CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 870-886
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Haversine-Based Geographically Weighted Panel Regression of Human Development in Gorontalo (2016–2025)

Debora Dwi Kurniawati, Henny Pramoedyo*, Suci Astutik, and Friansyah Gani

Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Indonesia

Abstract

Spatial disparities in human development suggest that aggregate progress may mask important regional differences in how education, health, and economic conditions influence outcomes. This study analyzes spatial variation in the contributions of education, health, and expenditure to the Human Development Index (HDI) across districts and cities in Gorontalo Province, Indonesia, during 2016–2025. Balanced panel data are examined using a fixed-effects (FE) model as a global benchmark and a Geographically Weighted Panel Regression (GWPR) framework to capture spatial heterogeneity. This research constitutes an empirical application of GWPR, integrating a fixed-effects panel structure with geographically weighted estimation. Local coefficients are obtained by applying spatially weighted least squares to within-transformed panel data, allowing parameters to vary across locations while remaining constant over time. A great-circle (Haversine) distance metric and an adaptive bisquare kernel are employed to construct spatial weights, with bandwidth selected through cross-validation (35 nearest neighbours). Compared with the global FE model, GWPR demonstrates improved explanatory performance and reveals substantial spatial variation. Education exhibits the strongest positive local effect, followed by health, while expenditure shows weaker and spatially varying influence. These findings support spatially sensitive, place-based development policies.

Keywords: Human Development Index; Geographically Weighted Panel Regression; Haversine Distance; Spatial Heterogeneity; Gorontalo.

1. Introduction

The Human Development Index (HDI) is widely used as a composite indicator of human well-being, summarizing achievements in education, health, and standard of living [1]. While HDI provides an overall measure of development performance, its aggregate structure may conceal important regional differences in how each component contributes to overall human development. In Indonesia, substantial disparities in HDI persist across districts and cities, even within the same province, reflecting uneven educational attainment, health outcomes, and economic conditions [2].

Recent developments in spatial econometrics emphasize that development processes are inherently spatial and heterogeneous. Spatial heterogeneity implies that relationships among development dimensions may vary across locations rather than remain constant. As argued

*Corresponding author. E-mail: hennyp@ub.ac.id

Submitted: February 02, 2026 Reviewed: March 04, 2026 Accepted: April 03, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.40886

by [3], imposing homogeneous parameters across regions can obscure locally specific dynamics and potentially misinform policy design. This concern is particularly relevant in geographically diverse provinces, where demographic, economic, and institutional characteristics differ markedly across districts and cities.

Panel data models are frequently employed in regional development studies because they allow researchers to control for unobserved time-invariant characteristics while exploiting temporal variation. However, conventional panel regression frameworks typically assume spatial homogeneity in regression coefficients. Such an assumption may be restrictive when the relative contribution of education, health, and expenditure components to HDI differs across space.

To capture both temporal dynamics and unobserved regional characteristics, panel data modelsare widely applied in studies of human development. Panel frameworks allow researchers to control for unobserved heterogeneity while exploiting time variation in the data. However, conventional panel regression models typically assume spatial homogeneity of regression coefficients, which may be overly restrictive in the presence of strong regional heterogeneity. Recent developments in spatial panel econometrics relax this assumption by allowing spatial variation in relationships among variables, thereby providing a more realistic representation of regional development processes [4].

Because HDI is constructed from these three dimensions, this study does not interpret education, health, and expenditure as external causal determinants. Instead, the analysis is framed as an assessment of spatially varying contributions of local components within the HDI structure. The objective is to examine how the relative strength of each dimension differs across districts and cities, thereby revealing local development profiles that global-average specifications may mask.

To accommodate both spatial heterogeneity and temporal structure, this study employs Geographically Weighted Panel Regression (GWPR). This method extends geographically weighted regression into a panel framework, allowing regression coefficients to vary spatially while accounting for time effects [5]. Spatial relationships are defined using the Haversine distance, which accounts for Earth's curvature and provides more accurate interregional distance measurement than planar approximations. By explicitly specifying a distance-based weighting scheme with an adaptive kernel, the model captures realistic geographic proximity among districts and cities.

Although spatial panel approaches have been applied in human development research [6, 7], clear empirical evaluation of spatial heterogeneity and model adequacy remains essential. In this study, spatial heterogeneity is assessed by examining the variability of local coefficient estimates across districts and cities and by testing the statistical significance of local parameters using location-specific t-statistics. Model selection within the panel framework is conducted using the Chow and Hausman tests to determine the appropriate global benchmark specification. Furthermore, model fit and bandwidth selection in the Geographically Weighted Panel Regression framework are evaluated using a cross-validation criterion, with the optimal adaptive bandwidth chosen as the one that minimizes the cross-validation score. By combining specification testing, local parameter inference, and cross-validation-based model comparison, this study provides a systematic assessment of spatial heterogeneity and model performance.

Motivated by these considerations, this study analyzes spatial variation in the local contributions of education, health, and expenditure components to HDI across districts and cities in Gorontalo Province during 2016–2025 using a Haversine-distance-based GWPR approach. By integrating panel structure, spatial heterogeneity, and an explicit distance specification, this research provides empirical evidence to support region-sensitive, place-based human development policy formulation.

2. Methods

This study employs secondary data from Statistics Indonesia (*Badan Pusat Statistik*, BPS) for Gorontalo Province, comprising balanced panel data at the district and city levels for the period 2016–2025. All regencies and municipalities in Gorontalo Province are included to ensure comprehensive spatial coverage. The dependent variable (Y) is the Human Development Index (HDI), which reflects achievements in education, health, and living standards. In contrast, the explanatory variables represent key socioeconomic factors associated with human development, including mean years of schooling (X_1), life expectancy at birth (X_2), and real per capita expenditure (X_3). These variables capture human capital accumulation, health conditions, and economic capacity, which are commonly linked to variations in human development across regions. All variables are sourced from official BPS publications to ensure data reliability, consistency across regions, and comparability over time.

Table 1: Research Variables

Variable	**Symbol**	**Unit**	**Source (Year)**
Human Development Index (HDI)	Y	Index (0–100)	BPS (2016–2025)
Mean Years of Schooling	X_1	Years	BPS (2016–2025)
Life Expectancy at Birth	X_2	Years	BPS (2016–2025)
Real Per Capita Expenditure	X_3	Rupiah	BPS (2016–2025)

The analytical procedure using the Haversine distance-driven Geographically Weighted Panel Regression (GWPR) in this study is carried out through the following steps:

1. **Global panel model estimation**: A global fixed-effects panel regression model is first estimated as specified in Eq. (6) to examine the overall relationship between the Human Development Index (HDI) and the explanatory variables across all districts and time periods. This model serves as a benchmark for evaluating the spatially varying specification defined in Eq. (9).
2. **Spatial data preparation**: Geographic coordinates for each district and city in Gorontalo Province are obtained from the centroids of their administrative boundaries. These coordinates correspond to (u_i, v_i) in Eq. (9) and serve as spatial references for modeling location-specific relationships.
3. **Spatial distance computation**: Pairwise spatial distances d_{ij} between districts and cities are computed using the Haversine distance defined in Eq. (10). The resulting distance matrix provides the basis for constructing spatial weights in the GWPR framework.
4. **Construction of spatial weighting matrix**: An adaptive kernel function is applied to the Haversine distance matrix to construct the diagonal spatial weighting matrix $\mathbf{W}(u_i, v_i)$ used in the local weighted least squares estimator defined in Eq. (11). This procedure assigns larger weights to geographically closer observations while allowing the effective number of neighbors to vary across locations.
5. **Bandwidth selection**: The optimal bandwidth is selected using a cross-validation (CV) criterion defined as
$$b^* = \arg\min_b \text{CV}(b),$$
where the chosen bandwidth minimizes the CV score and determines the spatial weighting structure in Eq. (11).
6. **Local GWPR estimation**: The GWPR model specified in Eq. (9) is estimated using the weighted least squares estimator given in Eq. (11) to obtain location-specific regression coefficients for each district and city.
7. **Model evaluation and interpretation**: Model performance is evaluated by comparing the GWPR results in Eq. (9) with the global fixed-effects panel model in Eq. (6) using AICc and goodness-of-fit measures. The spatial distribution of local parameter estimates is subsequently analyzed to identify regional variations in the determinants of HDI.

2.1. Spatial Test

2.1.1. Multicollinearity Diagnostic

To evaluate multicollinearity among the explanatory variables, this study employs the Variance Inflation Factor (VIF), a widely used diagnostic for detecting collinearity in multiple regression models [8]. The VIF measures the extent to which the variance of an estimated regression coefficient is inflated due to linear dependence among the explanatory variables. For a given predictor X_k, The VIF is computed according to Eq. (1).:

$$\mathrm{VIF}(X_k) = \frac{1}{1 - R_k^2}, \tag{1}$$

where R_k^2 denotes the coefficient of determination from the regression of X_k on the remaining explanatory variables. A VIF value exceeding 10 signals the presence of serious multicollinearity, whereas values below this cutoff indicate that multicollinearity is not problematic. This diagnostic confirms that the predictors incorporated in the model are not highly correlated with one another.

2.1.2. Spatial Autocorrelation Test Using Moran's I

Spatial dependence can be evaluated using *Moran's I*, a widely applied global spatial statistic designed to detect spatial clustering, dispersion, or random patterns among observations [9]. In regression analysis, Moran's I is commonly applied to model residuals to assess whether the assumption of independence is violated due to spatial autocorrelation.

In global panel regression frameworks, Moran's I is typically used to examine potential violations of cross-sectional independence arising from spatial interactions among regions. In geographically weighted models, including Geographically Weighted Panel Regression (GWPR) and Moran's I may likewise be applied to locally estimated residuals to evaluate whether the spatial weighting structure adequately captures spatial heterogeneity and mitigates residual spatial dependence.

Moran's I is defined as:

$$I = \frac{n}{\mathbf{W}} \cdot \frac{\sum_i \sum_j w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{\sum_i (e_i - \bar{e})^2}, \tag{2}$$

where n denotes the number of spatial units, w_{ij} represents the spatial weight between locations i and j, e_i is the residual at location i, $\bar{e}$ is the mean residual, and $W = \sum_i \sum_j w_{ij}$ denotes the sum of all spatial weights.

Statistical inference is conducted using the standardized test statistic:

$$Z_I = \frac{I - \mathbb{E}[I]}{\sqrt{\mathrm{Var}(I)}}, \tag{3}$$

where $\mathbb{E}[I]$ and $\mathrm{Var}(I)$ denote the expected value and variance of Moran's I under the null hypothesis of spatial randomness. A statistically insignificant result indicates the absence of residual spatial autocorrelation.

Within the locally weighted regression framework, spatial dependence is explicitly incorporated through a kernel-based spatial weighting matrix that assigns greater influence to geographically proximate observations. Consequently, insignificant residual spatial autocorrelation suggests that the local estimation procedure has effectively accounted for spatial heterogeneity.

2.1.3. Breusch–Pagan Heteroscedasticity Test

To evaluate whether the variance of the residuals is constant, the The Breusch-Pagan *(BP)* test was applied. This test detects heteroskedasticity by regressing the squared residuals on the independent variables [10]. The BP statistic is defined as:

$$\mathrm{BP} = \sum_{i=1}^{n} \hat{u}_i^2 \mathbf{z}_i^\top \left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{z}_i. \tag{4}$$

where $\hat{u}_i$ denotes the residuals of the regression model and Z represents the matrix of independent variables. Under the null hypothesis of homoskedasticity, the BP statistic follows a chi-square distribution with degrees of freedom equal to the number of predictors. A p-value greater than 0.05 indicates acceptance of the null hypothesis and suggests that the model is not heteroskedastic.

2.2. Penel Regression

Panel regression is a modeling framework for analyzing observations collected across multiple units and over time, allowing researchers to account for both cross-sectional heterogeneity and temporal dynamics. Recent econometric literature highlights that panel data models offer greater flexibility, mitigate multicollinearity, and improve the identification of causal effects by exploiting information within and between units [11]. The general specification is as follows:

$$y_{it} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{kit} + \varepsilon_{it}, \tag{5}$$

where $i = 1, \ldots, N$ denotes cross-sectional units and $t = 1, \ldots, T$ denotes time periods. Panel regression is appropriate when unobserved individual characteristics influence the dependent variable, because fixed-effects estimation can absorb such unit-specific heterogeneity. Indeed, even in recent spatial-dynamic and nonparametric panel models, fixed-effects specifications remain widely employed to control for unobserved time-invariant (or slowly varying) factors across units [12]. To address this issue, the Fixed-Effects Model (FEM) is widely used because it effectively controls for unobserved time-invariant factors.

2.2.1. Fixed Effect Model (FEM)

Fixed Effects Model (FEM) allows each cross-sectional unit to have its own intercept α_i, capturing unobserved time-invariant characteristics that may correlate with the explanatory variables. The FEM specification is:

$$y_{it} = \alpha_i + \sum_{k=1}^{p} \beta_k x_{kit} + \varepsilon_{it}. \tag{6}$$

Within (demeaning) transformation.
To eliminate α_i, we apply the within transformation:

$$y^*_{it} = y_{it} - \bar{y}_i, \qquad x^*_{kit} = x_{kit} - \bar{x}_{ki},$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \bar{x}_{ki} = \frac{1}{T}\sum_{t=1}^{T} x_{kit}.$$

The transformed model becomes:

$$y^*_{it} = \sum_{k=1}^{p} \beta_k x^*_{kit} + \varepsilon^*_{it}.$$

Stacking the transformed observations over all units and time periods yields

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*.$$

The fixed-effects (within) estimator is given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{FE}} = (\mathbf{X}^{*\top}\mathbf{X}^*)^{-1}\mathbf{X}^{*\top}\mathbf{y}^*.$$

with $\mathbf{X}^* \in \mathbb{R}^{NT\times p}$ and $\mathbf{y}^* \in \mathbb{R}^{NT\times 1}$. and $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ denote the stacked within-transformed regression and the dependent variable across all units. Fixed-Effects Model remains consistent even when individual effects correlate with regressors, making it the preferred choice in empirical applications with substantial heterogeneity between units [13].

2.3. Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is a local regression technique used to model spatially varying relationships between a response variable and predictor variables [14]. Unlike global linear regression, which assumes that model parameters are constant across all locations, GWR allows regression coefficients to vary spatially, enabling the model to capture local heterogeneity [15].

Let y_i denote the response variable at location i, and let (u_i, v_i) represent the geographical coordinates of that location. The GWR model is expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ki} + \varepsilon_i, \tag{7}$$

where $\beta_k(u_i, v_i)$ are location-specific regression coefficients and ε_i is the random error term.

2.3.1. Parameter Estimation of GWR

Local parameter estimates are obtained using weighted least squares [16]. The estimator of the local regression coefficients at location (u_i, v_i) is given by:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left(\mathbf{X}^\top \mathbf{W}(u_i, v_i)\mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W}(u_i, v_i)\mathbf{y}. \tag{8}$$

In this formulation,

$$\mathbf{X} \in \mathbb{R}^{n\times p}$$

denotes the matrix of explanatory variables, where n represents the number of spatial observations and p denotes the number of predictors. The response variable is represented by

$$\mathbf{y} \in \mathbb{R}^{n\times 1},$$

while

$$\mathbf{W}(u_i, v_i) \in \mathbb{R}^{n\times n}$$

is a diagonal spatial weighting matrix whose elements w_{ij} represent the spatial weights assigned to observations based on their geographic distance from location (u_i, v_i). The spatial weighting structure ensures that nearby observations exert greater influence on the estimation of the local parameters than more distant observations.

2.3.2. Bandwidth Selection in GWR

The bandwidth parameter b determines the spatial extent over which observations influence the estimation of local parameters in the GWR model. A small bandwidth produces highly localized estimates but may lead to unstable coefficients due to limited information, whereas a large bandwidth results in smoother estimates but may obscure local spatial variation.

To obtain reliable estimates, the bandwidth is typically selected using a data-driven optimization criterion such as the Akaike Information Criterion corrected (AICc) or the Cross-Validation (CV) score. The optimal bandwidth is defined as the value that minimizes the selected criterion:

$$b^* = \arg\min_b CV(b),$$

where $CV(b)$ represents the cross-validation score evaluated at bandwidth b. This procedure balances the trade-off between model bias and variance, ensuring that the spatial weighting scheme appropriately captures local spatial relationships.

2.4. Geographically Weighted Panel Regression (GWPR)

Geographically Weighted Panel Regression (GWPR) extends the classical Fixed Effects Model (FEM) by allowing slope parameters to vary across geographic regions while controlling for unobserved unit-specific and time-invariant heterogeneity. This integration of a panel-data structure with geographically weighted estimation enables the capture of spatially heterogeneous relationships that vary across time and location—an advantage over traditional global panel models in empirical regional analyses [17]. For example, applications of GWPR in socioeconomic contexts, such as poverty modeling and regional development, have demonstrated its ability to reveal local variations in variable effects that global models cannot capture.

Formally, the GWPR model is specified as:

$$y_{it} = \alpha_i + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, x_{kit} + \varepsilon_{it}, \tag{9}$$

where y_{it} is the dependent variable for spatial unit i at time t, (u_i, v_i) denote the geographic coordinates of unit i, α_i captures the unit-specific fixed effect, $\beta_k(u_i, v_i)$ are spatially varying slope coefficients for explanatory variables x_{kit}, and ε_{it} is the idiosyncratic error term. In GWPR, the fixed-effects transformation removes α_i before local estimation, isolating spatial variation in the slope coefficients without confounding by unobserved unit heterogeneity [18].

This formulation can be viewed as an extension of the classical Geographically Weighted Regression (GWR), in which local parameters are estimated using spatially weighted least squares. In the GWPR framework, the within transformation is first applied to remove the fixed effect, after which spatial weighting is imposed on the transformed observations during local parameter estimation. Consequently, the spatial weighting matrix operates on the within-transformed panel data, linking the GWR estimator to the panel-data structure.

2.4.1. Haversine Distance for Distance-Based Spatial Weighting

In the GWPR framework, spatial heterogeneity is captured through a location-specific weighting scheme, where geographically closer observations receive larger weights via a distance-decay kernel. Recent spatial econometric work emphasizes that the specification of distance-based weight matrices and their decay structure can materially affect inference and the interpretation of spillover reach [19, 20].

To operationalize geographic proximity among districts and cities in Gorontalo Province, this study computes inter-regional distances using the Haversine formula, which measures great-circle distance on the Earth's surface using latitude–longitude coordinates. The use of Haversine distance as a geographically meaningful distance metric is common in applied spatial modeling and optimization settings that require realistic inter-location distance measurement [21].

Compared with planar Euclidean distance, the Haversine metric accounts for the Earth's curvature and therefore provides a more accurate measure of inter-regional proximity when geographic coordinates are expressed in latitude and longitude. This is particularly relevant in regional studies covering dispersed administrative units, where projection-based distortions may affect distance measurement. By providing consistent great-circle distances, the Haversine specification contributes to greater stability in the construction of the spatial weighting matrix and reduces the risk of biased parameter estimates arising from distance mismeasurement.

The Haversine distance between two regions i and j is defined as:

$$d_{ij} = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_i - \phi_j}{2}\right) + \cos(\phi_i)\cos(\phi_j)\sin^2\left(\frac{\lambda_i - \lambda_j}{2}\right)}\right), \tag{10}$$

where R denotes the Earth's radius, ϕ_i and ϕ_j are the latitudes of regions i and j, and λ_i and λ_j are their longitudes (in radians). The resulting Haversine distance matrix is then used to construct the kernel-based spatial weighting matrix for GWPR estimation.

2.4.2. Parameter Estimation of GWPR

Parameter estimation in GWPR is achieved using a locally weighted least squares (LWLS) approach applied to the fixed-effects–transformed panel data. After the within transformation, a spatial weighting matrix $\mathbf{W}(u_i, v_i)$ is constructed for each location i, where diagonal elements reflect the geographic proximity between the focal area and other spatial units. The local parameter estimator is obtained by minimizing the spatially weighted sum of squared residuals at location (u_i, v_i), yielding the following weighted least squares solution:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left(\tilde{\mathbf{X}}^\top \mathbf{W}(u_i, v_i)\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \mathbf{W}(u_i, v_i)\tilde{\mathbf{y}}, \tag{11}$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{X}}$ denote the dependent variable and regressor matrix after within transformation. This estimator applies weighted least squares locally at each geographic location, consistent with the approach used in recent GWPR applications where local coefficients are obtained through spatial weights that decline with distance.

The spatial weights in GWPR are typically derived from kernel functions such as Gaussian or bisquare kernels, with a bandwidth parameter controlling the degree of spatial smoothing. Practical implementations commonly select the bandwidth using data-driven criteria, such as cross-validation (CV), to balance estimation bias and variance. These weighted kernel functions enable GWPR to adaptively emphasize nearby observations, thereby capturing local spatial variation in relationships among observations.

By estimating location-specific coefficients as functions of spatial position, GWPR explicitly models spatial heterogeneity in regression relationships while retaining the advantages of panel data structures, enabling more nuanced inference for processes that vary both across space and over time.

2.4.3. Bandwidth Selection in GWPR

In Geographically Weighted Panel Regression (GWPR), the bandwidth parameter controls the spatial scale over which local parameter estimates are computed, and it plays a crucial role in balancing bias and variance in the model. An inappropriate choice of bandwidth may lead to over-smoothing, which masks local spatial variation, or to under-smoothing, which produces unstable coefficients.

To select an appropriate bandwidth, this study adopts an information-theoretic approach inspired by recent methodological developments in the GWR literature. Specifically, the optimal bandwidth b^* is defined as the value that minimizes an information complexity criterion, which balances model fit with model complexity:

$$b^* = \arg\min_b \ \mathrm{IC}(b). \tag{12}$$

Here, $\mathrm{IC}(b)$ represents an information criterion evaluated at bandwidth b, such as the corrected Akaike Information Criterion (AICc) or similar. This formulation aligns with the strategy proposed in *Bandwidth Selection in Geographically Weighted Regression Models via Information Complexity Criteria*, which uses an information complexity measure to identify a bandwidth that yields superior model performance by accounting for both goodness-of-fit and complexity simultaneously [22].

Once the optimal bandwidth is determined, it is incorporated into the spatial weighting matrix used for local estimation of regression coefficients:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = \left(\mathbf{X}^\top \mathbf{W}(u_i, v_i; b)\,\mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{W}(u_i, v_i; b)\,\mathbf{y}, \tag{13}$$

where $\mathbf{W}(u_i, v_i; b)$ is a diagonal matrix of spatial weights dependent on the distance between observations and the selected bandwidth. By determining the bandwidth using an information criterion, the model adaptively identifies the spatial scope that best captures local relationships in the data, thereby enhancing the interpretability and robustness of local parameter estimates [22].

3. Results and Discussion

This section presents the empirical findings obtained from the analysis. The discussion begins with exploratory data diagnostics to examine the distribution of the variables and detect potential outliers. Subsequently, multicollinearity and panel model specification tests are conducted to determine the appropriate global model. Finally, the Geographically Weighted Panel Regression (GWPR) model is estimated to investigate spatial heterogeneity in the determinants of the Human Development Index across districts and cities in Gorontalo Province.

3.1. Outlier

To examine potential extreme observations in the dataset, boxplots were constructed for all variables included in the analysis. These plots provide a visual summary of the distribution and allow identification of potential outliers before the regression analysis is conducted.

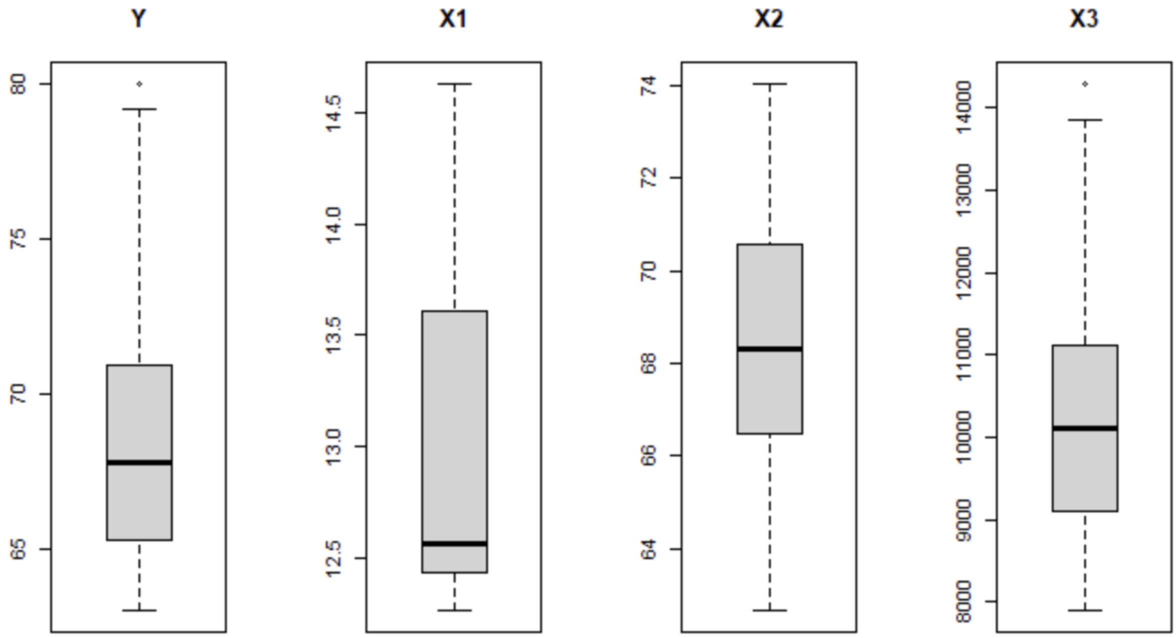


Fig. 1: Outlier

The boxplots of the four variables (Y, X_1, X_2, and X_3) illustrate their distribution and variability. Variable Y has a median around 67, X_1 around 12.5–13, X_2 around 68, and X_3 near 10,000, with respective interquartile ranges reflecting their dispersion. All variables show symmetric distributions with no significant outliers, indicating that the data are relatively consistent and free from extreme values. These results provide a clear understanding of the data distribution, serving as a basis for further statistical analysis.

3.2. Multicollinearity Test

Multicollinearity is assessed using the Variance Inflation Factor (VIF). A VIF value greater than 10 generally indicates severe multicollinearity, while values between 5 and 10 suggest moderate multicollinearity.

Table 2: Variance Inflation Factor (VIF) Results

Variable	VIF
X_1	3.228326
X_2	2.426777
X_3	2.669934

Interpretation. The Variance Inflation Factor (VIF) values for the independent variables X_1, X_2, and X_3 are 3.228, 2.427, and 2.670, respectively.Since all values are below the commonly used threshold of 10, there is no indication of problematic multicollinearity in the global panel model. This suggests that the independent variables are not strongly linearly correlated with one another; therefore, the regression coefficients can be interpreted reliably without concern for collinearity-induced distortion.

Multicollinearity diagnostics were evaluated for the global specification. Given that the same set of explanatory variables is used in the GWPR estimation,the absence of severe multicollinearity in the global model supports the stability of the local parameter estimates.

3.3. Panel Data Model Selection

To identify the most appropriate panel data specification, both the Chow test and the Hausman test were applied. The Chow test is used to compare the Common Effects Model (CEM) and the Fixed Effects Model (FEM), whereas the Hausman test is used to choose between the Fixed Effects Model (FEM) and the Random Effects Model (REM).

Table 3: Results of Chow and Hausman Tests

Test	Statistic	Value	p-value
Chow (CEM vs FEM)	F	12.456	6.449×10^{-8}
Hausman (FEM vs REM)	χ^2	39.77	1.192×10^{-8}

As shown in Table 3, both tests strongly reject the null hypothesis at the 5% significance level. The Chow test indicates the presence of significant individual effects, implying that the pooled model is inappropriate and that a fixed-effects model should be used. Furthermore, the Hausman test indicates that the Random Effects Model is inconsistent, confirming that the Fixed Effects Model is the preferred specification.

Overall, these findings demonstrate that unobserved heterogeneity across cross-sectional units plays a crucial role in explaining variations in the dependent variable. By controlling for time-invariant individual characteristics, the Fixed Effect Model provides more reliable and unbiased estimates of the effects of X_1, X_2, and X_3 on Y. Therefore, all subsequent analyses and discussions in this study are based on the Fixed Effect Model.

3.4. Residual Assumption Tests

After selecting the Fixed Effects Model (FEM) as the preferred panel-data specification, residual diagnostic tests were conducted to assess the model's adequacy. In the context of panel data estimation, the consistency of the fixed-effects estimator does not require normality of the residuals. However, diagnostic tests for normality, serial correlation, and heteroskedasticity are performed to evaluate the validity of statistical inference and the reliability of hypothesis testing.

Table 4: Summary of Residual Diagnostic Tests

Test	Statistic	Value	p-value
Shapiro–Wilk (Normality)	W	0.98671	0.7586
Wooldridge (Serial Correlation)	F	43.269	2.33×10^{-8}
Breusch–Pagan (Heteroskedasticity)	BP	23.608	3.016×10^{-5}

3.4.1. Normality of Residuals

The normality of the residuals was examined using the Shapiro–Wilk test. As shown in Table 4, the test yields a statistic of $W = 0.98671$ with a p-value of 0.7586. Since the p-value is greater than the 5% significance level, the null hypothesis of normality cannot be rejected. This indicates that the residuals from the Fixed-Effects Model are normally distributed.

3.4.2. Serial Correlation

Serial correlation in the FEM residuals was tested using Wooldridge's test for autocorrelation in fixed-effects panel models. The result shows an F-statistic of 43.269 with a p-value of 2.33×10^{-8}. As the p-value is far below 0.05, the null hypothesis of no serial correlation is rejected. This finding indicates that the residuals exhibit significant autocorrelation.

3.4.3. Heteroskedasticity

The presence of heteroskedasticity was examined using the Breusch–Pagan test. The test produces a BP statistic of 23.608 with a p-value of 3.016×10^{-5}. Since the p-value is smaller than the 5% significance level, the null hypothesis of homoskedasticity is rejected. This result indicates that the residual variance is not constant across observations, implying heteroskedasticity in the FEM.

Although serial correlation and heteroskedasticity were detected in the global FEM specification, the subsequent analysis employs the GWPR framework, which allows for spatially varying relationships and localized estimation. This approach reduces reliance on global homogeneity assumptions and provides more flexible modeling of spatial variability.

3.5. Geographically Weighted Panel Regression (GWPR)

Prior to estimating the *Geographically Weighted Panel Regression* (GWPR) model, the data are first transformed using the *within estimator* (demeaning) approach. This transformation is achieved by subtracting each variable's individual time-series mean to eliminate time-invariant individual effects.

3.5.1. Haversine Distance Matrix

The Haversine distance is a metric for measuring the shortest distance between two points on Earth's surface, given their latitude and longitude, while accounting for Earth's curvature. Unlike the Euclidean distance, which assumes a flat surface, the Haversine distance is better suited for spatial analysis across geographically distributed regions.

In the context of *Geographically Weighted Panel Regression* (GWPR), the Haversine distance is employed to construct the spatial distance matrix among all observational units (districts/cities). This matrix serves as the basis for defining spatial weights in the kernel function used to estimate local regression parameters.

The Haversine distance matrix is computed using the longitude and latitude coordinates of each region and implemented in the R software environment. The resulting distance matrix is reported in Table 5, while the complete matrix is provided in Appendix Y.

Table 5: Haversine Distance Matrix Between Districts/Cities

	1	2	3	4	5	6
1	0.00	103136.53	14967.27	16545.33	78509.72	85207.47
2	103136.53	0.00	116173.42	110739.73	24812.89	188328.86
3	14967.27	116173.42	0.00	23815.84	91835.85	72623.38
4	16545.33	110739.73	23815.84	0.00	85941.98	80517.32
5	78509.72	24812.89	91835.85	85941.98	0.00	163715.81
6	85207.47	188328.86	72623.38	80517.32	163715.81	0.00

Based on Table 5, the diagonal elements are zero, indicating that the distance of each region to itself is zero. The off-diagonal elements vary across region pairs, reflecting differences in their geographic locations. Since the geographic coordinates are time-invariant, the Haversine distance matrix remains the same across all time periods in the panel dataset.

This distance matrix is subsequently used to construct the spatial weighting matrix in the GWPR model with an adaptive kernel function. The use of the Haversine distance enhances spatial weighting accuracy by accounting for the Earth's curvature when coordinates are expressed in latitude and longitude. This reduces potential distortion from planar distance approximations and ensures more geographically consistent kernel weights. Given the spatial scale of the study area, the Haversine metric was considered more appropriate than alternative planar measures for constructing the spatial weighting matrix.

3.5.2. Kernel Weighting Function and Optimal Bandwidth (Haversine)

In Geographically Weighted Panel Regression (GWPR), the choice of bandwidth controls the balance between local model flexibility and overall smoothness. An optimal bandwidth is selected using the Cross-Validation (CV) criterion, where the best bandwidth corresponds to the minimum CV value.

In this study, bandwidth selection is conducted using an adaptive *bisquare kernel* with the Haversine distance matrix. Several candidate bandwidths are evaluated, and the optimal value is determined based on the lowest CV score.

Table 6: CV Scores and Candidate Adaptive Bandwidths (Haversine Distance)

CV Score	Bandwidth (Nearest Neighbours)
2.168208	44
2.001027	**35**
2.164795	28
2.001027	37

Based on Table, the minimum CV value of 2.001027 is achieved at an adaptive bandwidth of 35 nearest neighbours. This bandwidth is therefore selected as the optimal value for the GWPR model using Haversine distance. Because the Haversine distance matrix is time-invariant, the spatial weighting matrix is identical for all periods. Consequently, the local coefficient estimates vary across locations but remain comparable over time.

The selected bandwidth was compared with nearby candidate values, and the resulting local coefficient patterns remained substantively stable, indicating robustness to minor bandwidth variation. Because the spatial weights are derived from time-invariant geographic coordinates, the weighting structure remains constant across periods, ensuring temporal comparability while capturing spatial heterogeneity across locations.

3.5.3. GWPR Equation and Interpretation for 2025

The Geographically Weighted Panel Regression (GWPR) model was estimated using the Weighted Least Squares (WLS) method with an adaptive bisquare kernel and a Haversine distance matrix. Since the spatial weight matrix is time-invariant, the local parameter estimates remain constant over time for each location. Thus, the following system represents the GWPR equations for all districts/cities in 2025.

Boalemo (2025)

$$\hat{Y}_{2025} = 3.36212 + 1.589256X_{1,2025} + 0.411065X_{2,2025} + 0.001614X_{3,2025} + \varepsilon_{2025} \tag{14}$$

Bone Bolango (2025)

$$\hat{Y}_{2025} = -44.2860 + 3.058633X_{1,2025} + 1.035924X_{2,2025} + 0.000195X_{3,2025} + \varepsilon_{2025} \tag{15}$$

Gorontalo (2025)

$$\hat{Y}_{2025} = 3.32995 + 1.586533X_{1,2025} + 0.411542X_{2,2025} + 0.001617X_{3,2025} + \varepsilon_{2025} \tag{16}$$

Gorontalo Utara (2025)

$$\hat{Y}_{2025} = 3.38137 + 1.592084X_{1,2025} + 0.410828X_{2,2025} + 0.001610X_{3,2025} + \varepsilon_{2025} \tag{17}$$

Kota Gorontalo (2025)

$$\hat{Y}_{2025} = -43.6963 + 3.232783X_{1,2025} + 0.991933X_{2,2025} + 0.000203X_{3,2025} + \varepsilon_{2025} \quad (18)$$

Pohuwato (2025)

$$\hat{Y}_{2025} = -3.77430 + 1.873136X_{1,2025} + 0.494075X_{2,2025} + 0.001388X_{3,2025} + \varepsilon_{2025} \quad (19)$$

The system of equations shows that the relationship between the dependent variable and its predictors differs across districts and cities in 2025. In all locations, X_1 has the largest positive coefficient, indicating that it is the dominant factor influencing Y. The variable X_2 also contributes positively, although with smaller magnitudes. The coefficients of X_3 are positive but very small, suggesting that its effect on Y is relatively weak compared to X_1 and X_2.

These spatially varying coefficients confirm the presence of spatial heterogeneity and justify using GWPR rather than a global panel regression model. The equations are reported for 2025 as a representative year. Given the time-invariant spatial weights and fixed-effects structure, the local parameter estimates remain stable across periods, with temporal variation reflected in the explanatory variables.

3.5.4. Local Parameter Significance Test of the GWPR Model

The significance of the local parameters in the Geographically Weighted Panel Regression (GWPR) model was evaluated using the t-statistic at each district/city and year. A parameter is considered statistically significant at the 5% significance level if $|t| > 1.96$.

Table 7 presents the local t-statistics and significance results for all districts/cities from 2016 to 2025. The results indicate that mean years of schooling (X_1) and life expectancy at birth (X_2) are statistically significant across all districts and cities, while real per capita expenditure (X_3) is not significant in Gorontalo City.

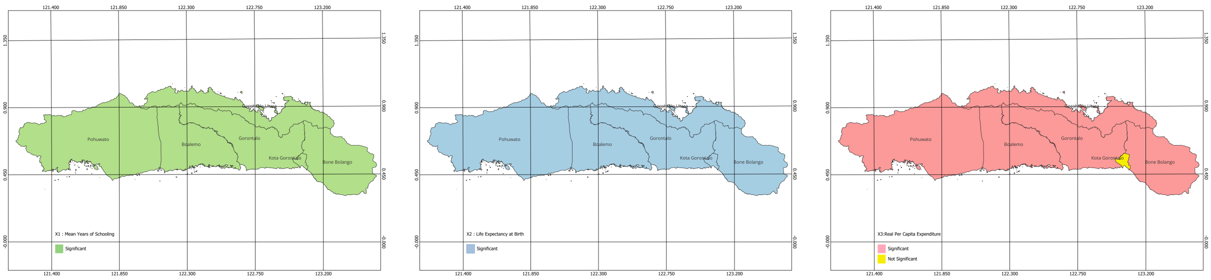
Table 7: Local Parameter Significance Test in the GWPR Model for Gorontalo Province

ID	Regency/City	Variable	Significance
1	Boalemo	X_1	Significant
		X_2	Significant
		X_3	Significant
2	Bone Bolango	X_1	Significant
		X_2	Significant
		X_3	Significant
3	Gorontalo	X_1	Significant
		X_2	Significant
		X_3	Significant
4	North Gorontalo	X_1	Significant
		X_2	Significant
		X_3	Significant
5	Gorontalo City	X_1	Significant
		X_2	Significant
		X_3	Not Significant
6	Pohuwato	X_1	Significant
		X_2	Significant
		X_3	Significant

Table 7 presents the results of the local parameter significance tests for the Geographically Weighted Panel Regression (GWPR) model across the six districts and cities in Gorontalo Province. The results show that mean years of schooling (X_1) and life expectancy at birth (X_2) are statistically significant in all regions, indicating that education and health consistently contribute to variations in the Human Development Index (HDI).

In contrast, real per capita expenditure (X_3) is not statistically significant in Gorontalo City, suggesting that the role of economic capacity in explaining HDI may vary across spatial contexts. This finding indicates that the influence of certain socioeconomic variables may differ between urban and non-urban areas.

To further illustrate the spatial heterogeneity captured by the GWPR model, Fig. 2 presents the spatial distribution of the estimated local regression coefficients for the three explanatory variables. The maps provide a visual representation of how the effects of education, health, and expenditure on HDI vary across districts and cities in Gorontalo Province.



(a) Local Coefficients of X_1 (Mean Years of Schooling) **(b)** Local Coefficients of X_2 (Life Expectancy at Birth) **(c)** Local Coefficients of X_3 (Real Per Capita Expenditure)

Fig. 2: Spatial Distribution of Local GWPR Coefficients for Gorontalo Province, 2016–2025

Fig. 2 indicates clear spatial variation in the magnitude of the estimated coefficients across the study area. The influence of mean years of schooling (X_1) appears relatively strong in several districts, highlighting the importance of education in improving human development outcomes.

Similarly, life expectancy at birth (X_2) demonstrates a positive and consistent contribution across locations, reflecting the role of health conditions in shaping regional development. In contrast, the effect of real per capita expenditure (X_3) is comparatively smaller and shows greater spatial variability, supporting the earlier finding that its statistical significance differs across regions.

4. Conclusion

This study investigates spatial heterogeneity in human development across districts and cities in Gorontalo Province during the period 2016–2025 using a Haversine distance-driven Geographically Weighted Panel Regression (GWPR) approach. By integrating panel data techniques with geographically weighted estimation, this study provides a more nuanced understanding of how socioeconomic determinants influence the Human Development Index (HDI) across space and over time.

The results reveal substantial spatial variation in the determinants of HDI across Gorontalo Province. Mean years of schooling (X_1) consistently exhibits the largest positive effect on HDI in all districts and cities, indicating that education remains the most influential driver of human development at the local level. Life expectancy at birth (X_2) also shows a positive and statistically significant contribution across locations, highlighting the importance of health conditions in shaping human development outcomes. In contrast, real per capita expenditure (X_3) has a relatively small effect and is not statistically significant in Gorontalo City, suggesting that the role of economic capacity in improving HDI may differ across spatial contexts.

The use of a Haversine distance-based weighting scheme, combined with an adaptive bisquare kernel, enables the GWPR model to capture spatial interactions more accurately by accounting for Earth's curvature and realistic geographic proximity among regions. Compared to the global Fixed Effects Model, the GWPR framework provides richer local insights by revealing spatially varying relationships that would otherwise be obscured under the assumption of parameter homogeneity.

In terms of model performance, the GWPR specification shows improved explanatory capability relative to the global FEM, as indicated by better model fit and cross-validation results reported in the estimation section. This supports the relevance of allowing spatially varying parameters in modeling HDI determinants.

The use of the Haversine-based weighting scheme is primarily conceptually motivated, as it accounts for Earth's curvature when coordinates are expressed in latitude and longitude. The present study does not conduct a formal empirical comparison across alternative distance metrics.

Overall, the findings underscore the importance of adopting place-based and region-sensitive development policies in Gorontalo Province. Policy interventions aimed at improving human development should prioritize education and health improvements while considering local socioeconomic characteristics and spatial heterogeneity. The methodological framework employed in this study can also be extended to other regions to support spatially informed development planning and evaluation.

"The dataset and code analyzed during the current study are publicly available in the Badan Pusat Statistik Gorontalo 2025[1]."

CRediT Authorship Contribution Statement

Debora Dwi Kurniawati: Conceptualization, methodology, writing—original draft. **Friansyah Gani**: Conceptualization, methodology, writing—original draft. **Henny Pramoedyo**: Supervision, validation, editing. **Suci Astutik**: Supervision, validation, editing.

Declaration of Generative AI and AI-assisted Technologies

The authors declare that no generative AI tools were used to generate or modify the data, results, or analysis of this study. The AI tools were used only for grammar and formatting improvement.

Declaration of Competing Interest

The authors declare no competing financial or personal interest.

Funding and Acknowledgments

This research received no external funding. The authors thank BPS Gorontalo for providing the dataset.

Data and Code Availability

The data used in this study are publicly available from the Badan Pusat Statistik (BPS) at https://gorontalokota.bps.go.id/id/statistics-table/2/NTIjMg==/indeks-pembangunan-manusia-kabupaten-kota-.html. The code used for analysis in this study can be provided by the corresponding author upon reasonable request.

References

[1] Ni Made Shantia Ananda, Suyitno Suyitno, and Meiliyani Siringoringo. "Geographically Weighted Panel Regression Modelling of Human Development Index Data in East Kalimantan Province in 2017-2020". In: *Jurnal Matematika, Statistika dan Komputasi* 19.2 (2023), pp. 323–341. DOI: 10.20956/j.v19i2.23775.

[2] Ragdad Cani Miranti and Carlos Mendez. "Social and economic convergence across districts in Indonesia: A spatial econometric approach". In: *Bulletin of Indonesian Economic Studies* 59.3 (2023), pp. 421–445. DOI: 10.1080/00074918.2022.2071415.

[1] https://gorontalo.bps.go.id

[3] Gianfranco Piras and Mauricio Sarrias. "Heterogeneous spatial models in R: spatial regimes models". In: *Journal of Spatial Econometrics* 4.1 (2023), p. 4. DOI: 10.1007/s43071-023-00034-1.

[4] Tomohiro Ando, Kunpeng Li, and Lina Lu. "A spatial panel quantile model with unobserved heterogeneity". In: *Journal of econometrics* 232.1 (2023), pp. 191–213. DOI: 10.1016/j.jeconom.2021.08.004.

[5] Agus Rusgiyono and Alan Prahutama. "Geographically Weighted Panel Regression With Fixed Effect for Modeling the Number of Infant Mortality in Central Java, Indonesia". In: *Media Statistika* 14.1 (2021), pp. 10–20. DOI: 10.14710/medstat.14.1.10-20.

[6] Amelia Fadila Rahman, Syafriandi Syafriandi, Nonong Amalita, et al. "Geographically Weighted Panel Regression Modeling on Human Development Index in West Sumatra". In: *UNP Journal of Statistics and Data Science* 1.3 (2023), pp. 232–239. DOI: 10.24036/ujsds/vol1-iss3/63.

[7] Alfira Mulya Astuti, Afifurrahman Afifurrahman, and Habibi Ratu Perwira Negara. "Analysis of Human Development Index in West Nusa Tenggara Province with Spatial Panel Model". In: *(IJCSAM) International Journal of Computing Science and Applied Mathematics* 10.2 (2024), pp. 113–118. DOI: 10.12962%2Fj24775401.v10i2.21957.

[8] Román Salmerón-Gómez, Catalina B García-García, and José García-Pérez. "A redefined variance inflation factor: overcoming the limitations of the variance inflation factor". In: *Computational Economics* 65.1 (2025), pp. 337–363. DOI: 10.1007/s10614-024-10575-8.

[9] Daojun Zhang and Yu Zhang. "Moran's I of VRPAD: A human activity-sensitive spatial pattern index for vegetation restoration evaluation". In: *Journal of Environmental Management* 387 (2025), p. 125948. DOI: 10.1016/j.jenvman.2025.125948.

[10] Bülent Güloğlu et al. "Testing Homoskedasticity in Spatial Panel Data Models". In: *Econometrics and Statistics* (2024). DOI: 10.1016/j.ecosta.2024.04.003.

[11] Hande Karabiyik, Franz C Palm, and Jean-Pierre Urbain. "Econometric analysis of panel data models with multifactor error structures". In: *Annual Review of Economics* 11.1 (2019), pp. 495–522. DOI: 10.1146/annurev-economics-063016-104338.

[12] Mengqi Zhang and Boping Tian. "Estimation of the Non-Parametric Spatial Dynamic Panel Data Model with Fixed Effects". In: *Mathematics* 11.13 (2023), p. 2865. DOI: 10.3390/math11132865.

[13] Leslie E. Papke and Jeffrey M. Wooldridge. "A Simple, Robust Test for Choosing the Level of Fixed Effects in Linear Panel Data Models". In: *Empirical Economics* 64.6 (2023), pp. 2683–2701. DOI: 10.1007/s00181-022-02337-y.

[14] Zhu Huisheng et al. "Exploring groundwater potential: Combining GIS techniques with OLS, GWR, and exploratory regression". In: *Journal of Hydrology: Regional Studies* 61 (2025), p. 102564. DOI: 10.1016/j.ejrh.2025.102564.

[15] Bhaskar Mandal and Kaushalendra Prakash Goswami. "Evaluating the influence of biophysical factors in explaining spatial heterogeneity of LST: Insights from Brahmani-Dwarka interfluve leveraging Geodetector, GWR, and MGWR models". In: *Physics and Chemistry of the Earth, Parts a/b/c* 138 (2025), p. 103836. DOI: 10.1016/j.pce.2024.103836.

[16] Friansyah Gani, Henny Pramoedyo, and Achmad Efendi. "Spatial Variation of HDI in East Java: A Tricube-Based Geographically Weighted Regression–Flower Pollination Algorithm Modeling Approach". In: *Cauchy: Jurnal Matematika Murni dan Aplikasi* 11.1 (2026), pp. 239–254. https://ejournal.uin-malang.ac.id/index.php/Math/article/view/38007.

[17] Hafsah Kurnia, Irma Fauziah, and Madona Yunita Wijaya. "Pemodelan kasus tingkat kemiskinan di Indonesia periode 2015–2021 dengan model regresi panel terboboti geografis". In: *Majalah Ilmiah Matematika dan Statistika* 24.2 (2024). DOI: 10.19184/mims.v24i2.39392.

[18] Arjun Pratama, Suyitno, and Ika Purnamasari. "Pemodelan Persentase Penduduk Miskin di Provinsi Kalimantan Timur Menggunakan Model Geographically Weighted Panel Regression". In: *Jurnal MSA* 9.2 (2024), p. 21021. DOI: 10.24252/msa.v9i2.21021.

[19] Chang Tan, Michaela Kesina, and J Paul Elhorst. "Parameterizing Spatial Weight Matrices in Spatial Econometric Models". In: *Political Analysis* 33.1 (2025), pp. 49–63. DOI: 10.1017/pan.2024.16.

[20] J Paul Elhorst et al. "The distance decay effect and spatial reach of spillovers". In: *Journal of Geographical Systems* 26.2 (2024), pp. 265–289. DOI: 10.1007/s10109-024-00440-5.

[21] Karen Roberts-Licklider and Theodore Trafalis. "Optimizing Treatment Facility Locations in Oklahoma Using Haversine, Euclidean, Manhattan, and Chebyshev Distance Optimization". In: *Operations Research Forum*. Vol. 6. 3. Springer. 2025, p. 102. DOI: 10.1007/s43069-025-00448-7.

[22] Tuba Koç. "Bandwidth selection in geographically weighted regression models via information complexity criteria". In: *Journal of Mathematics* 2022.1 (2022), p. 1527407. DOI: 10.1155/2022/1527407.