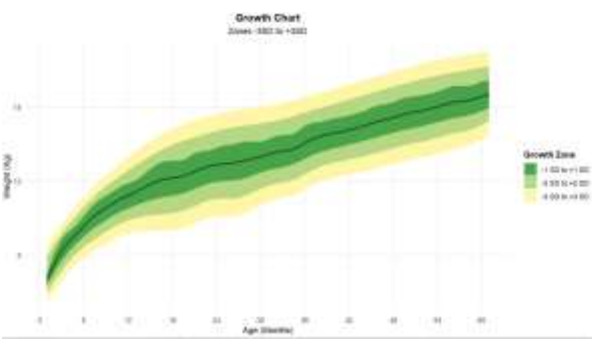

Figure 3. Growth chart estimation for male toddlers

Next, the model is obtained by estimating the parameter $\hat{\beta}$ using RStudio with the optimal bandwidth obtained previously. Based on the estimated $\hat{\beta}$, a local linear regression model is constructed to describe the relationship between toddler age and weight. Since the value of $\hat{\beta}$ differs across obsevations, the resulting model also varies accordingly. The estimated values of the local linear parameters $\hat{\beta}$ are presented below.

The estimated model of weight for male toddlers at two month is as follows:

$$\hat{Y} = 5.037 + 0.65(x - 2) \quad (19)$$

where $(x|2 - 1.40 < x < 2 + 1.40)$.

Based on this model, within the specified interval, each one-month increase in age is associated with an estimated increase of 0.65 kg in body weight. This coefficient represents the local rate of change of weight around the age of two months.

Table 3 presents the estimation results of the Local Linear Regression model using the Gaussian Kernel (GaussK) weight function for female toddlers. The optimal bandwidth is selected using the Cross-Validation (CV) method, and the corresponding performance metrics for each Z-score curve are reported.

Table 3. GaussK weight function for female toddler

Method	Z-Scores	h	MSE	MAPE	R^2
CV	(-3 SD)	3.00	0.06	2.32%	0.995
	(-2 SD)	3.00	0.05	1.95%	0.995
	(-1 SD)	2.07	0.04	1.51%	0.976
	(0 SD)	2.00	0.02	2.00%	0.995
	(1 SD)	1.52	0.05	1.75%	0.994
	(2 SD)	1.79	0.11	2.22%	0.987
	(3 SD)	1.78	0.14	2.50%	0.985

Table 3 presents the estimation results of the Gaussian Kernel (GaussK) model for female toddlers with bandwidth selected using the Cross-Validation (CV) method. The MAPE values range from 1.51% to 2.50% across all Z-score levels. Based on the MAPE criteria (MAPE ≤ 10%), these results fall within the category of highly accurate prediction. The MSE values are relatively small, with the lowest MSE observed at the median curve (0 SD) of 0.02. Meanwhile, the coefficient of determination (R^2) ranges from 0.976 to 0.995, indicating that more than 97% of the variability in female toddler weight is explained by the model across all Z-score curves. These results indicate that the Gaussian Kernel model with Cross-Validation bandwidth selection provides consistent estimation performance for female toddler weight growth at all Z-score levels. The estimated growth chart for female toddlers is shown in Figure 4, displaying the smoothed Z-score curves obtained from the Gaussian Kernel estimation using the CV method.

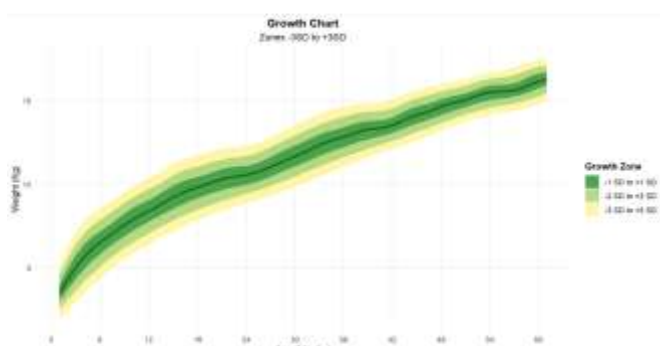

Figure 4. Growth chart estimation for female toddlers

The estimated model of weight for female toddlers at three month is as follows:

$$\hat{Y} = 6.889 + 0.362(x - 3) \quad (20)$$

where $(x|3 - 2 < x < 3 + 2)$.

Based on Equation 20, within the specified interval, a one-month increase in age is associated with an estimated increase of 0.362 kg in body weight. The coefficient 0.362 represents the local rate of change of weight around the age of three months.

The comparison results of children nutritional status based on Banyuwangi Standard Chart (Local) and WHO 2005 Standard Chart by sex (female and male) as given in Table 6.

Table 4. Percentage of nutritional status for female and male

Nutritional Status	Female		Male	
	Local	WHO 2005	Local	WHO 2005
Saverely underweight	1.90%	0.00%	0.00%	0.00%
Underweight	0.63%	0.98%	5.71%	1.48%
Normal weight	82.91%	84.72%	76.42%	84.18%
Overweight	14.56%	14.30%	17.86%	14.34%

Table 4 indicates that both the Banyuwangi Local and WHO 2005 standard charts show normal weight as the predominant nutritional status for both sexes. However, the local standard generally classifies fewer children as normal weight and more as underweight or overweight, particularly among males. In females, the local chart identifies a small proportion of severely underweight cases not detected by the WHO standard, while in males it detects higher proportions of underweight and overweight. This suggests that the Banyuwangi Local Standard Chart may be more sensitive to population-specific nutritional variations than the WHO 2005 reference.

4. Conclusion

This study examined toddler weight growth in Banyuwangi using the Gaussian Kernel (GaussK) method with Cross-Validation (CV) for bandwidth selection. The results indicate that weight increases with age for both male and female toddlers. The CV-based Gaussian Kernel model produced low prediction errors for both groups. For male toddlers, MAPE values ranged from 1.28% to 3.71% with R^2 between 0.983 and 0.996. For female toddlers, MAPE ranged from 1.51% to 2.50% with R^2 between 0.976 and 0.995. Based on the MAPE criterion ($\leq$ 10%), the estimation results fall within the highly accurate prediction category. Greater stability is observed around the median Z-score (0 SD), while more variability appears at extreme Z-score levels. The comparison between the Banyuwangi Local and WHO 2005 standards shows differences in nutritional status classification, indicating the potential relevance of population-specific growth references. Future research could consider additional variables such as nutritional status or environmental factors that may influence growth patterns, providing a more comprehensive model for toddler development.

Acknowledgment

The authors would like to express their gratitude to Universitas Airlangga and the Faculty of Science and Technology for supporting this community service activity through the “PROGRAM KEMITRAAN MASYARAKAT (PKM)” funding, under contract number 4041/B/UN3.FST/PM.01.01/2025. The authors also thank to Mrs. Lailatul Muniroh, S.K.M., M.Kes., a lecturer in the Department of Nutrition, Faculty of Public Health, Universitas Airlangga, for providing materials and expertise on toddler nutrition education. Furthermore, the authors extend their appreciation to the Head of the Songgon Community Health Center (Puskesmas), along with the staff and integrated healthcare center (Posyandu) volunteers, for their support, attention, and positive feedback. We also express our gratitude and appreciation to the Head of Songgon Village for providing the village office as the venue for this activity. Finally, thanks are due to the Community Service Team for their continuous support and contributions, ensuring this program was implemented smoothly and provided significant benefits.

5. References

[1] Badan Pusat Statistika (BPS), “Jumlah Penduduk Menurut Kelompok Umur dan Jenis Kelamin, 2025,” 2025.

[2] Central Intelligence Agency (CIA), “Total fertility rate – Country Comparison. ,” The World Factbook.

[3] Amaluddin • 12 June 2025 15:00, “2.269 Anak di Banyuwangi Masih Menderita Stunting,” MetroTV.

[4] L. Asra Laily, S. Indarjo, J. Ilmu Kesehatan Masyarakat, F. Ilmu Keolahragaan, and U. Negeri Semarang, “Literature Review: Dampak Stunting terhadap Pertumbuhan dan Perkembangan Anak,” 2023, doi: 10.15294/higeia/v7i3/63544.

[5] Selendang Sutra Nusantara, “Anggaran Rp10 Miliar Untuk Stunting Di Banyuwangi, KNPS Jatim Siap Kolaborasi Dengan Dinkes,” Selendangsutra.id.

[6] N. L. SUKERNI, I. K. G. SUKARSA, and N. L. P. SUCIPTAWATI, “PENDEKATAN REGRESI SPLINE UNTUK MEMODELKAN POLA PERTUMBUHAN BERAT BADAN BALITA,” *E-Jurnal Matematika*, vol. 7, no. 3, p. 259, Sep. 2018, doi: 10.24843/MTK.2018.v07.i03.p212.

[7] Syahrul Khairati, Siti Maisyaroh Fitri Siregar, Sri Wahyuni, and N. Nurhasanah, “Hubungan berat badan lahir rendah dengan kejadian stunting: Tinjauan Literatur,” *Haga Journal of Public Health*

(HJPH), vol. 1, no. 3, pp. 105–112, Jul. 2024, doi: 10.62290/hjph.v1i3.30.

[8] N. Chamidah *et al.*, “Modeling the Growth of Toddler Height in Banyuwangi for Health Assessment Using Local Linear Regression,” 2025. [Online]. Available: www.ijeais.org/ijaar

[9] M. Fatekurohman, S. N. Khasanah, Y. S. Dewi, M. Fatekurohman, S. N. Khasanah, and Y. S. Dewi, “NONPARAMETRIC REGRESSION MODELING USING THE SPLINE APPROACH TO STUNTING CASES IN INDONESIA,” *BAREKENG: J. Math. & App*, vol. 19, no. 2, pp. 697–0708, 2025, doi: 10.30598/barekengvol19iss2pp0697-0708.

[10] N. Chamidah, B. Lestari, T. N. Larasati, and L. Muniroh, “Designing Z-score standard growth charts based on height-for-age of toddlers using local linear estimator for determining stunting,” 2024, p. 030002. doi: 10.1063/5.0225156.

[11] M. Rifada, N. Chamidah, R. A. Ningrum, and L. Muniroh, “Stunting determinants among toddlers in Probolinggo district of Indonesia using parametric and nonparametric ordinal logistic regression models,” *Communications in Mathematical Biology and Neuroscience*, 2023, doi: 10.28919/cmbn/6690.

[12] M. de Onis, A. W. Onyango, E. Borghi, A. Siyam, C. Nishida, and J. Siekmann, “Development of a WHO growth reference for school-aged children and adolescents,” *Bull. World Health Organ.*, vol. 85, no. 09, pp. 660–667, Sep. 2007, doi: 10.2471/BLT.07.043497.

[13] T. J. Cole, “The LMS method for constructing normalized growth standards,” *Eur. J. Clin. Nutr.*, vol. 44, no. 1, pp. 45–60, 1990.

[14] M. Salibian-Barrera, “Robust nonparametric regression: Review and practical considerations,” *Econom. Stat.*, Apr. 2023, doi: 10.1016/j.ecosta.2023.04.004.

[15] L. Selk and J. Gertheiss, “Nonparametric regression and classification with functional, categorical, and mixed covariates,” *Adv. Data Anal. Classif.*, vol. 17, no. 2, pp. 519–543, Jun. 2023, doi: 10.1007/s11634-022-00513-7.

[16] A. M. Sadek and L. A. Mohammed, “Evaluation of the Performance of Kernel Non-parametric Regression and Ordinary Least Squares Regression,” *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 3, p. 1352, Sep. 2024, doi: 10.62527/joiv.8.3.2430.

[17] Y. Linke, I. Borisov, P. Ruzankin, V. Kutsenko, E. Yarovaya, and S. Shalnova, “Multivariate Universal Local Linear Kernel Estimators in Nonparametric Regression: Uniform Consistency,” *Mathematics*, vol. 12, no. 12, p. 1890, Jun. 2024, doi: 10.3390/math12121890.

[18] D. Botani, N. Kareem, T. Ali, and B. Sedeeq, “Optimizing bandwidth parameter estimation for non-parametric regression using fixed-form threshold with Dmey and Coiflet wavelets,” *Hacettepe Journal of Mathematics and Statistics*, vol. 54, no. 3, pp. 1094–1106, Jun. 2025, doi: 10.15672/hujms.1605499.

[19] H. Faridah, D. U. Haq, N. Budiantara, J. Dwi, and T. Purnomo, “PEMILIHAN PARAMETER OSILASI OPTIMAL MENGGUNAKAN GENERALIZED CROSS-VALIDATION (GCV) PADA REGRESI NONPARAMETRIK DERET FOURIER,” vol. 14, no. 2, pp. 577–587, 2025, doi: 10.14710/j.gauss.14.2.577-587.

[20] D. Barreiro-Ures, R. Cao, M. Francisco-Fernández, and R. Fernández-Casal, “Bagging cross-validated bandwidth selection in nonparametric regression estimation with applications to large-sized samples,” *Comput. Stat. Data Anal.*, vol. 213, p. 108257, Jan. 2026, doi: 10.1016/j.csda.2025.108257.

[21] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[22] H. Khoshvaght, R. R. Permala, A. Razmjou, and M. Khiadani, “A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction,” *J. Environ. Chem. Eng.*, vol. 13, no. 6, p. 119675, Dec. 2025, doi: 10.1016/j.jece.2025.119675.

[23] T. O. Hodson, T. M. Over, and S. S. Foks, “Mean Squared Error, Deconstructed,” *J. Adv. Model. Earth Syst.*, vol. 13, no. 12, Dec. 2021, doi: 10.1029/2021MS002681.

[24] R. I. Rauf, M. A. Alrasheedi, R. Sadiq, and A. M. A. Aldawsari, “Evaluating Predictive Accuracy of Regression Models with First-Order Autoregressive Disturbances: A Comparative Approach Using Artificial Neural Networks and Classical Estimators,” *Mathematics*, vol. 12, no. 24, p. 3966, Dec. 2024, doi: 10.3390/math12243966.

[25] S. G. . Makridakis, S. C. . Wheelwright, and R. J. . Hyndman, *Forecasting : methods and applications*. John Wiley & Sons, 1998.