MethodsX 17 (2026) 104228

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

A multi-predictor nonparametric local linear generalized Poisson regression model for overdispersed count data☆,☆☆

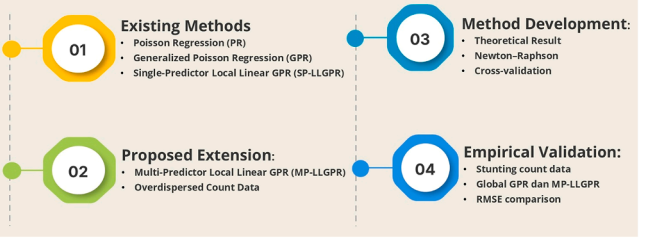
Darnah [a], M. Fathurahman [a,*], Iriyani Kamaruddin [b], Suyitno [a], Toha Saifudin [c], Mardiana Idayu Ahmad [d], Erni Tri Astuti [e], Ratna Kusuma [f], Rahmawati Munir [g], Made Dirgantara [h], Fidia Deny Tisna Amijaya [i], Imam Rofi'i [j], Andi Syaifullah Al Maslulil Agri [k], Feni Arista Bahni [a]

[a] *Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*
[b] *Nutrition Study Program, Faculty of Public Health, Mulawarman University, Samarinda, Indonesia*
[c] *Statistics Study Program, Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia*
[d] *School of Industrial Technology, Universiti Sains Malaysia, 11800 Penang, Malaysia*
[e] *Polytechnic of Statistics STIS, Jakarta, Indonesia*
[f] *Biology Study Program, Department of Biology, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*
[g] *Physics Study Program, Department of Physics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*
[h] *Physics Study Program, Faculty of Mathematics and Natural Sciences, Palangka Raya University, Palangka Raya, Indonesia*
[i] *Mathematics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*
[j] *Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia*
[k] *Environmental Science Study Program, Department of Biology, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda, Indonesia*

HIGHLIGHTS

- Develops a multi-predictor extension of nonparametric local linear generalized Poisson regression model.
- Implements leave-one-out cross-validation (CV) based bandwidth selection for data-driven local estimation.
- The method is empirically validated in healthcare application using stunting count data in Kalimantan, Indonesia.

GRAPHICAL ABSTRACT



☆ Related research article None.☆☆ For a published article: None.

* Corresponding author.
E-mail address: fathur@fmipa.unmul.ac.id (M. Fathurahman).

https://doi.org/10.1016/j.mex.2026.104228
Received 28 August 2026; Accepted 22 October 2026
Available online 23 October 2026
2215-0161/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

ARTICLE INFO

Keywords:
Nonparametric regression
Local linear estimator
Generalized Poisson distribution
Overdispersed
Count data

ABSTRACT

Generalized Poisson regression is widely used for modeling overdispersed count data; however, existing nonparametric local linear generalized Poisson regression methods are generally restricted to a single predictor, limiting their applicability to datasets involving multiple predictors. This study extends the nonparametric local linear generalized Poisson regression framework to accommodate multiple predictors, namely multi-predictor local linear generalized Poisson regression (MP-LLGPR). The proposed method provides a flexible approach for modeling overdispersed count response variable with locally varying relationships involving multiple predictors. The method is empirically validated in a healthcare application using stunting count data from 44 regencies and cities in Kalimantan, Indonesia. The empirical validation demonstrates the applicability of the proposed method for modeling overdispersed count data with multiple predictors and provides a comparison with the corresponding global generalized Poisson regression model.

Specifications table

Subject area	Mathematics and Statistics
More specific subject area	Statistical Modeling; Nonparametric Regression; Generalized Poisson Regression; Local Linear Regression; Overdispersed Count Data.
Name of your method	Multi-predictor Nonparametric Local Linear Generalized Poisson Regression (MP-LLGPR)
Name and reference of original method	Darnah, Memi Nor Hayati, Sri Wahyuningsih, Iriyani Kamaruddin, Suyitno, Andrea Tri Rian Dani, Rito Goejantoro, Desi Yuniarti, Fidia Deny Tisna Amijaya, Ika Purnamasari, Meiliyani Siringoringo, Surya Prangga, Ratna Kusuma, Rahmawati Munir, "Locally kernel weighted maximum likelihood estimator for local linear multi-predictor poisson regression", MethodsX, 14(103,258), 2025, https://doi.org/10.1016/j.mex.2025.103258. Darnah, M. I. Utoyo, and N. Chamidah, "Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression Approach Based on Local Linear Estimator," in IOP Conference Series: Earth and Environmental Science, Institute of Physics Publishing, Apr. 2019. doi:10.1088/1755–1315/243/1/012,023. D. Darnah, M. Imam Utoyo, and N. Chamidah, "Estimation of the Bi-response Poisson Regression Model Based on Local Linear Approach," International Journal of Academic and Applied Research, vol. 3, no. 5, pp. 14–18, 2019, [Online]. Available: http://www.ijeais.org/ijaar Darnah, I. Utoyo, and N. Chamidah, "Bi-response Poisson regression model for modeling effect of early marriage on maternal and infant mortality," AIP Conf Proc, vol. 2554, no. 1, p. 030,005, Jan. 2023, doi: 10.1063/5.0104404.
Resource availability	None

Background

Poisson regression is one of the most widely used statistical methods for modeling count data because it relates the expected count to a set of explanatory variables through a log-link function. However, the classical Poisson regression model relies on the equidispersion assumption, in which the mean and variance of the response variable are equal. In practice, this assumption may be violated because of overdispersion or underdispersion, which can affect parameter estimation, standard errors, and statistical inference [1–8] . To address this limitation, Wang and Famoye [9] introduced the generalized Poisson regression (GPR) model, which extends the classical Poisson regression by incorporating a dispersion parameter and provides a flexible framework for modeling equidispersed, overdispersed, and underdispersed count data.

Regression models can be developed using either parametric or nonparametric approaches. Parametric regression assumes a predetermined functional form for the relationship between the response and predictor variables [10–18] . However, this assumption may be restrictive when the underlying relationships are complex or nonlinear. Nonparametric regression provides a flexible alternative by estimating the regression function from the observed data without imposing a specific parametric functional form [19–21].

Various smoothing approaches have been developed for nonparametric regression, including kernel, spline, penalized spline, truncated spline, local polynomial, and local linear estimators [21–43] . Among these approaches, local linear estimation provides a flexible mechanism for approximating an unknown regression function locally and has been applied to Poisson regression models for modeling nonlinear relationships between count responses and predictor variables [1,2,11,12,35,44].

The local linear approach has also been extended to generalized Poisson regression to accommodate nonlinear relationships in overdispersed count data. Astuti [22,45] developed a local likelihood-based generalized Poisson regression approach; however, the formulation is restricted to a single predictor variable, limiting its applicability when the response is simultaneously associated with multiple predictors. In addition, the existing approach determines the smoothing parameter using the root mean square error (RMSE), whereas bandwidth selection can instead be formulated as a cross-validation problem. These limitations motivate an extension of the local linear generalized Poisson regression framework to accommodate multiple predictors and to use a data-driven cross-validation criterion for bandwidth selection.

Accordingly, this study develops a MP-LLGPR method for modeling overdispersed count data. The proposed extension incorporates multiple predictors within a local linear generalized Poisson framework, while bandwidth selection is performed using CV. The method provides a unified framework for locally estimating nonlinear relationships between an overdispersed count response and multiple predictors. Its empirical performance is validated using stunting count data from 44 regencies and cities in Kalimantan, Indonesia. The stunting dataset used to demonstrate the applicability and model performance of the proposed method rather than as a substantive analysis of determinants of stunting.

Method details

Poisson regression model

The response variable in a Poisson regression model is assumed to follow a Poisson distribution, which belongs to the exponential family of distributions. This model is classified as a generalized linear model consisting of three components, that is random component, systematic component, and link function. Let Ybe the random variable for count data and it follows a Poisson distribution, the probability density function (PDF) is

$$P(Y=y|\theta)=f(y,\theta)=\frac{\theta^{y}\exp(-\theta)}{y!},\ y=0,\ 1,\ 2,\ \ldots \tag{1}$$

with mean and variance,

$$E(Y)=Var(Y)=\theta \tag{2}$$

If Eq. (1) rewritten in the form of the PDF of the exponential family of distributions, then [46]

$$f(y,\theta)=exp[y\log(\theta)-\theta-\log(y_i!) \tag{3}$$

where $Y$is a random component, $E(Y)\ =\theta$, and $g(\theta)=log(\theta)$ is canonical link function. The link function relates the response variable Y and the k predictors X, yielding the Poisson regression model as follows:

$$\theta_i=\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right) \tag{4}$$

with $\boldsymbol{x}_i^T=[1\quad x_{1i}\quad x_{2i}\quad \ldots\quad x_{ki}]$ is the i th observation result data vector of the predictor variables and $\boldsymbol{\beta}=[\beta_0\quad \beta_1\quad \beta_2\quad \ldots\quad \beta_k]^T$. Poisson regression model given by Eq. (4) is nonlinear regression [47]. A commonly used parameter estimation method for nonlinear regression models, such as Weibull regression model [48,49] and ordinal logistic regression model [50], is maximum likelihood estimation (MLE) by maximizing likelihood function. The likelihood function in Eq. (1) is defined as

$$\ell(\boldsymbol{\beta})=\prod_{i=1}^{n}f(y_i,\theta_i)=\prod_{i=1}^{n}\frac{\exp\left[y_i\boldsymbol{x}_i^T\boldsymbol{\beta}-\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right]}{y_i!} \tag{5}$$

and the log-likelihood function is

$$L(\boldsymbol{\beta})=log(l(\boldsymbol{\beta}))=\sum_{i=1}^{n}\left(y_i\boldsymbol{x}_i^T\boldsymbol{\beta}-\exp\left(y_i\boldsymbol{x}_i^T\boldsymbol{\beta}\right)-\log(y_i!)\right) \tag{6}$$

The vector $\boldsymbol{\beta}$ can be obtained by maximizing log-likelihood function in Eq. (6) and setting its first derivative equal to zero, yielding

$$\frac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\boldsymbol{0} \tag{7}$$

Since Eq. (7) has no closed-form analytical solution, a numerical method is required to obtain the maximum likelihood estimator. The Newton-Raphson iterative algorithm is commonly used for this purpose.

Generalized Poisson regression model

According to Ismail and Jemain [8], let $(Y_i,x_i),i=1,\ 2,\ \cdots,\ n$ be a count random variable following a generalized Poisson distribution, the corresponding probability density function is defined as

$$f(y;\theta,\varphi)=\left(\frac{\theta}{1+\varphi\theta}\right)^{y}\frac{(1+\varphi y)^{y-1}}{y!}\exp\left(-\frac{\theta(1+\varphi y)}{1+\varphi\theta}\right);y=0,\ 1,\ 2,\cdots \tag{8}$$

The mean value in generalized Poisson distribution is

$$E(Y)=\theta \tag{9}$$

and the variance is

$$\mathrm{Var}(Y) = \theta(1+\varphi\theta)^2 \tag{10}$$

The parameter φ in Eq. (8) represents the dispersion parameter. When $\varphi = 0$, the model reduces to the Poisson distribution, when $\varphi < 0$, it indicates underdispersion, and when $\varphi > 0$, it indicates overdispersion relative to Poisson distribution [23,47]. The mean of the response variable $E(Y|\boldsymbol{x}) = \theta(\boldsymbol{x})$in the global generalized Poisson regression is assumed to depend on some predictors through a link function

$$\theta(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\boldsymbol{\beta}) \tag{11}$$

where$\boldsymbol{\beta}$ is the parameter vector of the regression model. The parameters are estimated using MLE. Based on Eq. (8), the likelihood function can be expressed

$$\ell(\theta,\varphi,y_i) = \prod_{i=1}^{n}\left(\frac{\theta}{1+\varphi\theta}\right)^{y_i}\frac{(1+\varphi y_i)^{y_i-1}}{y_i!}\exp\left(-\frac{\theta(1+\varphi y_i)}{1+\varphi\theta}\right) \tag{12}$$

and log-likelihood function is

$$L(\boldsymbol{x}^T\boldsymbol{\beta},\varphi,y_i) = \sum_{i=1}^{n}\left\{y_i\ln\left(\frac{\boldsymbol{x}_i^T\boldsymbol{\beta}}{1+\varphi\boldsymbol{x}_i^T\boldsymbol{\beta}}\right) + (y_i-1)\ln(1+\varphi y_i) - \ln(y_i!) - \frac{\boldsymbol{x}_i^T\boldsymbol{\beta}(1+\varphi y_i)}{1+\varphi\boldsymbol{x}_i^T\boldsymbol{\beta}}\right\} \tag{13}$$

The maximum likelihood estimator for $\boldsymbol{\beta}$and φ are obtained by solving the maximum likelihood equations.

$$\frac{\partial L}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{y_i - exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{(1+\varphi exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))^2} = \mathbf{0} \tag{14}$$

$$\frac{\partial L}{\partial\varphi} = \sum_{i=1}^{n}\left\{\frac{-y_i exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{1+\varphi exp(\boldsymbol{x}_i^T\boldsymbol{\beta})} + \frac{y_i(y_i-1)}{1+\varphi y_i} - \frac{exp(\boldsymbol{x}_i^T\boldsymbol{\beta})(y_i - exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))}{(1+\varphi exp(\boldsymbol{x}_i^T\boldsymbol{\beta}))^2}\right\} = 0 \tag{15}$$

Since Eq. (14) and (15) are nonlinear in parameters, the solution can be obtained by using some iterative methods such as Newton Raphson.

Nonparametric local linear generalized Poisson regression model for a predictor variable

In the nonparametric local linear generalized Poisson regression model, rather than assuming a specific regression function, the relationship between the mean response and a predictor variable is represented by a smooth nonparametric regression function as follows:

$$E(Y_i|x) = \theta_i(x) = \exp[m(x_i)] \tag{16}$$

Suppose m is a continuous function that has a $(p+1)$-th derivative at the point x_0. Then, for any point $x_i$in the neighborhood of x_0, the function m can be expanded into the following Taylor series of degree p.

$$m(x_i) \approx m(x_0) + m'(x_0)(x_i - x_0) + \frac{m''(x_0)(x_i-x_0)^2}{2!} + \cdots + \frac{m^{(p)}(x_0)(x_i-x_0)^p}{p!} = \boldsymbol{x}_i^{*T}\boldsymbol{\beta} \tag{17}$$

where $\boldsymbol{x}_i^* = \begin{bmatrix}1 & (x_i-x_0) & (x_i-x_0)^2 & \cdots & (x_i-x_0)^p\end{bmatrix}^T$, $\boldsymbol{\beta} = \begin{bmatrix}\beta_0 & \beta_1 & \beta_2 & \cdots & \beta_p\end{bmatrix}^T$, with $\boldsymbol{\beta}_\nu = s^\nu(x_0)/\nu!$ and $\nu = 0, 1, 2, \ldots, p$.

The regression curve at point x_0 in nonparametric regression is estimated using the locally kernel weighted maximum likelihood estimator method, in which the model parameters are first estimated. In Eq. (17), if p = 0, the technique is called local constant and if p = 1, it is known as local linear model [23].

Based on Eq. (8) and (17), we obtain the local kernel weighted log-likelihood function

$$L_p(\boldsymbol{x}_i^{*T}\boldsymbol{\beta},\varphi,y_i) = \sum_{i=1}^{n}K_h(x_i-x_0)\left\{y_i\ln\left(\frac{\boldsymbol{x}_i^{*T}\boldsymbol{\beta}}{(1+\varphi\boldsymbol{x}_i^{*T}\boldsymbol{\beta})}\right) + (y_i-1)\ln(1+\varphi y_i) - \ln(y_i!) - \frac{\boldsymbol{x}_i^{*T}\boldsymbol{\beta}(1+\varphi y_i)}{1+\varphi\boldsymbol{x}_i^{*T}\boldsymbol{\beta}}\right\} \tag{18}$$

where K is kernel function and h > 0 represent the bandwidth. The bandwidth h controls smoothness of the fit curve. If h is too small, the fit becomes too noisy, and the variance increases. Conversely, if h is too large, the fit becomes over-smoothed and important feature may be distorted or lost completely. That is, the fit will have large bias. The bandwidth must be chosen to compromise this bias-variance trade-off. The CV method is a commonly used technique to determine the optimal bandwidth by minimizing the CV value.

Based on Eq. (18), the estimators $\boldsymbol{\beta}$and φ are obtained by solving the (p + 2) equations.

$$\frac{\partial L_p}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}K_h(x_i-x_0)\frac{y_i - \exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta})}{(1+\varphi\exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta}))^2}\boldsymbol{x}_i = \mathbf{0} \tag{19}$$

$$\frac{\partial L_p}{\partial\varphi} = \sum_{i=1}^{n}K_h(x_i-x_0)\left\{\frac{-y_i\exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta})}{1+\varphi\exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta})} + \frac{y_i(y_i-1)}{1+\varphi y_i} - \frac{\exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta})(y_i - \exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta}))}{(1+\varphi\exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta}))^2}\right\} = 0 \tag{20}$$

The solutions to Eqs. (19) and (20) can be obtained using the Newton–Raphson method.

Proposed multi-predictor nonparametric local linear generalized Poisson regression (MP-LLGPR) model

Let $(y_i, \boldsymbol{x}_i)$, $i = 1, 2, \cdots, n$ be the observed data pairs, where y_i is the count response that follows the generalized Poisson distribution and $\boldsymbol{x}_i = (\, x_{i1} \quad x_{i2} \quad x_{i3} \quad \ldots \quad x_{ik}\,)$ is a vector of k predictor variables. The dependence of the mean response on the k predictors is described by a smooth nonparametric regression function

$$\theta(\boldsymbol{x}_i) = \exp[m(\boldsymbol{x}_i)] \tag{21}$$

where $m(\boldsymbol{x}_i)$ is an unknown function, from the exponential family function, and continuous function that has a $(p+1)$-th derivative at the point $x_0$that we aim to estimate using locally kernel weighted maximum likelihood estimator method.

The assigned weights are determined by the relative distances between the observed predictor values and the target point x_0, such that observations closer to x_0 receive larger weights than those farther away. The kernel-weighted likelihood function at the point x_0, which depends on the polynomial degree p and the bandwidth [2], can be expressed

$$\ell(\beta, \varphi; \boldsymbol{x}_0) = \prod_{i=1}^{n} f(y_i; \theta(\boldsymbol{x}_i), \varphi)^{\prod_{j=1}^{k} K_{h_j}(x_{ij} - x_{0j})} \tag{22}$$

where $h_j$denotes the bandwidth associated with the j-th predictor. By approximating $m(\boldsymbol{x}_i)$using a first-order Taylor expansion, we obtain

$$m(\boldsymbol{x}_i) = \beta_0 + \sum_{j=1}^{k} \beta_j (x_{ij} - x_{0j}) \tag{23}$$

Eq. (23) can be expressed in vector form as follows

$$m(\boldsymbol{x}_i) = \boldsymbol{x}_i^{*T} \boldsymbol{\beta} \tag{24}$$

where $\boldsymbol{\beta}$is the vector of local regression coefficients. The complete parameter vector of the MP-LLGPR model is defined as

$$\psi = \begin{bmatrix} \boldsymbol{\beta} \\ \varphi \end{bmatrix} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k, \varphi)^T \tag{25}$$

The local mean parameter is given by $\theta(\boldsymbol{x}_i) = \exp(\boldsymbol{x}_i^{*T}\boldsymbol{\beta})$. Based on Eq. (21), Eq. (8) can be expressed

$$f(y_i; \theta(\boldsymbol{x}_i), \varphi) = \left(\frac{\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right)^{y_i} \frac{(1+\varphi y_i)^{y_i - 1}}{y_i!} \exp\left(-\frac{\theta(\boldsymbol{x}_i)(1+\varphi y_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right) \tag{26}$$

Definition 1. Let we have the Kernel-weighted likelihood function in Eq. (22) and the probability density function of y_i in Eq. (26) then the local likelihood function can be written as

$$\ell(y_i; \theta(\boldsymbol{x}_i), \varphi) = \prod_{i=1}^{n} \left(\frac{\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right)^{y_i} \frac{(1+\varphi y_i)^{y_i - 1}}{y_i!} \exp\left(-\frac{\theta(\boldsymbol{x}_i)(1+\varphi y_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right)^{\prod_{j=1}^{k} K_{h_j}(x_{ij} - x_{0j})} \tag{27}$$

and the local log-likelihood function is

$$\begin{aligned} L(y_i; \theta(\boldsymbol{x}_i), \varphi) &= \ln(\ell(y_i; \theta(\boldsymbol{x}_i), \varphi)) = \ln\left(\prod_{i=1}^{n} \left\{ \left(\frac{\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right)^{y_i} \frac{(1+\varphi y)^{y_i - 1}}{y_i\,!} \exp\left(-\frac{\theta(\boldsymbol{x}_i)(1+\varphi y_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right) \right\}^{\prod_{j=1}^{k} K_{h_j}(x_{ij} - x_{0j})}\right) \\ &= \sum_{i=1}^{n} \left(\prod_{j=1}^{k} K_{h_j}(x_{ij} - x_{0j})\right) \left\{ y_i \ln\left(\frac{\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}\right) + (y_i - 1)\ln(1+\varphi y_i) - \ln(y_i!) - \frac{\theta(\boldsymbol{x}_i)(1+\varphi y_i)}{1+\varphi\theta(\boldsymbol{x}_i)} \right\} \end{aligned} \tag{28}$$

Theorem. Suppose the multi-predictor nonparametric local linear generalized Poisson regression (MP-LLGPR) model specified in Eq. (21), the regression function$m(\boldsymbol{x}_i)$defined in Eq. (24), and the log-likelihood function presented in Eq. (28) can be expressed $L(y_i; \theta(\boldsymbol{x}_i), \varphi) = L$, the local weighted maximum likelihood estimators of $\boldsymbol{\beta}$and φat the point x_0 are obtained by solving the following system of equation

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left(\prod_{j=1}^{k} K_{h_j}(x_{ij} - x_{0j})\right) \left\{ \frac{y_i - \theta(\boldsymbol{x}_i)}{(1+\varphi\theta(\boldsymbol{x}_i))^2} \boldsymbol{x}_i^* \right\} = \boldsymbol{0} \tag{29}$$

$$\frac{\partial L}{\partial \varphi}=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{-\frac{(y_i)\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}+\frac{y_i(y_i-1)}{1+\varphi y_i}-\frac{\theta(\boldsymbol{x}_i)(y_i-\theta(\boldsymbol{x}_i))}{(1+\varphi\theta(\boldsymbol{x}_i))^2}\right\}=0 \tag{30}$$

Proof. The maximum likelihood estimator for the MP-LLGPR model can be obtained by maximizing the log-likelihood function presented in Eq. (29). The first derivative of Eq. (29) with respect to $\boldsymbol{\beta}$ using the chain rule can be expressed as

$$\frac{\partial L}{\partial \boldsymbol{\beta}}=\frac{\partial L}{\partial \theta(\boldsymbol{x}_i)}\frac{\partial \theta(\boldsymbol{x}_i)}{\partial \boldsymbol{\beta}} \tag{31}$$

where

$$\begin{aligned}\frac{\partial L}{\partial \theta(\boldsymbol{x}_i)}&=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{y_i\left(\frac{1+\theta(\boldsymbol{x}_i)}{\theta(\boldsymbol{x}_i)}\right)\left(\frac{1}{(1+\varphi\theta(\boldsymbol{x}_i))^2}\right)-(1+\varphi y_i)\left(\frac{1}{(1+\varphi\theta(\boldsymbol{x}_i))^2}\right)\right\}\\&=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{\frac{y_i-\theta(\boldsymbol{x}_i)}{\theta(\boldsymbol{x}_i)(1+\varphi\theta(\boldsymbol{x}_i))^2}\right\}\end{aligned} \tag{32}$$

and

$$\frac{\partial \theta(\boldsymbol{x}_i)}{\partial \boldsymbol{\beta}}=\frac{\partial}{\partial \boldsymbol{\beta}}\left(\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)=\boldsymbol{x}_i^*\theta(\boldsymbol{x}_i) \tag{33}$$

Eq. (32) and (33) substituted into Eq. (31), we obtained

$$\frac{\partial L}{\partial \boldsymbol{\beta}}=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{\frac{y_i-\theta(\boldsymbol{x}_i)}{(1+\varphi\theta(\boldsymbol{x}_i))^2}\boldsymbol{x}_i^*\right\}=\mathbf{0} \tag{34}$$

Furthermore, the first derivative of Eq. (28) with respect to φ can be expressed as

$$\begin{aligned}\frac{\partial L}{\partial \varphi}&=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{-y_i\frac{\partial}{\partial\varphi}\ln(1+\varphi\theta(\boldsymbol{x}_i))+(y_i-1)\frac{\partial}{\partial\varphi}\ln(1+\varphi y_i)-\frac{\partial}{\partial\varphi}\frac{1+\varphi y_i}{1+\varphi\theta(\boldsymbol{x}_i)}\right\}\\&=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{-\frac{y_i\theta(\boldsymbol{x}_i)}{1+\varphi\theta(\boldsymbol{x}_i)}+\frac{y_i(y_i-1)}{1+\varphi y_i}-\frac{\theta(\boldsymbol{x}_i)(y_i-\theta(\boldsymbol{x}_i))}{(1+\varphi\theta(\boldsymbol{x}_i))^2}\right\}=0\end{aligned} \tag{35}$$

Eq. (34) and (35) are implicit and therefore require an iterative procedure to obtain their solutions. In this study, the Newton–Raphson iterative method [41] is employed to estimate the parameters.

Definition 2. Let the first derivatives of the local log-likelihood function with respect to the parameter$\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2,\cdots,\beta_k)^T$and φ, as shown in Eq. (34) and (35), $x_{ij};\ j=1,2,\ldots,k$ is predictor variable, the gradient vector is

$$\mathbf{g}(\psi)=\frac{\partial L}{\partial \psi}=\begin{bmatrix}\frac{\partial L}{\partial \boldsymbol{\beta}}\\ \frac{\partial L}{\partial \varphi}\end{bmatrix} \tag{36}$$

and the Hessian matrix is

$$\mathbf{H}(\psi)=\frac{\partial^2 L}{\partial\psi\partial\psi^T}=\begin{bmatrix}\mathbf{H}_{\beta\beta} & \mathbf{H}_{\beta\varphi}\\ \mathbf{H}_{\varphi\beta} & \mathbf{H}_{\varphi\varphi}\end{bmatrix} \tag{37}$$

The elements of the Hessian matrix in Eq. (37) can be expressed

$$\mathbf{H}_{\beta\beta}=\frac{\partial^2 L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}=-\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\theta(\boldsymbol{x}_i)\left\{\frac{1}{(1+\varphi\theta(\boldsymbol{x}_i))^2}+\frac{2\varphi(y_i-\theta(\boldsymbol{x}_i)}{(1+\varphi\theta(\boldsymbol{x}_i))^3}\boldsymbol{x}_i^*\boldsymbol{x}_i^{*T}\right\} \tag{38}$$

$$\mathbf{H}_{\varphi\varphi}=\frac{\partial^2 L}{\partial\varphi^2}=\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{\frac{y_i(\theta(\boldsymbol{x}_i))^2}{(1+\varphi\theta(\boldsymbol{x}_i))^2}-\frac{y_i^2(y_i-1)}{(1+\varphi y_i)^2}+\frac{2(\theta(\boldsymbol{x}_i))^2(y_i-\theta(\boldsymbol{x}_i))}{(1+\varphi\theta(\boldsymbol{x}_i))^3}\right\} \tag{39}$$

$$\mathbf{H}_{\beta\varphi}=\frac{\partial^2 L}{\partial\boldsymbol{\beta}\partial\varphi}=-2\sum_{i=1}^{n}\left(\prod_{j=1}^{k}K_{h_j}\left(x_{ij}-x_{0j}\right)\right)\left\{\frac{\theta(\boldsymbol{x}_i)(y_i-\theta(\boldsymbol{x}_i)}{(1+\varphi\theta(\boldsymbol{x}_i))^3}\boldsymbol{x}_i^*\right\} \tag{40}$$

$$\mathbf{H}_{\varphi\boldsymbol{\beta}}=\mathbf{H}_{\beta\varphi}^T \tag{41}$$

The Fisher information matrix is constructed as follows

$$\mathbf{I}(\psi) = -E\left[\frac{\partial^2 L}{\partial\psi\partial\psi^T}\right] \tag{42}$$

Using the gradient vector and Hessian matrix obtained from Eq. (36) and (37), the Newton–Raphson iterative procedure is employed to estimate the model parameters.

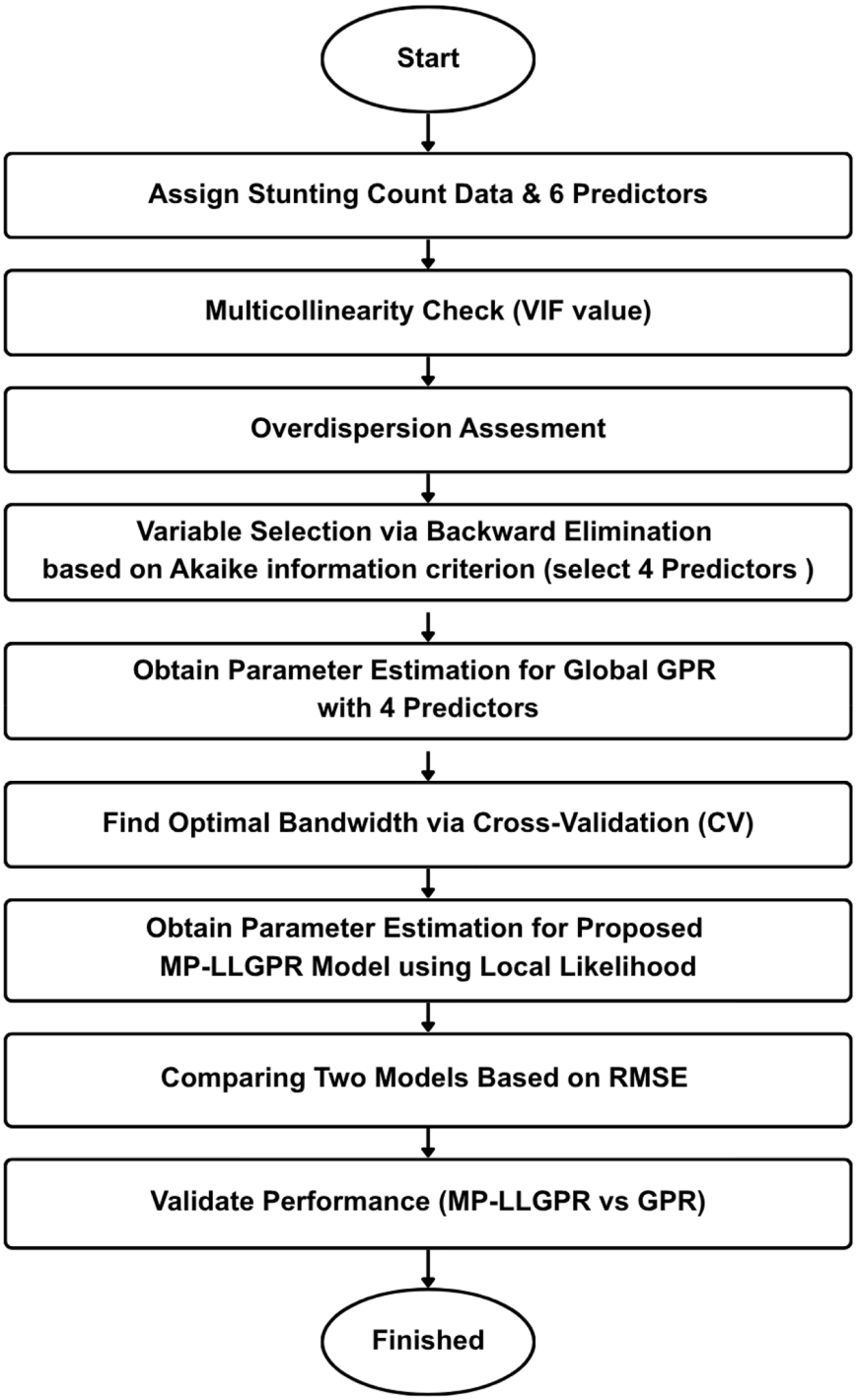


Fig. 1. Flowchart of the MP-LLGPR methodology.

$$\psi^{(m+1)} = \psi^{(m)} - \mathbf{H}^{-1}\left(\psi^{(m)}\right)\mathbf{g}(\psi^{(m)}); m = 0, 1, 2, .. \tag{43}$$

The iterative procedure begins with the selection of initial parameter estimates

$$\widehat{\psi}^{(0)}(x_0) = \left[\widehat{\beta}_0^{(0)}(x_0), \widehat{\beta}_1^{(0)}(x_0), \cdots, \widehat{\beta}_k^{(0)}\left(x_{0j}\right), \widehat{\varphi}^{(0)}\right]^T \tag{44}$$

The initial parameter vector given in Eq. (44) and continues until convergence is achieved, the stopping criterion is $\| \widehat{\psi}^{(m+1)} - \widehat{\psi}^{(m)} \| \leq \varepsilon$, where ε is a small positive constant. The initial value of the local regression coefficients are obtained from the global Poisson regression model, while the initial dispersion parameter is set to $\varphi = 0$. Thus, the initial parameter vector is defined as

$$\widehat{\psi}^{(0)} = \left(\boldsymbol{\beta}_{Poisson}^T, 0\right)^T \tag{45}$$

Definition 4. Suppose the multi-predictor local linear generalized Poisson regression (MP-LLGPR) model is specified in Eq. (21), with the local linear predictor given in Eq. (24). For a given bandwidth vector$\boldsymbol{h} = (h_1, h_2, \cdots, h_k)^T$ the CV criterion is defined as

$$CV(\boldsymbol{h}) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \widehat{\theta}_{-i}(x_i; \boldsymbol{h})]^2 \tag{46}$$

where $\widehat{\theta}_{-i}(\boldsymbol{x}_i; \boldsymbol{h}) = \exp\left(\boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}_{-i}\right)$denotes the leave-one-out estimate of the mean parameter obtained by excluding the i th observation. The optimal bandwidth vector is obtained by minimizing the CV criterion,

$$\boldsymbol{h}_{opt} = \arg\min CV(\boldsymbol{h}) \tag{47}$$

Method validation

Empirical validation

We evaluated the proposed method using empirical validation data on stunting case counts across 44 regencies/cities in Kalimantan, Indonesia, in 2024. The dataset consists of 1 response variable, namely the number of stunting cases (Y), and 6 initial predictor variables, including the number of integrated community-based health service posts or posyandu (X_1), the number of nutrition health workers (X_2), the percentage of monthly per capita food expenditure (X_3), the percentage of the poor population (X_4), the percentage of infants aged less than six months receiving exclusive breastfeeding (X_5), and the number of low birth weight infants (X_6). The full dataset was compiled from the official publication of the Central Bureau of Statistics (BPS) and the Indonesian Ministry of Health. Fig. 1 presents a flowchart illustrating the overall computational and validation methodology of this study.

Prior to model estimation, descriptive statistical analysis was performed to summarize the initial data characteristics. The resulting descriptive statistics are presented in Table 1. The response variable (Y) displays a mean of 119 and a standard deviation of 43.60 (variance = 1900.96), where the sample variance substantially exceeds the mean, indicating initial empirical evidence of overdispersion.

Before fitting the Poisson regression model, potential multicollinearity among the predictor variables was assessed using the variance inflation factor (VIF). The resulting VIF values for all predictor variables are presented in Table 2.

All predictor variables yielded VIF values below 10 (Table 2), confirming the absence of multicollinearity. Consequently, all six initial predictor variables were included in the full Poisson regression model specification. Parameter estimation for the full Poisson regression model was performed using MLE via the Newton–Raphson algorithm, with the resulting estimates summarized in Table 3.

The initial six-predictor Poisson regression model was subsequently assessed for dispersion. The dispersion ratio, calculated as the ratio of the deviance to its residual degrees of freedom, was 9.034, which was substantially greater than 1. This result indicates that the equidispersion assumption of the Poisson model was not satisfied, providing evidence of overdispersion in the stunting case counts. Therefore, the GPR framework was considered appropriate for accommodating the observed dispersion.

To obtain a parsimonious model, backward elimination based on the AIC was applied to the six-predictor GPR model. This variable selection procedure reduced the predictor set from six to four predictors. The selected four-predictor GPR model yielded an AIC value of 438.6039 and was subsequently used as the baseline for developing the proposed MP-LLGPR model. The parameter estimates for the selected four-predictor global GPR model, along with their corresponding significance tests, are summarized in Table 4. All retained predictors were statistically significant at the 5% level, establishing a validated global baseline prior to localized bandwidth

Table 1
Descriptive Summary of the Research Variables.

Variable	Average	Minimum	Maximum	Standar Deviation
The stunting cases ($\mathbf{Y}$)	119	31	242	43.60
The number of integrated community-based health service posts or posyandu ($\mathbf{X}_1$)	321	36	1688	277.74
The number of nutrition health workers ($\mathbf{X}_2$)	58	12	177	33.62
The percentage of monthly per capita food expenditure ($\mathbf{X}_3$)	52.22	41.43	61.98	5.14
The percentage of the poor population ($\mathbf{X}_4$)	5.66	2.23	10.75	2.00
The percentage of infants aged less than six months receiving exclusive breastfeeding ($\mathbf{X}_5$)	64.00	24.80	86.60	14.94
The number of low birth weight infants ($\mathbf{X}_6$)	202	0.00	727	168.29

Table 2
VIF values of predictor variables.

Variable	VIF value	Decision
The number of integrated community-based health service posts or posyandu ($\mathbf{X}_1$)	1.899	No multicollinearity
The number of nutrition health workers ($\mathbf{X}_2$)	1.230	No multicollinearity
The percentage of monthly per capita food expenditure ($\mathbf{X}_3$)	1.275	No multicollinearity
The percentage of the poor population ($\mathbf{X}_4$)	1.097	No multicollinearity
The percentage of infants aged less than six months receiving exclusive breastfeeding ($\mathbf{X}_5$)	1.654	No multicollinearity
The number of low birth weight infants ($\mathbf{X}_6$)	1.217	No multicollinearity

Table 3
Parameter estimates of Poisson regression model using MLE.

Variable	Parameter	Estimate	P-value
Intercept	β_0	2.859	$< 2 \times 10^{-16}$
The number of integrated community-based health service posts or posyandu ($\mathbf{X}_1$)	β_1	2.860×10^{-4}	3.42×10^{-8}
The number of nutrition health workers ($\mathbf{X}_2$)	β_2	3.452×10^{-3}	3.46×10^{-10}
The percentage of monthly per capita food expenditure ($\mathbf{X}_3$)	β_3	2.029×10^{-2}	3.21×10^{-11}
The percentage of the poor population ($\mathbf{X}_4$)	β_4	4.130×10^{-2}	1.14×10^{-7}
The percentage of infants aged less than six months receiving exclusive breastfeeding ($\mathbf{X}_5$)	β_5	2.296×10^{-3}	0.0235
The number of low birth weight infants ($\mathbf{X}_6$)	β_6	7.577×10^{-4}	3.45×10^{-14}

optimization and MP-LLGPR model fitting.

The optimal bandwidth parameter was determined by minimizing the CV criterion across a specified candidate grid. The resulting bandwidth was applied to the Gaussian kernel function to calculate localized spatial weights for each observation relative to every estimation point. Under this kernel specification, nearby observations receive higher weights, whereas those further away receive exponentially lower weights. The optimal bandwidth and its corresponding CV value are presented in Table 5.

Based on the optimal bandwidth in Table 5, the parameters of the MP-LLGPR model were estimated . The resulting optimal bandwidth vector was $\boldsymbol{h}_{opt} = (320, 60, 9, 200)^T$,with a minimum CV value of 3546.3172. The complete set of local maximum likelihood parameter estimates obtained at each observation point, representing the 44 regencies and cities of Kalimantan in 2024, is provided in **Supplementary Table S1**.

To validate the analytical performance and superiority of the proposed MP-LLGPR model over the baseline GPR global model, goodness-of-fit and prediction accuracy were evaluated using RMSE. The final comparative results are summarized in Table 6.

A comparison based on the RMSE further highlights the improvement achieved by the proposed method (Table 6). The global GPR model produced an RMSE of 47.6230, whereas the MP-LLGPR model yielded a substantially lower RMSE of 25.7018. This represents a 46.03% reduction in RMSE following the incorporation of the localized nonparametric specification. As visually confirmed in Fig. 2, the lower RMSE of the MP-LLGPR model indicates that its fitted values are closer to the observed stunting case counts. These findings demonstrate that allowing parameter estimates to vary locally across the predictor space enables the proposed method to capture heterogeneous relationships between the response variable and multiple predictors across the observed data.

Software

The proposed MP-LLGPR model was implemented in R. The computational procedure included bandwidth selection using a grid search based on leave-one-out CV and local parameter estimation using the Newton–Raphson algorithm. Model performance was evaluated using the RMSE.

Limitations

The proposed MP-LLGPR method may become computationally demanding as the number of predictors increases because bandwidth optimization is performed in a multidimensional predictor space. In addition, its performance may be affected by the curse of

Table 4
Parameter estimates of the selected global GPR model using MLE.

Variable	Parameter	Estimate	P-value
Intercept	β_0	2.7649	< 0.001
The number of integrated community-based health service posts or posyandu ($\mathbf{X}_1$)	β_1	0.0003	0.0442
The number of nutrition health workers ($\mathbf{X}_2$)	β_2	0.0037	0.0154
The percentage of monthly per capita food expenditure ($\mathbf{X}_3$)	β_3	0.0219	0.0093
The number of low birth weight infants ($\mathbf{X}_6$)	β_6	0.0008	0.0044
Dispersion	φ	0.7351	< 0.001
AIC		438.6039	

Table 5
Optimal bandwidth and CV criterion for the MP-LLGPR model.

Variable	Optimal Bandwidth
The number of integrated community-based health service posts or posyandu ($\mathbf{X}_1$)	320
The number of nutrition health workers ($\mathbf{X}_2$)	60
The percentage of monthly per capita food expenditure ($\mathbf{X}_3$)	9
The number of low birth weight infants ($\mathbf{X}_6$)	200
Minimum CV	3546.3172

Table 6
Performance comparison between global GPR and MP-LLGPR models.

Model	RMSE
GPR global	47.6230
MP-LLGPR	**25.7018**

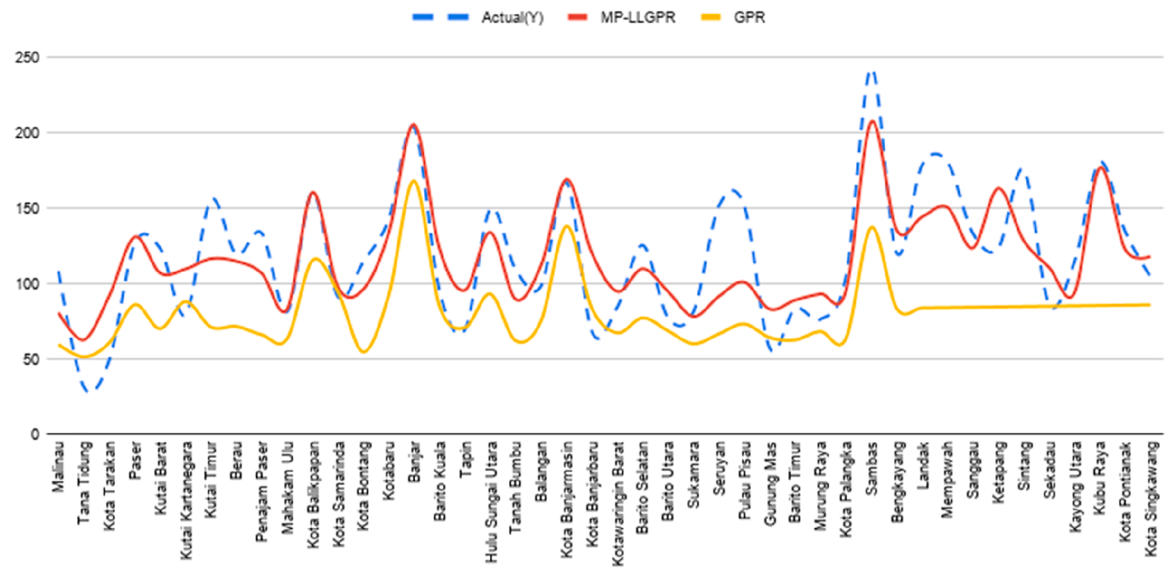


Fig. 2. Comparison of observed stunting counts and fitted mean counts from the global GPR and MP-LLGPR models.

dimensionality when applied to high-dimensional data.

Ethics statements

This research, entitled "A Multi-predictor Nonparametric Local Linear Generalized Poisson Regression Model for Overdispersed Count Data", did not involve human participants, animals, or personal data, and thus did not require formal ethical approval. The study focuses on the development and application of statistical methodologies, specifically designed to enhance generalized Poisson regression models using a local linear approach. The research adheres to the highest standards of academic integrity and transparency, with no conflicts of interest, and aims to contribute to the advancement of robust statistical frameworks for health data analysis.

CRediT author statement

Darnah: Conceptualization, Methodology, Validity test, Writing-Preparation of the first draft, and Supervision. M.Fathurahman: Conceptualization and Methodology. Iriyani Kamaruddin: Methodology and Validity test. Suyitno: Conceptualization, Methodology, and Validity test. Toha Saifudin: Methodology and Validity test. Mardiana: Writing-Review and Editing. Erni Tri Astuti: Methodology and Validity test. Ratna Kusuma: Supervision and Funding Acquisition. Rahmawati Munir: Validity test and Project Administration. Made Dirgantara: Writing-Review and Editing. Fidia Deny Tisna Amijaya: Methodology and Validity test. Imam Rofi'i: Validity test and Project Administration. Andi Syaifullah Al Maslulil Agri: Editing and Project Administration. Feni Arista Bahni: Editing and Project Administration.

Supplementary material and/or additional information [OPTIONAL]
None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, for supporting this work through the Regular Fundamental Research Grant (contract number: 238/C3/DT.05.00/PL-BARU/2026). Appreciation is also extended to the Institute for Research and Community Services (LPPM) at Mulawarman University (contract number: 607/DST/UN17.L1/PT.00.03/2026).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2026.104228.

Data availability

The authors do not have permission to share data.

References

[1] M.I.U. Darnah, N. Chamidah, Modeling of maternal mortality and infant mortality cases in East Kalimantan using Poisson Regression approach based on local linear estimator, in: IOP Conference Series: Earth and Environmental Science, 2019, https://doi.org/10.1088/1755-1315/243/1/012023.
[2] Darnah, et al., Locally kernel weighted maximum likelihood estimator for local linear multi-predictor poisson regression, MethodsX 14 (2025) 103258, https://doi.org/10.1016/j.mex.2025.103258.
[3] M. Amin, M.N. Akram, M. Amanullah, On the James-Stein estimator for the poisson regression model, Commun. Stat. - Simul. Comput. 51 (10) (2022) 5596–5608, https://doi.org/10.1080/03610918.2020.1775851.
[4] I.E. Okorie, E. Afuecheta, S. Nadarajah, A. Bright, A.C. Akpanta, A Poisson regression approach for assessing morbidity risk and determinants among under five children in Nigeria, Sci. Rep. 14 (1) (2024), https://doi.org/10.1038/s41598-024-72373-4.
[5] P. Schober, T.R. Vetter, Count data in medical research: poisson regression and negative binomial regression, Anesth. Analg. 132 (5) (2021) 1378–1379, https://doi.org/10.1213/ANE.0000000000005398.
[6] W. Iqbal, Y.M. Tang, K.Y. Chau, M. Irfan, M. Mohsin, Nexus between air pollution and NCOV-2019 in China: application of negative binomial regression analysis, Process. Saf. Environ. Prot. 150 (2021) 557–565, https://doi.org/10.1016/j.psep.2021.04.039.
[7] M. Amin, M.N. Akram, B.M.G. Kibria, A new adjusted Liu estimator for the Poisson regression model, Concurr. Comput. Pract. Exp. 33 (20) (2021), https://doi.org/10.1002/cpe.6340.
[8] N. Ismail, A.A. Jemain, Handling overdispersion with negative binomial and generalized poisson regression models, Casualty Actuar. Soc. Forum (2007) 103–158.
[9] W. Wang, F. Famoye, Modeling household fertility decisions with generalized poisson regression, J. Popul. Econ. 10 (3) (1997) 273–283, https://doi.org/10.1007/s001480050043.
[10] S. Brown, A. Heathcote, On the use of nonparametric regression in assessing parametric regression models, J. Math. Psychol. 46 (6) (2002) 716–730, https://doi.org/10.1006/jmps.2002.1421.
[11] M.I.U. Darnah, N. Chamidah, Estimation of the Bi-response Poisson regression model based on local linear approach, Int. J. Acad. Appl. Res. 3 (5) (2019) 14–18.
[12] S. Mestiri, A. Farhat, Using non-parametric count model for credit scoring, J. Quant. Econ. 19 (1) (2021) 39–49, https://doi.org/10.1007/s40953-020-00208-w.
[13] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1998, https://doi.org/10.1002/9781118625590.
[14] G. Nurhuda, Wasono, D.A. Nohe, Nonparametric regression modeling based on spline truncated estimator on simulation data, J. Mat. Stat. dan Komputasi 19 (1) (2022) 172–182, https://doi.org/10.20956/j.v19i1.21534.
[15] H.F.F. Mahmoud, Parametric versus semi and nonparametric regression models, Int. J. Stat. Probab. 10 (2) (2021) 90, https://doi.org/10.5539/ijsp.v10n2p90.
[16] I. Sriliana, I.N. Budiantara, V. Ratnasari, A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application, Symmetry 14 (12) (2022) 2687, https://doi.org/10.3390/sym14122687.
[17] M.W. Khattak, A. Pirdayani, P. De Winne, T. Brijs, H. De Backer, Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model, Accid. Anal. Prev. 151 (2021) 105964, https://doi.org/10.1016/j.aap.2020.105964.
[18] M. Al Haris, P.R. Arum, Negative binomial regression and generalized poisson regression models on the number of traffic accidents in Central Java, BAREKENG J. Ilmu Mat. dan Terap. 16 (2) (2022) 471–482, https://doi.org/10.30598/barekengvol16iss2pp471-482.
[19] R.L. Eubank, Nonparametric Regression and Spline Smoothing, Marcel Dekker, New York, 1999.
[20] G. Wahba, Improper priors, spline smoothing and the problem of guarding against model errors in regression, J. R. Stat. Soc. Ser. B 40 (3) (1978) 364–372, https://doi.org/10.1111/j.2517-6161.1978.tb01050.x.
[21] A.T.R. Dani, V. Ratnasari, I.N. Budiantara, Optimal knots point and bandwidth selection in modeling mixed estimator nonparametric regression, IOP Conf. Ser. Mater. Sci. Eng. 1115 (1) (2021) 012020, https://doi.org/10.1088/1757-899X/1115/1/012020.
[22] N. Setia, A. Islamiyati, E.T. Herdiani, Analysis of maternal mortality rate using multivariate adaptive generalized poisson regression spline, Commun. Math. Biol. Neurosci. 2025 (98) (2025) 1–18, https://doi.org/10.28919/cmbn/9312.
[23] E.T. Astuti, I.N. Budiantara, S. Sunaryo, M. Dokhi, Statistical modeling for mortality data using local generalized poisson regression model, Int. J. Appl. Math. Stat. 33 (3) (2013) 92–101.
[24] E. Ana, N. Chamidah, P. Andriani, and B. Lestari, "Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression," 2019, p. 012067. doi: 10.1088/1742-6596/1397/1/012067.

[25] L.R. Cheruiyot, Local linear regression estimator on the boundary correction in nonparametric regression estimation, J. Stat. Theory. Appl. 19 (3) (2020) 460, https://doi.org/10.2991/jsta.d.201016.001.
[26] N.C. Nidhomuddin, A. Kurniawan, Confidence interval of the parameter on multipredictor biresponse longitudinal data analysis using local linear estimator for modeling of case increase and case fatality rates COVID-19 in Indonesia: a theoretical discussion, Commun. Math. Biol. Neurosci. 2022 (2022) 1–12, https://doi.org/10.28919/cmbn/6900.
[27] N. Chamidah, Y.S. Yonani, E. Ana, B. Lestari, Identification the number of mycobacterium tuberculosis based on sputum image using local linear estimator, Bull. Electr. Eng. Informatics 9 (5) (2020) 2109–2116, https://doi.org/10.11591/eei.v9i5.2021.
[28] C. Adam, I. Gijbels, Local polynomial expectile regression, Ann. Inst. Stat. Math. 74 (2) (2022) 341–378, https://doi.org/10.1007/s10463-021-00799-y.
[29] S.E. Ahmed, D. Aydin, E. Yilmaz, Semiparametric time-series model using local polynomial: an application on the effects of financial risk factors on crop yield, J. Risk. Financ. Manage 15 (3) (2022) 141, https://doi.org/10.3390/jrfm15030141.
[30] M. Ulya, N. Chamidah, T. Saifudin, Predicting the sweetness level of avomango (Gadung Klonal 21) using multi-predictor local polynomial regression, in: IOP Conference Series: Earth and Environmental Science, 2021 012009, https://doi.org/10.1088/1755-1315/733/1/012009.
[31] M.D. Cattaneo, M. Jansson, X. Ma, Simple local polynomial density estimators, J. Am. Stat. Assoc. 115 (531) (2020) 1449–1455, https://doi.org/10.1080/01621459.2019.1635480.
[32] W. Cui, M. Wei, Strong consistency of kernel regression estimate, Open. J. Stat. 03 (3) (2013) 179–182, https://doi.org/10.4236/ojs.2013.33020.
[33] W.H. Wong, On the consistency of cross-validation in kernel nonparametric regression, Ann. Stat. 11 (4) (1983) 1136–1141, https://doi.org/10.1214/AOS/1176346327.
[34] A.G. Skretna, S. Kar, M. Piotrowska, J. Leszczynski, The kernel-weighted local polynomial regression (KwLPR) approach: an efficient, novel tool for development of QSAR/QSAAR toxicity extrapolation models, J. Cheminform. 13 (1) (2021) 9, https://doi.org/10.1186/s13321-021-00484-5.
[35] G. Wahba, Y. Wang, Spline function, Encycl. Stat. Sci. (2014) 1–27, https://doi.org/10.1002/9781118445112.stat02388.
[36] M.B. Adiguzel, M.A. Cengiz, Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria, Heliyon 9 (9) (2023) e19964, https://doi.org/10.1016/j.heliyon.2023.e19964.
[37] X. Lin, D. Zhang, Inference in generalized additive mixed models by using smoothing splines, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (2) (1999) 381–400, https://doi.org/10.1111/1467-9868.00183.
[38] M. Roozbeh, M. Maanavi, N.A. Mohamed, Penalized least squares optimization problem for high-dimensional data, Int. J. Nonlinear Anal. Appl. 14 (1) (2023) 245–250, https://doi.org/10.22075/ijnaa.2022.26998.3470.
[39] Kosmaryati, M.D. Kartikasari, Analysis of penalized semiparametric regression model on Bi-response longitudinal data, in: Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019), 2020, pp. 172–178, https://doi.org/10.2991/assehr.k.201010.025.
[40] L.N. Berry, N.E. Helwig, Cross-validation, information theory, or Maximum likelihood? A comparison of tuning methods for penalized splines, Stats 4 (3) (2021) 701–724, https://doi.org/10.3390/stats4030042.
[41] I.W. Sudiarsa, I.N. Budiantara, S.W. Purnami, Combined estimator fourier series and spline truncated in multivariable nonparametric regression, Appl. Math. Sci. 9 (100) (2015) 4997–5010, https://doi.org/10.12988/ams.2015.55394.
[42] I.U. Darnah, N. Chamidah, Bi-response Poisson regression model based on local linear for modeling effect of early marriage on the child- girl to maternal and infant mortality in East Java, in: AIP Conference Proceedings, 2022 030005, https://doi.org/10.1063/5.0104404.
[43] A. Baíllo, A. Grané, Local linear regression for functional predictor and scalar response, J. Multivar. Anal. 100 (1) (2009) 102–111, https://doi.org/10.1016/j.jmva.2008.03.008.
[44] M. Kohler, A. Krzyżak, Asymptotic confidence intervals for Poisson regression, J. Multivar. Anal. 98 (5) (2007) 1072–1094, https://doi.org/10.1016/j.jmva.2006.07.006.
[45] E.T. Astuti, I.N. Budiantara, S. Sunaryo, M. Dokhi, On the confidence band of local likelihood estimates in generalized Poisson regression model, in: The 3rd Basic Science International Conference (BaSIC), 2013, p. 2013.
[46] R.E. Walpole, R.H. Myers, S.L. Myers, and K.E. Ye, Probability & statistics for engineers & scientists ninth edition. Boston: Pearson, 2011.
[47] G. Zuo, K. Fu, X. Dai, L. Zhang, Generalized poisson hurdle model for count data and its application in ear disease, Entropy 23 (9) (2021) 1206, https://doi.org/10.3390/e23091206.
[48] Suyitno, et al., Geographically weighted Weibull regression modeling on dissolved oxygen data to analyze river water quality in East Kalimantan, MethodsX 16 (2026) 103745, https://doi.org/10.1016/j.mex.2025.103745.
[49] D. Suyitno, A.T.R. Dani, N.T. Oktavia, Modeling the hospitalization time of stroke patients at Abdul Wahab Sjahranie Hospital Samarinda using the Weibull Regression Model, MethodsX 14 (2025) 103082, https://doi.org/10.1016/j.mex.2024.103082.
[50] B. Yadav, L. Jeyaseelan, M. Sappani, T. Mani, S. George, S.I. Bangdiwala, Modeling zero-inflated count data using generalized Poisson and ordinal logistic regression models in medical research, Oman. Med. J. 39 (1) (2024) e586, https://doi.org/10.5001/omj.2024.41.