CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 325-339
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Spatial Analysis of Child Violence Victims in West Java in 2024 Using Geographically Weighted Negative Binomial Regression

Suliyanto*, Dita Amelia, Lisa Amanda Putri, and Aurellia Calista Anggakusuma

Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia

Abstract

Violence against children remains a critical issue in Indonesia, with West Java consistently reporting high numbers of reported child violence victims. This study examines socioeconomic factors associated with the count of reported child violence victims across 27 districts and cities in West Java in 2024, using secondary administrative data obtained from Open Data Jabar. The explanatory variables include poverty rate (X_1), average years of schooling (X_2), number of divorce cases (X_3), Labor Force Participation Rate (X_4), and Open Unemployment Rate (X_5). Diagnostic tests indicate the presence of spatial heterogeneity and overdispersion, supporting the application of a Geographically Weighted Negative Binomial Regression (GWNBR) model with a child population offset. Model performance comparison based on in-sample fit criteria shows that the GWNBR model provides superior fit (deviance = 42.94, AICc = 99.87) compared to the global Negative Binomial Regression model (deviance = 48.27, AIC = 105.63). The GWNBR results reveal substantial spatial variation: average years of schooling (X_2) is statistically significant across all 27 regions, while the number of divorce cases (X_3) is significant in 23 regions. Poverty rate (X_1) shows localized significance in 16 regions. Labor force participation rate (X_4) and unemployment rate (X_5) each exhibit significance in 6 regions, though with distinct spatial patterns. These findings highlight geographically varying risk structures that cannot be adequately captured by global models and underscore the importance of spatially adaptive modeling for informing region-specific child protection policies. Although the analysis relies on reported administrative data that may not fully represent the true underlying prevalence of child violence, the results provide valuable spatial insights relevant to policy development aligned with SDG 3, SDG 4, and SDG 16.

Keywords: Child Violence; Spatial Analysis; West Java; GWNBR; Negative Binomial; Spatial Heterogeneity; Kernel and Bandwidth; Count Data; Overdispersion.

1. Introduction

Violence against children remains a critical and persistent social problem in Indonesia, despite the implementation of various protection programs, regulatory frameworks, and child-centered national initiatives. Under Law No. 23 of 2002 on Child Protection, violence against children is defined broadly as any act that causes physical, psychological, or sexual suffering, including neglect, coercion, threats, and unlawful deprivation of liberty [1]. This definition aligns with

*Corresponding author. E-mail: suliyanto@fst.unair.ac.id

Submitted: January 12, 2026 Reviewed: January 26, 2026 Accepted: February 12, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.40390

the World Health Organization, which characterizes violence against children as all forms of violence directed at individuals under 18 years of age by parents, caregivers, peers, or strangers [2]. National evidence confirms the magnitude of the problem. The 2024 National Survey on the Life Experiences of Children and Adolescents reported that 33.64% of adolescents aged 13–17 had experienced at least one form of violence [3]. In the same year, administrative records from the SIMFONI-PPA system documented 31,947 reported victims of violence nationwide, including 2,259 victims involving children [4]. While these figures underscore the persistence of child victimization, they must be interpreted cautiously, as administrative data are inherently shaped by differences in reporting mechanisms, institutional capacity, service accessibility, and public awareness across regions.

West Java, Indonesia's most populous province, consistently records among the highest numbers of reported child violence victims. According to data from the West Java Office of Women's Empowerment, Child Protection, and Family Planning (DP3AKB), a total of 3,084 violence victims were reported in 2024, with approximately 95% involving children [5]. However, the observed spatial distribution of reported child violence victims across districts and cities reflects reported data rather than the true underlying prevalence of child violence. Several regions report zero or very low counts, which may indicate underreporting rather than the absence of violence. This is consistent with the findings of [6], who reported that the prevalence of violence experienced by adolescents is substantially higher than officially recorded figures, as many victims go unreported or are not captured in administrative systems [6]. Such patterns are closely linked to variations in administrative capacity, availability of reporting services, community awareness, and social stigma. Consequently, underreporting represents an important data limitation that must be explicitly acknowledged when analyzing spatial patterns of child violence using administrative records.

While underreporting complicates the accurate assessment of child violence, existing evidence suggests that the risk of violence is also shaped by complex and spatially heterogeneous socioeconomic conditions. Previous empirical studies have identified poverty, educational attainment, number of divorce cases, labor force participation, and unemployment as key determinants associated with increased child vulnerability [7]. Socioeconomic stressors can elevate parenting stress, weaken household stability, and reduce protective supervision, thereby increasing the likelihood of violence against children [8, 9]. Importantly, the magnitude and direction of these effects are not spatially uniform. Differences in urbanization, population density, local labor markets, cultural norms, and access to social services may cause the same socioeconomic factor to exert varying influences across regions. Ignoring this geographic heterogeneity risks oversimplifying the underlying risk structure and limits the effectiveness of policy responses.

Despite growing academic attention to child protection issues in Indonesia, much of the existing literature remains dominated by qualitative, legal, or purely descriptive analyses focusing on typologies of violence, psychological consequences, or regulatory frameworks [10]. Spatially explicit quantitative studies at the district or city level remain limited, and national-level statistical analyses often rely on aggregated data that obscure substantial sub-provincial variation [11]. From a methodological perspective, data on child violence victims are typically recorded as non-negative integers and frequently exhibit overdispersion, where the variance exceeds the mean [12]. This characteristic violates the assumptions of standard Poisson regression models and necessitates alternative modeling strategies that can accommodate both overdispersion and spatial heterogeneity.

Geographically Weighted Negative Binomial Regression (GWNBR) provides a robust analytical framework for addressing these challenges. By allowing regression coefficients to vary across geographic locations, GWNBR captures local variations in the relationships between socioeconomic determinants and child violence victimization while appropriately handling overdispersed count data [13]. In this study, spatial dependence and heterogeneity were formally evaluated using Moran's I and Breusch–Pagan statistics applied to model residuals, with the global Negative Binomial Regression (NBR) model serving as a benchmark [14]. The presence of spatial heterogeneity supports the use of GWNBR to uncover localized risk patterns that cannot be adequately represented by global models.

Accordingly, this study applies the GWNBR model to examine spatial variations in the determinants of child violence victims across 27 districts and cities in West Java using cross-sectional data for 2024. Five key socioeconomic indicators poverty rate, average years of schooling, number of divorce cases, Labor Force Participation Rate (LFPR), and Open Unemployment Rate (OUR) are analyzed to identify region-specific effects. By generating a spatially explicit risk profile, this study aims to support evidence based policy formulation through regional risk ranking, prioritization of targeted interventions, and more efficient allocation of child protection resources. In this way, the findings contribute to more operational and outcome oriented policy insights aligned with the objectives of Sustainable Development Goals SDG 3 (Good Health and Well-being), SDG 4 (Quality Education), and SDG 16 (Peace, Justice, and Strong Institutions) [15].

2. Methods

This section describes the overall workflow of the study, including data preparation, spatial modeling using GWNBR, parameter estimation, and model evaluation. Each subsection explains the methodological steps applied to ensure robust spatial analysis and reliable inference.

2.1. Data and Variables

This study uses secondary cross-sectional data from Open Data Jabar [16], covering 27 districts/cities in West Java Province for 2024. The response variable (Y) is the total number of reported child violence victims in each district/city. Geographic coordinates of district/city centroids were initially obtained in the WGS 1984 geographic coordinate system (EPSG:4326) and subsequently projected into the Universal Transverse Mercator (UTM) Zone 48S coordinate system (EPSG:32748). This transformation ensures that all Euclidean distance and bandwidth computations are performed in meters using projected planar coordinates.

As the data come from administrative records, zero counts in some regions may reflect underreporting rather than absence of violence. To account for population differences, the child population (aged 0–17 years) is included as a logarithmic offset, log(Child Population), in all regression models, allowing interpretation in terms of risk rather than absolute counts. Variable definitions, units, and sources are summarized in Table 1.

Table 1: Definition of Variables Used in the Study

Variable	Description	Unit	Role in Model	Source
Y	Total number of reported child violence victims	Count (victims)	Response variable	Open Data Jabar
X_1	Poverty rate	Percentage (%)	Explanatory variable	BPS/Open Data Jabar
X_2	Average years of schooling	Years	Explanatory variable	BPS
X_3	Number of divorce cases	Count (cases)	Explanatory variable	Religious Court Statistics
X_4	Labor Force Participation Rate (LFPR)	Percentage (%)	Explanatory variable	BPS
X_5	Open Unemployment Rate (OUR)	Percentage (%)	Explanatory variable	BPS
Offset	Child population (aged 0–17 years)	Count (persons)	log(Child Population)	BPS

Predictor variables were analyzed in their original measurement scales without standardization, allowing regression coefficients to be directly interpreted as effects per unit change in the original variables.

2.2. Multicollinearity Test

Multicollinearity among explanatory variables was examined using the Variance Inflation Factor (VIF), defined as

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \ldots, p,$$

where R_j^2 is the coefficient of determination obtained from regressing the j-th explanatory variable on all remaining predictors. A VIF value exceeding 10 indicates serious multicollinearity [17].

2.3. Poisson Regression

Poisson regression was initially employed as a baseline model for count data [18]. A random variable Y follows a Poisson distribution with mean μ if its probability mass function is

$$P(Y = y; \mu) = \frac{e^{-\mu}\mu^y}{y!}, \quad y = 0, 1, 2, \ldots$$

with $E(Y) = \mu$ and $\text{Var}(Y) = \mu$. The Poisson regression model with a log-link function and exposure adjustment is specified as

$$\log(\mu_i) = \log(\text{Child Population}_i) + \mathbf{x}_i^T \boldsymbol{\beta},$$

where $\mathbf{x}_i$ denotes the vector of explanatory variables and $\boldsymbol{\beta}$ represents the regression parameters. Parameter estimation is conducted using Maximum Likelihood Estimation (MLE).

2.4. Overdispersion Test

Overdispersion was assessed using dispersion statistics based on the residual deviance and Pearson chi-square obtained from the Poisson regression model [19].

$$\phi_1 = \frac{D^2}{df}, \quad D^2 = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right],$$

$$\phi_2 = \frac{\chi^2}{df}, \quad \chi^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i},$$

where $\hat{\mu}_i$ is the fitted mean from the Poisson model and $df = n - p - 1$. Values of ϕ_1 or ϕ_2 greater than one indicate extra-Poisson variation, suggesting that the Poisson model is inadequate and motivating the use of the Negative Binomial Regression model.

2.5. Negative Binomial Regression

Negative Binomial Regression (NBR) assumes that the response variable follows a Poisson–Gamma mixture distribution with probability mass function.

$$f(y; \mu, \theta) = \frac{\Gamma(y + 1/\theta)}{\Gamma(1/\theta)y!}\left(\frac{1}{1 + \theta\mu}\right)^{1/\theta}\left(\frac{\theta\mu}{1 + \theta\mu}\right)^{y}.$$

The mean response is modeled as

$$\log(\mu_i) = \log(\text{Child Population}_i) + \beta_0 + \sum_{k=1}^{p}\beta_k X_{ki}.$$

Parameter estimation is performed using MLE. Simultaneous significance testing is conducted using the likelihood ratio test, while partial significance testing is based on the Wald z-statistic [20].

2.6. Spatial Dependence and Heterogeneity Tests

Spatial effects were examined to justify the use of a geographically weighted modeling approach. Two aspects were considered: spatial dependence and spatial heterogeneity. Spatial dependence was assessed using Moran's I statistic applied to the *residuals of the global Negative Binomial Regression (NBR) model* [21]. Testing Moran's I on residuals allows the detection of remaining spatial autocorrelation after accounting for the effects of explanatory variables. A statistically significant result indicates that the global model fails to fully capture the underlying spatial structure.

A contiguity-based spatial weight matrix was employed using the *queen contiguity* criterion, where two regions are considered neighbors if they share either a common boundary or vertex. The spatial weights w_{ij} were row-standardized prior to analysis.

Moran's I statistic is defined as

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{S_e^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}},$$

where e_i denotes the residual of the global NBR model at location i, $\bar{e}$ is the mean of the residuals, and S_e^2 is the residual variance. The standardized test statistic is given by

$$Z_I = \frac{I - E(I)}{\sqrt{\mathrm{Var}(I)}}.$$

Statistical significance for Moran's I was evaluated at the 10% significance level ($\alpha = 0.10$), consistent with the threshold used for local parameter inference in the GWNBR model. Spatial heterogeneity was evaluated using the Breusch–Pagan (BP) test to assess whether the variance of the residuals varies systematically across space [22]. The BP test statistic follows a chi-square distribution with degrees of freedom equal to the number of explanatory variables. The BP test was also assessed at the 10% significance level.

Significant results from Moran's I or the BP test indicate the presence of spatial effects that are not adequately addressed by a global model. Consequently, these findings provide a statistical justification for employing the Geographically Weighted Negative Binomial Regression (GWNBR) model to account for spatially varying relationships.

2.7. Geographically Weighted Negative Binomial Regression

To accommodate spatial heterogeneity, the Geographically Weighted Negative Binomial Regression (GWNBR) model was employed [23]. The model is defined as follows.

$$y_i \sim \mathrm{NB}\Bigg(\exp\Big(\sum_{p=1}^{P} \beta_p(u_i, v_i) x_{ip} + \log(\text{Child Population}_i) \Big),\ \theta(u_i, v_i) \Bigg),$$

where y_i represents the total number of reported child violence victims in location i, x_{ip} are the predictor variables, $\beta_p(u_i, v_i)$ are location-specific regression coefficients, $\theta(u_i, v_i)$ is the overdispersion parameter in the location i, and Child Population$_i$ is the total number of children in the location i used as offset. Parameter estimation is carried out using Maximum Likelihood Estimation (MLE).

Including the offset allows the model to account for differences in population at risk, effectively modeling the rate of child violence per child rather than absolute counts. The model allows coefficients to vary locally across districts/cities, capturing spatial non-stationarity in the effects of socioeconomic factors. The significance of local parameters can be evaluated using standardized Z-statistics, with a predictor considered significant if $|Z| > 1.645$ at the 10% significance level, consistent with previous applications of GWNBR.

Spatial weighting in the GWNBR model is based on distances between district and city centroids. Although centroid coordinates were initially obtained in the WGS 1984 geographic coordinate system (EPSG:4326), distance calculations were performed after transforming the coordinates into the UTM Zone 48S projected coordinate system (EPSG:32748), so that all Euclidean distances and bandwidths were computed in meters rather than angular degrees.

2.8. Distance, Bandwidth, and Weighting Function

The spatial distance between locations was calculated using Euclidean distance [24]. For this purpose, centroid coordinates were expressed in the UTM Zone 48S projected coordinate system (EPSG:32748), such that the coordinate pairs (u_i, v_i) represent planar locations in meters rather than angular degrees.

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2},$$

where (u_i, v_i) and (u_j, v_j) denote the projected planar coordinates of locations i and j, respectively.

The optimal adaptive bandwidth was selected through cross-validation (CV) by minimizing the model deviance [25]:

$$b_{\text{opt}} = \arg\min_b \sum_{i=1}^{n} D\left(y_i, \hat{y}_{\neq i}(b)\right),$$

where $\hat{y}_{\neq i}(b)$ is the predicted value at location i using a model fitted with bandwidth b excluding that observation, and $D(\cdot)$ denotes the deviance. This approach follows the standard methodology for geographically weighted models [24]. An adaptive bi-square kernel weighting function was applied [26]:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b_i}\right)^2\right)^2, & d_{ij} < b_i, \\ 0, & d_{ij} \geq b_i. \end{cases}$$

2.9. Model Evaluation

Model fit was evaluated using the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) [27, 28]. The AIC is defined as

$$\text{AIC} = -2\ell(\hat{\theta}) + 2k,$$

where $\ell(\hat{\theta})$ denotes the maximized log-likelihood function and k is the number of estimated parameters. The AICc is a finite-sample correction of AIC and is defined as

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1},$$

where n is the sample size. Smaller values of AIC and AICc indicate a better model fit.

In-sample model performance was additionally assessed using the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) [29]. The MAE is defined as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|,$$

while the RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

Lower values of MAE and RMSE indicate better in-sample fit. MAE and RMSE are reported as complementary in-sample error measures for count outcomes.

3. Results and Discussion

This section presents the modeling results, starting with global model estimation followed by geographically weighted analysis and spatial pattern interpretation. The discussion then elaborates on the implications of the findings and their relevance to regional policy and intervention strategies.

3.1. Distribution of Reported Child Violence Victims

In 2024, a total of 460 reported child violence victims were recorded across 27 regencies and cities in West Java, corresponding to an average of 17.04 victims per region. The spatial distribution of reported victims exhibits marked heterogeneity. Several regions namely Sukabumi Regency, Kuningan Regency, Majalengka Regency, Purwakarta Regency, Pangandaran Regency, Cimahi City, Tasikmalaya City, and Banjar City reported zero victims, whereas Bekasi City recorded the highest number, with 109 reported child violence victims. The total of 460 reported victims refers specifically to district/city-level aggregated administrative records obtained from Open Data Jabar and therefore differs from figures reported by DP3AKB at the provincial level.

The spatial distribution of child violence victims is visualized in Fig. 1, revealing marked disparities across regions and suggesting potential spatial clustering patterns.

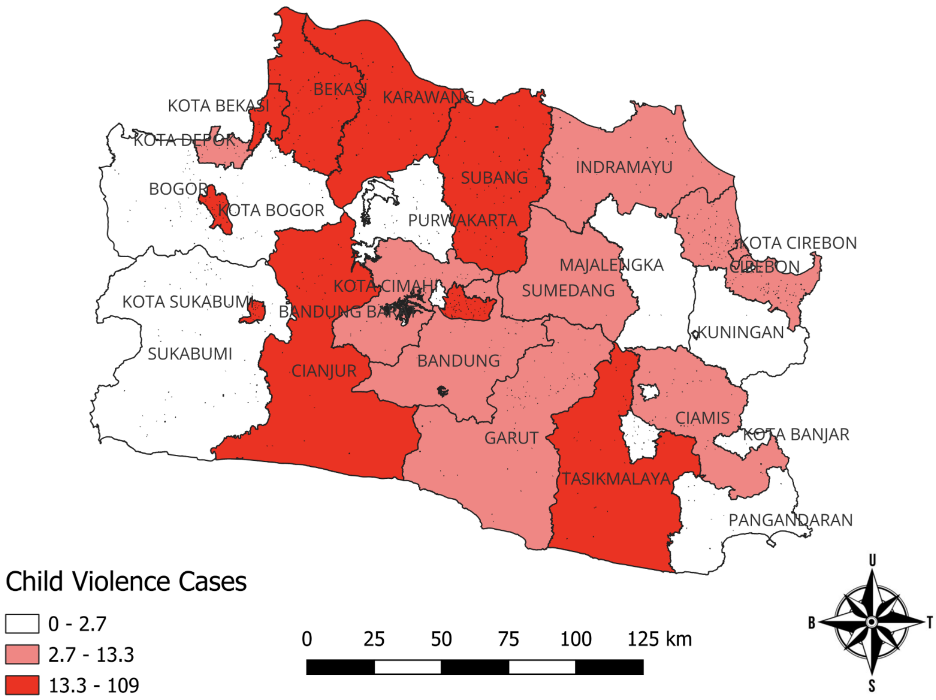

Fig. 1: Distribution of Child Violence Victims in West Java

It is important to note that the data used in this study are derived from administrative records. Therefore, zero counts observed in several districts may not necessarily indicate the absence of child violence victims. Instead, they may reflect differences in reporting capacity, public awareness, access to child protection services, and institutional readiness across regions. This potential underreporting constitutes an inherent limitation of administrative and ecological data and is explicitly acknowledged in the interpretation of the results.

To ensure comparability across regions with different population sizes, all count regression models incorporate the logarithm of the child population (aged 0–17 years) as an offset term. Consequently, the analysis focuses on the incidence of reported child violence rather than absolute case counts.

3.2. Multicollinearity Test

Multicollinearity among explanatory variables was examined using the Variance Inflation Factor (VIF). The results are presented in Table 2.

Table 2: VIF Values of Predictor Variables

Variable	VIF
X_1	6.597257
X_2	4.598448
X_3	4.756418
X_4	1.982776
X_5	1.858050

All VIF values are below 10, indicating the absence of multicollinearity. Therefore, all predictor variables were retained for further analysis.

3.3. Poisson Regression Model

The Poisson regression parameter estimates are summarized in Table 3. A Poisson regression model with a log link function and an offset for the child population was initially estimated. Parameter estimates are presented in Table 3.

Table 3: Poisson Regression Parameter Estimates with Offset

Parameter	Estimate	z-value	p-value
Intercept	-36.9800	-15.053	$< 2 \times 10^{-16}$
X_1 (Poverty rate)	0.0031	2.367	0.0179
X_2 (Average years of schooling)	1.1990	15.153	$< 2 \times 10^{-16}$
X_3 (Number of divorce cases)	0.0004	7.578	3.50×10^{-14}
X_4 (LFPR/TPAK)	0.1529	5.532	3.16×10^{-8}
X_5 (Open unemployment rate)	0.4077	7.586	3.31×10^{-14}
Residual deviance = 715.95 df = 21 AIC = 811.67			

The estimated Poisson regression model is expressed as

$$\begin{aligned}\ln(\mu_i) = \ln(\text{Child Population}_i) &- 36.9800 + 0.0031X_{1i} + 1.1990X_{2i} + 0.0004X_{3i} \\ &+ 0.1529X_{4i} + 0.4077X_{5i}.\end{aligned}$$

At the 10% significance level, all explanatory variables are statistically significant. Poverty rate, average years of schooling, number of divorce cases, labor force participation rate, and open unemployment rate all have a significant positive association with the number of reported child violence victims after controlling for child population size through the offset term.

The overdispersion statistic based on the residual deviance is

$$\phi_1 = \frac{715.95}{21} \approx 34.09 > 1,$$

indicating substantial overdispersion. Therefore, the Poisson model is inadequate and the Negative Binomial Regression model is subsequently employed.

3.4. Negative Binomial Regression Model

To address overdispersion and variations in population exposure across regions, a Negative Binomial Regression (NBR) model with an offset for child population (aged 0–17 years) was applied. The offset was specified in logarithmic form to model the incidence rate of child victims per population at risk.

The parameter estimates of the NBR model are presented in Table 4.

Table 4: Negative Binomial Regression Parameter Estimates

Parameter	Estimate	z-value	p-value
Intercept	-26.4900	-2.640	0.0083
X_1 (Poverty Rate)	0.0001	0.020	0.9844
X_2 (Average Years of Schooling)	1.0210	3.587	0.0003
X_3 (Number of Divorces Cases)	0.0006	2.197	0.0280
X_4 (Labor Force Participation Rate)	0.0309	0.255	0.7984
X_5 (Open Unemployment Rate)	0.3838	1.525	0.1273
AIC = 105.63 $\quad \theta = 0.5268$			

Based on Table 4, at the 10% significance level, average years of schooling (X_2) and the number of divorce cases (X_3) have statistically significant effects on the number of reported child violence victims. In contrast, poverty rate (X_1), labor force participation rate (X_4), and open unemployment rate (X_5) are not statistically significant in the global NBR model (all $p > 0.10$).

This result indicates that regions with higher average years of schooling and higher number of divorce cases tend to have higher incidence rates of child victims. The positive coefficient for unemployment rate (X_5) suggests a potential association, but it is not statistically significant at the 10% level ($p = 0.127$). The estimated Negative Binomial Regression model with an offset is expressed as

$$\begin{aligned}\ln(\mu_i) = \ln(\text{Child Population}_i) &- 26.4900 + 0.0001X_{1i} + 1.0210X_{2i} + 0.0006X_{3i} \\ &+ 0.0309X_{4i} + 0.3838X_{5i}.\end{aligned}$$

This result indicates that regions with higher average years of schooling and higher number of divorce cases tend to have higher incidence rates of child victims, after accounting for differences in child population size. The unemployment rate shows a positive but statistically nonsignificant association.

3.5. Spatial Dependence and Heterogeneity Tests

The Breusch–Pagan test yielded $BP = 16.459$ with $p = 0.00565$, indicating spatial heterogeneity. Moran's I test produced a p-value of 0.07522, revealing spatial dependence at the 10% significance level. These results justify the application of a geographically weighted model.

3.6. Geographically Weighted Negative Binomial Regression

The Geographically Weighted Negative Binomial Regression (GWNBR) model was employed to assess spatial non-stationarity in socioeconomic effects on child violence across districts and cities in West Java, with local parameters estimated for each spatial unit. An adaptive bisquare kernel was applied, and distance computations were performed in projected coordinate space (meters) to ensure that Euclidean distances are meaningful. The optimal bandwidth of 20,547.87 meters (approximately 20.5 km) represents the local influence of neighboring regions rather than angular degrees. The spatial weight matrix was constructed using the district distance matrix for local estimation. The model includes an offset for the total child population in each district, allowing estimation of child violence rates per child rather than absolute counts.

The significance of the local parameters was evaluated using standardized Z-statistics, with a predictor considered significant if $|Z| > 1.645$ at the 10% significance level. Estimated coefficients represent multiplicative effects on the expected count due to the log-link, meaning a one-unit increase in a predictor multiplies the expected count by $\exp(\beta)$ at a given location.

Model performance was evaluated by comparing the global Negative Binomial Regression (NBR) and the GWNBR using deviance and information criteria appropriate to each modeling framework. For consistency in reporting, AIC is reported for global models (Poisson and NBR), while AICc is reported for GWNBR to account for its effective number of local parameters. The global NBR model yielded a deviance of 48.27 and an AIC value of 105.63, while the GWNBR model exhibited a reduced deviance of 42.94 and a lower AICc value of 99.87. Overall, these results indicate that allowing regression coefficients to vary spatially improves in-sample model fit and more effectively captures local heterogeneity in socioeconomic effects.

Table 5: Model Performance Comparison Based on In-sample Fit Criteria

Model	Deviance	Information Criterion	Value
Global NBR	48.27	AIC	105.63
GWNBR	42.94	AICc	99.87

It should be noted that a substantial number of districts reported zero child violence victims, which may reflect underreporting rather than absence of victims. Therefore, these results should be interpreted cautiously, acknowledging the potential limitations of the data and the uncertainty in local estimates.

Table 6: Grouping of Regencies/Cities Based on GWNBR Model

Significant Variables	Regencies/Cities	Number
X_1, X_2, X_3	Cirebon, Garut, Indramayu, Kota Banjar, Kota Bekasi, Kota Cirebon, Kuningan, Majalengka, Subang, Sumedang	10
X_1, X_2, X_3, X_4, X_5	Bandung, Kota Bandung	2
X_1, X_2, X_3, X_5	Ciamis, Kota Tasikmalaya, Pangandaran, Tasikmalaya	4
X_2, X_3	Bandung Barat, Bekasi, Cianjur, Karawang, Kota Cimahi, Kota Depok, Purwakarta	7
X_2, X_4	Bogor, Kota Bogor, Kota Sukabumi, Sukabumi	4

The grouping of regencies/cities based on significant local predictors is summarized in Table 6. Ten regions show simultaneous significance of poverty rate (X_1), average years of schooling (X_2), and number of divorce cases (X_3), while seven regions exhibit significance only for X_2 and X_3. Four regions show significance for X_1, X_2, X_3, and unemployment rate (X_5), and four regions demonstrate significance for X_2 and labor force participation rate (X_4). Bandung City emerges as the only area where all predictors (X_1–X_5) are significant, reflecting a more complex socioeconomic structure compared to other districts.

As an illustration, the estimated number of child violence victims in Bandung using the GWNBR approach is presented in Table 7 as follows.

Table 7: Testing GWNBR Model Parameters in Bandung City

Parameter	Estimate	z-statistic
Offset: log(childpop)	(fixed)	–
$\hat{\beta}_0$	-5.0644	-13.9872
$\hat{\beta}_1$	0.2376	7.8759
$\hat{\beta}_2$	0.1532	6.0274
$\hat{\beta}_3$	0.00543	3.6873
$\hat{\beta}_4$	0.12846	4.2045
$\hat{\beta}_5$	0.00463	7.3066

Based on Table 7, the estimated number of child violence victims in Bandung City was obtained using the GWNBR model approach as follows

$$\begin{aligned} \ln(\hat{\mu}) = \ & \log(\text{Child Population}) - 5.0644 + 0.2376X_1 + 0.1532X_2 \\ & +0.00543X_3 + 0.12846X_4 + 0.00463X_5 \end{aligned} \tag{1}$$

Based on Table 7, each variable meets the absolute value of $|Z| > 1.645$ with a significance level of 10%, indicating that poverty rate (X_1), average years of schooling (X_2), number of divorce cases (X_3), labor force participation rate/TPAK (X_4), and unemployment rate/TPT (X_5) each have a significant effect on the rate of child violence victims per child population in Bandung City.

Based on the results of the GWNBR model estimation in equation Eq. (1), which includes an offset term for child population, the coefficients represent multiplicative effects on the expected rate of child violence victims per child. The interpretation is as follows. A one percentage point increase in the poverty rate (X_1) increases the expected rate of child violence victims by a factor of $\exp(0.2376) \approx 1.268$ times (a 26.8% increase), assuming other variables remain constant. An additional year in the average years of schooling (X_2) increases the expected rate by a factor of $\exp(0.1532) \approx 1.166$ times (a 16.6% increase). A one unit increase in the number of divorces cases (X_3) increases the expected rate by a factor of $\exp(0.00543) \approx 1.005$ times (a 0.5% increase). A one percentage point increase in the labor force participation rate (X_4) increases the expected rate by a factor of $\exp(0.12846) \approx 1.137$ times (a 13.7% increase). Lastly, a one percentage point increase in the unemployment rate (X_5) increases the expected rate by a factor of $\exp(0.00463) \approx 1.005$ times (a 0.5% increase).

Based on the results of modeling the number of child violence victims in West Java in Table 6 using thematic maps as presented in Fig. 2, the results are as follows:

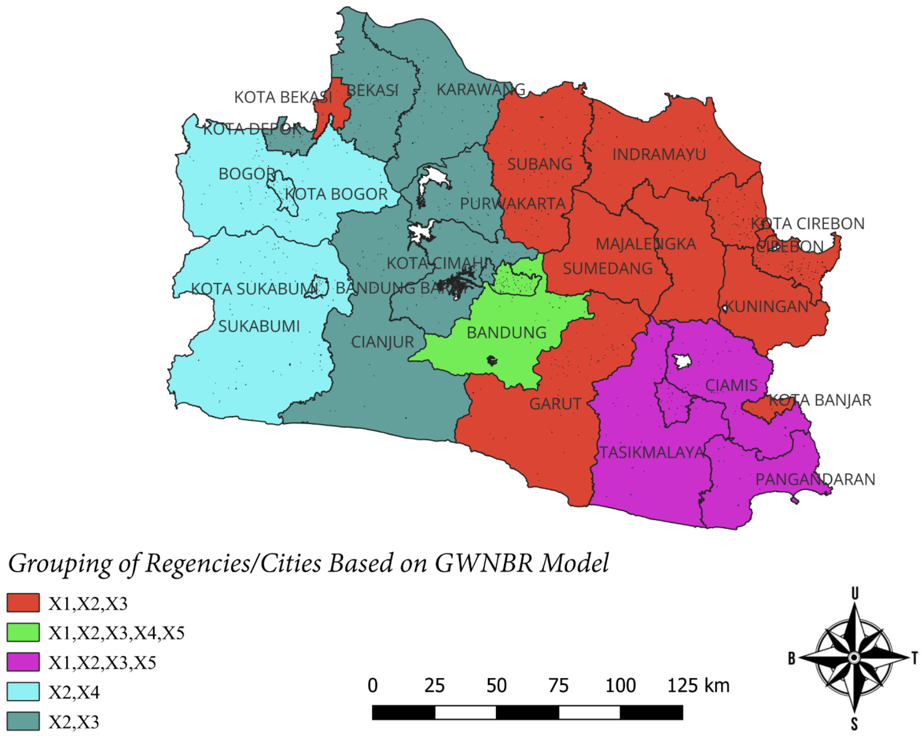

Fig. 2: Grouping of Regencies/Cities Based on Significant Variables in the GWNBR Model

Based on Fig. 2, the GWNBR modeling results identified five spatial patterns in West Java Province according to locally significant socioeconomic variables, highlighting spatial heterogeneity in factors influencing child violence. The red cluster (X_1, X_2, X_3) is significantly affected by poverty rate (X_1), average years of schooling (X_2), and number of divorces cases (X_3), indicating that basic economic conditions and family dynamics are the primary determinants in these areas. The green group (X_1, X_2, X_3, X_4, X_5) represents the regions with the most complex socioeconomic structure, where all variables are locally significant, suggesting that child violence is simultaneously influenced by poverty, education, family dynamics, participation in the labor force, and unemployment. The purple cluster (X_1, X_2, X_3, X_5) is influenced by poverty, education, divorce, and unemployment, while labor force participation (X_4) is not significant, implying that job stability rather than participation matters more in these areas. The light blue cluster (X_2, X_4) is significantly affected only by education and labor force participation, typically found in regions with relatively lower poverty, where human capital quality and economic engagement are the dominant determinants. The dark blue cluster (X_2, X_3) shows significant influence of education and divorce, indicating the predominance of social and family factors over macroeconomic conditions. Notably, Bandung City emerges as the sole urban area within the cluster where all variables are locally significant, reflecting complex urban dynamics and multidimensional interactions among socioeconomic factors.

3.7. Model Comparison

Model performance was evaluated using appropriate information criteria for each modeling framework: Akaike Information Criterion (AIC) was used for global models (Poisson and NBR), while the corrected Akaike Information Criterion (AICc) was used for the geographically weighted model (GWNBR) to account for its effective number of local parameters. In-sample prediction accuracy was assessed using mean absolute error (MAE) and root mean square error (RMSE); both are in-sample error measures rather than out-of-sample predictive accuracy metrics. The comparison results are summarized in Table 8.

Table 8: Model Performance Comparison Based on In-sample Fit Criteria

Model	Information Criterion	Deviance	MAE	RMSE
Poisson Regression	811.67 (AIC)	715.95	5.42	7.21
Negative Binomial Regression (NBR)	105.63 (AIC)	48.27	3.67	4.75
Geographically Weighted NBR (GWNBR)	99.87 (AICc)	42.94	2.31	3.19

Based on Table 8, the model comparison results demonstrate that the GWNBR provides the strongest performance among the three approaches evaluated. The GWNBR model yields the lowest AICc value (99.87), alongside the smallest deviance (42.94), MAE (2.31), and RMSE (3.19), indicating superior in-sample model fit compared to the Poisson and global Negative Binomial models. The substantial reduction in information criteria (AICc = 5.76 between NBR and GWNBR) provides strong evidence that incorporating spatial heterogeneity substantially improves in-sample model fit and explanatory performance. Furthermore, the lower MAE and RMSE values indicate smaller in-sample prediction errors, reflecting improved goodness-of-fit of the GWNBR model.

These findings confirm the importance of accounting for spatial non-stationarity when modeling child violence determinants. The superiority of GWNBR aligns with previous studies showing that geographically weighted approaches consistently outperform global models when spatial variation is present [30]. However, it is important to acknowledge that these comparisons are based on in-sample fit metrics. The model's performance in out-of-sample prediction requires further validation. Additionally, while GWNBR captures localized relationships, ecological inferences should be made cautiously due to potential underreporting variations across regions and the inherent limitations of cross-sectional administrative data.

4. Conclusion

This study demonstrates that the Geographically Weighted Negative Binomial Regression (GWNBR) model provides a relatively better representation of the spatial distribution of child violence victims in West Java compared to the Poisson and global Negative Binomial models. The GWNBR achieves the lowest AIC and AICc values and shows improved capacity to capture spatial heterogeneity, as supported by diagnostic assessments of overdispersion and spatial non-stationarity. The results reveal substantial local variation in the effects of education, divorce, poverty, labor force participation, and unemployment, indicating that the determinants of child violence are spatially non-uniform. Bandung City exhibits the most complex pattern with all five predictors being locally significant. From a methodological perspective, these findings highlight the importance of spatially adaptive modeling for overdispersed count data, particularly when analyzing administrative data with inherent limitations such as underreporting and zero-inflation. The inclusion of a child population offset enhances the epidemiological interpretability of the model by accounting for differential population sizes across regions. From a policy perspective, the identified spatial patterns suggest that interventions aimed at reducing child violence should be designed based on local socioeconomic conditions rather than implemented uniformly across regions. Future research may extend this framework by incorporating temporal dynamics, validating out-of-sample predictive performance, and addressing ecological inference limitations through multilevel or individual level data integration.

CRediT Authorship Contribution Statement

Suliyanto: Conceptualization, Methodology, Investigation, Writing–Original Draft Preparation. **Dita Amelia:** Data Curation, Formal Analysis, Validation, Writing–Review & Editing. **Lisa Amanda Putri:** Software, Visualization, Data Analysis, Writing–Review & Editing. **Aurellia Calista Anggakusuma:** Supervision, Project Administration, Methodological Review.

Declaration of Generative AI and AI-assisted Technologies

Generative AI and AI-assisted tools were employed solely for technical and linguistic support during the manuscript preparation process. ChatGPT was used to facilitate the conversion of content into LaTeX format and to address compilation-related issues, whereas DeepL assisted in translating certain sections into English. All analyses, interpretations, and final versions of the manuscript were critically reviewed and approved by the authors, who assume full responsibility for the accuracy, originality, and scientific integrity of the study.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research was funded by Universitas Airlangga. The authors would like to acknowledge Universitas Airlangga for providing financial support and institutional facilities that contributed to the completion of this study.

Data and Code Availability

The data underlying this study were collected from open-access public sources. The analytical code developed and used in this research can be provided by the corresponding author upon reasonable request. Availability of both data and code is governed by ethical standards and relevant data usage policies.

References

[1] Republik Indonesia. *Undang-Undang Republik Indonesia Nomor 23 Tahun 2002 tentang Perlindungan Anak*. https://peraturan.bpk.go.id/Home/Details/40172/uu-no-23-tahun-2002. Diakses: 10 Oktober 2025. 2014.

[2] World Health Organization. *Violence Against Children*. https://www.who.int/news-room/fact-sheets/detail/violence-against-children. Diakses: 10 Oktober 2025. 2022.

[3] SNPHAR. *National Survey on the Life Experiences of Children and Adolescents*. https://www.snphar.or.id/report-2024.pdf. Diakses: 11 Oktober 2025. 2024.

[4] Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. *SIMFONI-PPA Report*. https://simfoni.kemenpppa.go.id/report2024.pdf. Diakses: 12 Oktober 2025. 2024.

[5] D. Sutarya. *West Java Child Violence Report 2024*. https://dp3akb.jabarprov.go.id/laporan-kekerasan-anak-2024. Diakses: 20 Oktober 2025. 2025.

[6] Y. Syukriani, A. Noviandhari, N. Arisanti, E. P. Setiawati, V. K. Rusmil, M. Dhamayanti, and N. Sekarwana. "Cross-sectional survey of underreported violence experienced by adolescents: a study from Indonesia". In: *BMC Public Health* 22 (2022), p. 50. DOI: 10.1186/s12889-021-12427-8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8742168/.

[7] A. Yob, B. Rekan, and C. Candra. "Socioeconomic factors and child vulnerability". In: *Journal of Social Issues* 78.2 (2022), pp. 123–145. DOI: 10.1111/josi.12345.

[8] S. Farooq, R. Kumar, and L. Tan. "Parenting stress and child violence". In: *Child Abuse Review* 33.1 (2024), pp. 45–67. DOI: 10.1002/car.3456.

[9] P. Mehta, Y. Li, and A. Singh. "Labor market instability and child violence". In: *Social Science Research* 50.4 (2023), pp. 201–220. DOI: 10.1016/j.ssresearch.2023.104567.

[10] E. Pratiwi, H. Pramoedyo, S. Astutik, and F. Fauwziyah. "Modeling Geographically Weighted Negative Binomial Regression (GWNBR) on Stunting Incidence in Malang Regency". In: *Jurnal Matematika, Statistika dan Komputasi* 19.1 (2024), pp. 1–15. DOI: 10.20956/j.v19i1.21757.

[11] I. N. Fajril, B. Poerwanto, and H. Hardianti. "Geographically Weighted Negative Binomial Regression (GWNBR) Modeling in Infant Mortality Rate Cases in South Sulawesi". In: *Jurnal Ilmiah Teori dan Aplikasi Statistika* 18.1 (2024), pp. 1–17. DOI: 10.36456/jstat.vol18.no1.a10093.

[12] J. Chen, L. Liu, L. Xiao, C. Xu, and D. Long. "Integrative Analysis of Spatial Heterogeneity and Overdispersion of Crime with a Geographically Weighted Negative Binomial Model". In: *ISPRS International Journal of Geo-Information* 9.1 (2020), p. 60. DOI: 10.3390/ijgi9010060.

[13] D. S. Rini. "Geographically Weighted Negative Binomial Regression for Dengue Fever Cases in Bengkulu Province". In: *Prosiding Seminar Nasional Matematika (PRISMA)*. 2023. https://journal.unnes.ac.id/sju/prisma/article/view/19686.

[14] A. Ayunda, I. Husein, and U. Faigle. "Geographically Weighted Negative Binomial Regression (GWNBR) Modeling of Tuberculosis (TB) in North Sumatra". In: *ZERO: Jurnal Sains, Matematika dan Terapan* (2025). https://jurnal.uinsu.ac.id/index.php/zero/article/view/25006.

[15] UNICEF. *Sustainable Development Goals and Child Protection*. https://www.unicef.org/reports/sdg-child-protection-2021. Diakses: 10 Oktober 2025. 2021.

[16] Open Data Jabar. *West Java Open Data Platform*. https://data.jabarprov.go.id. Diakses: 10 Oktober 2025. 2024.

[17] D. N. Gujarati. *Basic Econometrics*. 5th. New York: McGraw-Hill Education, 2012.

[18] R. H. Myers. *Classical and Modern Regression with Applications*. 2nd. Duxbury Advanced Series in Statistics and Decision Sciences. Boston: Duxbury/Thomson Learning, 1990.

[19] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. 2nd. Monographs on Statistics and Applied Probability. Boca Raton, FL: Chapman & Hall/CRC, 1989.

[20] J. M. Hilbe. *Negative Binomial Regression*. 2nd. Cambridge, UK: Cambridge University Press, 2011.

[21] P. A. P. Moran. "Notes on continuous stochastic phenomena". In: *Biometrika* 37.1/2 (1950), pp. 17–23. DOI: 10.1093/biomet/37.1-2.17.

[22] T. S. Breusch and A. R. Pagan. "A simple test for heteroskedasticity and random coefficient variation". In: *Econometrica* 47.5 (1979), pp. 1287–1294. DOI: 10.2307/1911963.

[23] Z. Su, A. S. Fotheringham, and M. Charlton. "Geographically weighted negative binomial regression model predicts wildfire occurrence". In: *Forests* 10.5 (2019), p. 377. DOI: 10.3390/f10050377.

[24] A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, UK: John Wiley & Sons, 2002.

[25] C. Brunsdon, A. S. Fotheringham, and M. Charlton. "Geographically weighted regression: Modelling spatial non-stationarity". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3 (1998), pp. 431–443. DOI: 10.1111/1467-9884.00145.

[26] T. Nakaya, A. S. Fotheringham, C. Brunsdon, and M. Charlton. "Geographically weighted Poisson regression for disease association mapping". In: *Statistics in Medicine* 24.17 (2005), pp. 2695–2717. DOI: 10.1002/sim.2124.

[27] H. Akaike. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

[28] C. M. Hurvich and C.-L. Tsai. "Regression and time series model selection in small samples". In: *Biometrika* 76.2 (1989), pp. 297–307. DOI: 10.1093/biomet/76.2.297.

[29] R. J. Hyndman and A. B. Koehler. "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4 (2006), pp. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001.

[30] A. R. da Silva and T. C. V. Rodrigues. "A SAS macro for geographically weighted negative binomial regression". In: *Proceedings of the SAS Users Group International Conference (SUGI)*. 2016. https://support.sas.com/resources/papers/proceedings16/SAS123-2016.pdf.