

Spatial Dynamics of Crime and Socio-economic Factors in East Java 2024: A Geographically Weighted Negative Binomial Regression Approach

Toha Saifudin*, Sulyanto, Safira Salsabila, Farhan Aldan Khairian, Salman Syaugi Sungkar, Aji Febriandi

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia

Abstract This study highlights that spatially targeted policies are essential for reducing crime in East Java, as key socio-economic drivers particularly unemployment, poverty, population density, and economic growth vary significantly across regions. This study investigates the spatial variability of crime in East Java, Indonesia, by examining how different socio-economic factors contribute to crime patterns using the Geographically Weighted Negative Binomial Regression (GWNBR) approach. The analysis covers 38 districts and cities, with crime counts as the response variable and six socio-economic indicators number of social assistance beneficiary households, open unemployment rate, number of poor residents, population density, Gini index, and economic growth rate as explanatory variables. The modeling process began with a Poisson regression but revealed substantial overdispersion, prompting the use of the Negative Binomial model. Among all models tested, GWNBR provided the best fit, reflected by the lowest AIC value (406.6973), significantly outperforming the Poisson and Negative Binomial models. Spatial diagnostic tests further confirmed that crime distribution in East Java is influenced by both local spatial interaction and regional disparities. Empirical results show that unemployment, poverty, population density, and economic growth are significant and positive determinants of crime in many areas, whereas the effects of social assistance beneficiaries and income inequality vary depending on local characteristics. Surabaya the region with the highest crime incidence exhibits significant positive contributions from all predictors, highlighting the complex socio-economic pressures in large urban environments. These findings underscore the importance of geographically targeted policies, such as labor-market interventions in regions where unemployment strongly influences crime, expanded welfare programs in areas where poverty has a dominant impact, and improved urban planning in densely populated cities. This study aligns with Indonesia's commitment to the Sustainable Development Goals (SDGs), particularly Goal 1 (No Poverty) and Goal 16 (Peace, Justice, and Strong Institutions), by offering evidence-based recommendations to support crime reduction and social welfare enhancement.

Keywords Crime Modelling, Poverty, Spatial Analysis, Geographically Weighted Negative Binomial Regression, Socio-Economic Factors, East Java

AMS 2010 subject classifications 62J12, 62P25, 62H11, 62-07

DOI: 10.19139/soic-2310-5070-3361

1. Introduction

Public safety is a fundamental aspect of social well-being, and crime remains one of the most visible indicators of insecurity within communities. Crime refers to behaviors that violate legal norms and disrupt social and economic stability, often resulting in significant losses for individuals and society [1]. A substantial body of research shows that variations in crime levels are strongly influenced by socio-economic conditions such as unemployment, poverty, income inequality, and population pressure. These factors may increase economic hardship, weaken social cohesion, and heighten exposure to environments where unlawful activities are more likely to occur [2, 3]. In

*Correspondence to: Toha Saifudin (Email: tohasaifudin@fst.unair.ac.id). Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia.

Indonesia, crime remains an important public concern, particularly in provinces with large populations and diverse socio-economic structures such as East Java.

Two prominent theoretical frameworks underpin crime analysis: routine activity theory and social disorganization theory. Routine activity theory posits that crime occurs when a motivated offender, a suitable target, and the absence of capable guardianship converge in time and space [4]. Social disorganization theory, on the other hand, attributes crime to weakened community institutions and social cohesion in areas characterized by poverty, residential instability, and ethnic heterogeneity [5]. These frameworks are particularly relevant to East Java, where rapid urbanization, economic disparities, and demographic pressure create fertile conditions for crime. The spatial concentration of crime in certain districts reflects both the uneven distribution of economic resources and the varying capacity of local communities to exercise informal social control.

Empirical findings highlight that socio-economic disparities play a critical role in shaping crime patterns. Factors including inequality in income distribution, educational attainment, labor-market instability, and the prevalence of households experiencing economic hardship have been identified as major contributors to crime trends in East Java and Indonesia more broadly [6]. Previous studies show that greater deprivation and inequality increase vulnerability to criminal behavior by restricting access to resources and limiting economic resilience [7].

Other structural indicators such as the unemployment rate, the Human Development Index, population density, and measures of income inequality are also widely recognized as determinants of crime. Higher unemployment may reduce access to income opportunities and increase the likelihood of individuals engaging in illegal activities as a coping mechanism [8]. Meanwhile, densely populated areas tend to experience greater social interaction, competition for limited resources, and pressure on public infrastructure, creating conditions that may elevate crime rates [9]. Income inequality has also been shown to intensify relative deprivation and social tension, contributing to increased involvement in crime across various regions [10]. Importantly, these relationships often vary across geographic space, highlighting the need for spatially sensitive analytical approaches.

East Java presents a particularly complex socio-economic landscape. As Indonesia's second most populous province, it is home to approximately 40 million residents across 38 districts and cities with diverse ethnic groups, including Javanese, Madurese, Tenggerese, and Osing communities, whose distinct cultural norms and social structures shape community cohesion and informal crime control mechanisms. The province has experienced sustained industrial growth concentrated in the Surabaya metropolitan area and the Gerbangkertosusila corridor, generating rapid urbanization and widening spatial inequality between urban cores and rural peripheries. Moreover, government social assistance programs, including conditional cash transfers, have been implemented unevenly across districts, creating variation in the social safety net that may differentially moderate the crime-promoting effects of poverty and unemployment. These regional specificities justify the application of spatially varying coefficient models rather than global regression frameworks.

This study aligns with national and global development priorities. At the national level, understanding the socio-economic context of crime supports Indonesia's Medium-Term Development Plan, which emphasizes poverty reduction, employment creation, and social welfare improvements. At the global level, the research contributes to the Sustainable Development Goals (SDGs), particularly Goal 1 (No Poverty) and Goal 16 (Peace, Justice, and Strong Institutions), by providing analytical insights into the mechanisms that shape crime patterns.

Although research on crime in Indonesia has expanded, many studies continue to rely on global regression models that assume uniform relationships across all regions. Such models overlook spatial heterogeneity and may obscure important local variations. Additionally, indicators representing household-level economic vulnerability such as families receiving government social assistance are rarely integrated into crime modeling despite their relevance for understanding regional deprivation. Furthermore, the use of count-data models with spatial extensions remains limited in the Indonesian criminological literature, with most existing studies employing ordinary least squares or standard Poisson specifications that do not account for overdispersion or spatial non-stationarity simultaneously.

To address these limitations, the present study applies the GWNBR model to examine spatial variations in the socio-economic determinants of crime in East Java. This method accommodates overdispersion in count data and allows model coefficients to vary across locations, resulting in more accurate and context-specific interpretations than traditional global models. The findings are expected to enhance the understanding of spatial crime patterns

and provide policymakers with evidence-based insights to formulate targeted strategies for crime prevention and regional safety improvement.

The remainder of this paper is organized as follows. Section 2 describes the data and research methodology, including variable definitions, model specifications, and the sequential analytical framework. Section 3 presents the empirical results and discussion, covering descriptive statistics, multicollinearity diagnostics, Poisson and Negative Binomial regression estimates, spatial diagnostic tests, GWNBR estimation, and model comparison. Section 4 concludes with key findings, policy implications, methodological limitations, and directions for future research.

2. Methods

2.1. Data Research Variable

This study utilizes two main categories of variables, namely the response variable (Y) and a set of predictor variables (X). All data are secondary and obtained from the East Java Provincial Statistics Agency (Central Statistics Agency or Badan Pusat Statistik (BPS)) for the year 2024. A detailed description of each variable used in the analysis is presented in the following table 1. The six predictor variables were selected based on theoretical grounding in routine activity and social disorganization frameworks, empirical precedent in the Indonesian crime literature, and data availability at the district level from official BPS sources. Potential omitted variable bias is acknowledged as a limitation and is discussed in Section 4.

Table 1. Research Variables

Notation	Variable	Scale	Unit
Y	Total Crime Cases	Ratio	Number of cases
X_1	Number of Social Assistance Beneficiary Households	Ratio	Households
X_2	Open Unemployment Rate	Ratio	Percent (%)
X_3	Number of Poor People	Ratio	Persons (thousands)
X_4	Population Density	Ratio	Persons per km ²
X_5	Gini Ratio (Income Inequality Index)	Ratio	Index (0–1)
X_6	Economic Growth Rate	Ratio	Percent (%)
u	Latitude	Interval	Degrees (°)
v	Longitude	Interval	Degrees (°)

The following discussion elaborates on the theoretical background and definitions of the key variables examined in this research.

1. Crime

Crime refers to behaviors formally categorized as unlawful for statistical reporting purposes. The classification follows the International Classification of Crime for Statistical Purposes (ICCS), which enables consistent definitions and comparable measurements across regions [11]. To account for population differences, crime rates are typically standardized as the number of incidents per 100,000 residents, ensuring that comparisons between areas remain unbiased. It is important to note that official crime statistics are subject to underreporting biases, as certain offenses such as domestic violence, petty theft, and white-collar crimes may go unreported due to social stigma, distrust in law enforcement, or lack of awareness. The data used in this study reflect recorded crime cases from official police reports compiled by BPS, and therefore the findings should be interpreted with the caveat that actual crime incidence may exceed reported figures. This limitation is consistent with aggregate administrative data used across the crime-modeling literature.

2. Socio-economic Factors Affecting Crime

Research in the field consistently demonstrates that a range of socio-economic factors including poverty, disparities in income distribution, unemployment, human capital conditions, and overall economic performance play a meaningful role in shaping crime patterns. The strength and direction of these relationships often vary by

location and type of crime. Evidence from the pandemic period, for instance, illustrates how changes in routine activities and mobility can alter criminal opportunities [12]. In Indonesia, empirical studies show that regions with higher poverty and greater inequality tend to experience more crime. On a global scale, a recent meta-analysis concludes that although the economic magnitude of inequality's effect on crime is relatively modest, it remains statistically significant and varies considerably across studies [13].

3. Poverty and Social Assistance Beneficiaries

Poverty is closely associated with heightened vulnerability to crime due to limited access to economic resources, increased exposure to unsafe environments, and weakened community oversight. Evidence from various Indonesian provinces demonstrates that higher levels of poverty and income inequality are linked to higher crime rates. Households registered as recipients of government social welfare programs such as conditional cash transfers are generally those experiencing economic hardship. The number of such beneficiary households can therefore serve as an indicator of spatial socio-economic vulnerability, particularly when detailed micro-level poverty data are unavailable [14]. Using both poverty counts and beneficiary-household data allows researchers to distinguish baseline poverty levels from the extent of social-assistance coverage, which may influence crime differently across regions.

4. Open Unemployment Rate

Higher unemployment tends to increase the likelihood of certain criminal activities because financial strain grows while informal social control weakens. Empirical findings from the early pandemic period in the United States show that abrupt increases in unemployment were followed by rises in gun violence and homicide [8]. Studies conducted in France and the United Kingdom also document positive relationships between unemployment and specific forms of crime, especially during economic downturns [15]. For this reason, the unemployment rate remains a central explanatory factor in crime research and should be interpreted based on local economic conditions and crime types. When incorporated into spatial models, it helps reveal how unemployment's influence varies geographically.

5. Number of Poor People

The total population living in poverty is a key indicator of the socio-economic condition of an area. The national statistical authority defines individuals as poor if their monthly per-capita expenditure falls below the poverty threshold. A larger poor population implies restricted access to essential needs such as food, clothing, education, and healthcare and reflects deeper structural disadvantages that may contribute to higher crime risks.

6. Population Density

Areas with high population density often experience greater exposure to crime because dense living conditions can intensify social interaction, competition for resources, and pressure on public facilities. Research in Java indicates that population density has a positive association with crime, both directly and through its links with poverty and unemployment. This relationship remains significant in East Java [9], although its magnitude varies depending on local welfare conditions and government interventions [16]. Studies from India and San Andres Island show similar patterns, with density interacting with factors such as literacy rates, income inequality, and social control [17]. These findings highlight the importance of population density as a core explanatory variable and support the use of spatial methods such as geographically weighted regression to capture non-uniform effects across locations [18].

7. Gini Index (Income Inequality)

Income inequality is theorized to generate feelings of relative deprivation, which can heighten tension, social conflict, and ultimately crime [7]. Empirical evidence aligns with this theory: higher levels of income inequality are associated with increased firearm-related homicides in the United States even after controlling for poverty and other socio-economic variables and contribute cumulatively to homicide rates across nations. Similar patterns are observed in Indonesia, where provinces with greater inequality tend to report more crime, although statistical significance differs across studies [10]. These observations confirm the Gini Index as a crucial indicator in both international and national crime research.

8. Economic Growth

Economic growth, typically assessed through changes in regional economic output, is generally expected to lower crime by generating employment, increasing incomes, and improving access to welfare services. However, the relationship is not always straightforward. Higher economic activity can also create additional opportunities

for crime, particularly in rapidly expanding urban environments [19]. Studies across Indonesian provinces show that unemployment consistently affects crime, whereas economic growth and income inequality do not always display significant effects. In East Java, economic growth tends to reduce crime, but the extent of its influence varies depending on local socio-economic conditions [14]. This indicates that although economic growth remains relevant, its impact is shaped by the inclusiveness and distribution of development.

2.2. Research Methodology

After the data are collected, the analysis proceeds as follows:

1. Perform data preprocessing on both the dependent and independent variables by merging the relevant BPS publication sources into a single Excel file.
2. Once the dataset is organized, the number of crime incidents for each district or city in East Java is modeled and estimated.
3. Detect multicollinearity
Assess multicollinearity among predictors using Variance Inflation Factor (VIF). A VIF greater than 10 signals serious multicollinearity issues. The VIF for the j -th predictor is:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (1)$$

where R_j^2 is the coefficient of determination of the regression of variable X_j against other independent variables.

4. Fit a Poisson regression model and check for overdispersion
A Poisson regression model is then fitted, expressed as:

$$y_i = \exp(x_i^T \beta) + \varepsilon_i \quad (2)$$

where μ represents the expected number of occurrences, β denotes the set of regression parameters, and ε_i refers to the random error component. A fundamental assumption of the Poisson model is that the mean and variance of the response variable are identical, a condition known as equidispersion. According to [20], deviations from this assumption referred to as overdispersion can occur when the data contain many zero counts, when there is unaccounted-for variation among observations, or when events are correlated. To assess whether the equidispersion assumption holds, the Pearson chi-square statistic is divided by the corresponding degrees of freedom:

$$\phi = \frac{X^2}{n - p - 1}, \quad X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\text{var}(\hat{\mu}_i)} \quad (3)$$

5. Negative Binomial Regression Analysis

If the Poisson model fails to accommodate overdispersion, a Negative Binomial model is used instead [21]. The Poisson–Gamma mixture yields:

$$E[Y] = \mu \quad \text{and} \quad V[Y] = \mu + \theta\mu^2 \quad (4)$$

and the probability mass function becomes:

$$f(y, \mu, \theta) = \frac{\Gamma(y + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta})y!} \left(\frac{\theta\mu}{1 + \theta\mu} \right)^y \left(\frac{1}{1 + \theta\mu} \right)^{\frac{1}{\theta}} \quad \text{with } y = 0, 1, 2, \dots \quad (5)$$

Parameter estimation, joint significance testing, and partial significance testing are performed in R.

6. Spatial Assumption Testing

Spatial dependence occurs when the value of an observation at one location is influenced by observations at nearby locations. To detect spatial dependence, Moran's I statistic is used [22]. The hypotheses for Moran's

I are:

$$\begin{aligned} H_0 : I &= 0 \quad (\text{no spatial dependency}) \\ H_1 : I &\neq 0 \quad (\text{there is spatial dependency}) \end{aligned}$$

The statistical test method applied is:

$$Z_{\text{statistic}} = \frac{\hat{I} - E(\hat{I})}{\sqrt{\text{var}(\hat{I})}} \quad (6)$$

With

$$\hat{I} = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Here, y_i represents the observed value at location i , y_j the value at location j , $\bar{y}$ is the mean of all observations; w_{ij} is the spatial weight describing proximity between locations; and n denotes the total number of regions. Spatial heteroskedasticity is examined to determine whether the variability of errors differs across locations. The Breusch–Pagan (BP) test is applied with the following hypotheses:

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \quad (\text{variance between locations is the same}) \\ H_1 : &\text{at least one } \sigma_i^2 \neq \sigma^2 \quad (\text{variance between locations is different}) \end{aligned}$$

Test statistic:

$$BP = \left(\frac{1}{2} \right) f^T Z (Z^T Z)^{-1} Z^T f \quad (8)$$

Decision: reject H_0 if the value of $BP > \chi_{(\alpha, k)}^2$ or p-value $< \alpha$, which means that heteroscedasticity occurs in the model.

7. Geographically Weighted Negative Binomial Regression (GWNBR)

Geographically Weighted Negative Binomial Regression is a statistical approach used to model data with spatial heterogeneity in cases where there is overdispersion. GWNBR originates from a negative binomial regression model that takes into account local parameters that are not the same for each specific region. Compared to alternative spatial approaches such as the spatial lag model and spatial error model, which impose global spatial structures on the data, GWNBR allows each location to have its own set of regression coefficients, capturing spatial non-stationarity more flexibly. Unlike machine learning methods (e.g., random forests or gradient boosting), GWNBR preserves interpretability of local coefficients and provides inferential statistics for each spatial unit, which is essential for evidence-based policy formulation. The mathematical structure of the GWNBR model is described as follows [23]:

$$y_i \sim NB \left[\exp \left(\sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} \right), \theta(u_i, v_i) \right], \quad i = 1, 2, 3, \dots, n \quad (9)$$

where,

y_i : Observation value to i

x_{ik} : Observed value of the k^{th} independent variable in the observation area (u_i, v_i)

$\beta_k(u_i, v_i)$: Regression coefficient of the k^{th} independent variable for each region (u_i, v_i)

$\theta(u_i, v_i)$: Dispersion parameter of the region (u_i, v_i)

Compute Euclidean distances between observation locations:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (10)$$

Determine the optimal bandwidth via cross-validation using Fixed Gaussian, Fixed Bisquare, Adaptive Gaussian, and Adaptive Bisquare kernels:

$$CV(b) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(b))^2 \quad (11)$$

The spatial weighting function used is the kernel function. This study applies several kernel functions, including:

Fixed Gaussian Kernel Function:

$$w_{ij}(u_i, v_i) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{h} \right)^2 \right] \quad (12)$$

Fixed Bisquare Kernel Function:

$$w_{ij}(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h} \right)^2 \right]^2, & \text{for } d_{ij} \leq h \\ 0, & \text{for } d_{ij} > h \end{cases} \quad (13)$$

Adaptive Gaussian Function:

$$w_j(u_i, v_i) = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{h_i} \right)^2 \right] \quad (14)$$

Adaptive Bisquare Function:

$$w_{ij}(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h_i} \right)^2 \right]^2, & \text{for } d_{ij} \leq h_i \\ 0, & \text{for } d_{ij} > h_i \end{cases} \quad (15)$$

8. Model Comparison Test

$$F_{statistic} = \frac{\text{deviance Model A}/df_A}{\text{deviance Model B}/df_B} \quad (16)$$

Suppose model A is a negative binomial model and model B is a GWNBR model. Reject H_0 if $F_{count} > F_{(\alpha, df_B)}$, which means that there is a significant difference between the negative binomial model and the GWNBR model, so it is necessary to test the GWNBR model parameters simultaneously and partially [24]

9. Calculate the estimated parameters of the GWR model using the optimal kernel function.
10. Parameter Significance Testing

$$t_{statistic} = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))} \quad (17)$$

H_0 is rejected if the test statistic $|t_{count}| > t_\alpha$. Rejecting H_0 means that the parameter has a significant effect on the response variable at each location.

11. Model Selection Using AIC and dispersion

To determine the best model, the Akaike Information Criterion is computed:

$$AIC = 2p - 2 \ln \hat{L}(\hat{\theta}) \quad (18)$$

where $\hat{L}$ is the log-likelihood value.

12. Interpretation of Results

The final step involves interpreting local coefficient estimates, identifying spatial patterns, and comparing regional variations in crime determinants.

3. Results and Discussion

3.1. Descriptive Statistics of Crime Counts and Their Potential Determinants

Descriptive statistics for crime counts and socio-economic indicators are presented in Table 2.

Table 2. Descriptive Statistics of Research Variables

Variable	Mean	Minimum	Maximum
Y	1,581.1	469	9,252
X_1	310,688	26,287	797,236
X_2	4.041	1.560	6.490
X_3	104.81	6.59	240.14
X_4	1,970	411	8,727
X_5	0.3318	0.2330	0.4350
X_6	4.631	1.670	5.760

Average crime incidents across districts and cities in East Java total 1,581 cases. . The lowest crime count is observed in Pacitan Regency with 469 cases, while Surabaya City registers the highest figure at 9,252 cases, reflecting its status as the province's largest urban center.

The mean number of households receiving government social assistance is 310,688, with Batu City having the smallest value (26,287 households) and Jember Regency having the largest (797,236 households). The unemployment rate across the region averages 4.04 percent, ranging from 1.56 percent in Pacitan Regency to 6.49 percent in Sidoarjo Regency.

For the poverty indicator, the average number of residents classified as poor is approximately 104.81 thousand people. Batu City records the lowest level (6.59 thousand), whereas Malang Regency has the highest (240.14 thousand). Population density varies considerably, with an average of 1,970 persons per square kilometer. Pacitan Regency shows the lowest density (411 persons/km²), while Surabaya again stands out with the highest density (8,727 persons/km²). Income inequality, reflected by the Gini index, averages 0.3318. Sampang Regency reports the lowest inequality (0.2330), whereas Madiun City records the highest (0.4350). Finally, the economic growth rate averages 4.63 percent, with Bojonegoro Regency at the lower end (1.67 percent) and Surabaya City achieving the highest growth rate at 5.76 percent.

Overall, these descriptive results illustrate substantial variation across districts and cities in East Java, suggesting that socio-economic conditions differ widely and may contribute to differing crime patterns across the region.

3.2. Multicollinearity Test

Multicollinearity was assessed using the Variance Inflation Factor (VIF), and the results are summarized in Table 3.

Table 3. VIF Values of Predictor Variables

	X_1	X_2	X_3	X_4	X_5	X_6
VIF	8.199	1.282	5.333	4.290	2.288	2.080

The VIF values for all predictors are below the commonly accepted threshold of 10. This indicates that none of the variables exhibit problematic collinearity, and all predictors can be retained for subsequent model estimation.

3.3. Poisson Regression Modeling

Following the multicollinearity assessment, a Poisson regression model was estimated to examine the relationship between crime counts and the explanatory variables. The parameter estimates are presented in Table 4.

Table 4. Estimation and Evaluation of Parameters in the Poisson Regression Model

Parameter	Estimation	Std. Error	Z_{Score}	P-Value	Significant
$\hat{\beta}_0$	5.120e+00	5.280e-02	96.970	< 2e-16	Significant
$\hat{\beta}_1$	3.196e-07	5.761e-08	5.547	2.9e-08	Significant
$\hat{\beta}_2$	1.005e-01	3.892e-03	25.824	< 2e-16	Significant
$\hat{\beta}_3$	7.684e-03	1.400e-04	54.891	< 2e-16	Significant
$\hat{\beta}_4$	1.912e-04	2.872e-06	66.579	< 2e-16	Significant
$\hat{\beta}_5$	3.303e+00	1.633e-01	20.231	< 2e-16	Significant
$\hat{\beta}_6$	3.191e-01	6.923e-03	6.090	< 2e-16	Significant
Deviance = 8017.8			df = 31		
AIC = 8373.365					

Initial Poisson regression yielded deviance of 8017.8 with AIC of 8373.365. All predictors showed statistical significance at 5% level. However, dispersion statistic $\phi = 258.639$ substantially exceeds 1, indicating severe overdispersion and violating Poisson assumptions. Based on Table 4, the deviance value obtained is $D(\hat{\beta}) = 8017.8$. At a 5% significance level, the critical chi-square value is $\chi^2_{(6;0.05)} = 12.592$. H_0 is rejected, indicating that at least one predictor variable significantly affects the response variable. At the same significance level (5%), the Z-critical value is $Z_{0.025} = 1.959$. According to Table 4, the $|Z\text{-values}|$ of parameters β_1 to β_6 are greater than 1.959, confirming that these predictors have partial significant effects on crime levels in East Java. This means that one or more variables number of beneficiary families (X_1), open unemployment rate (X_2), number of poor residents (X_3), population density (X_4), Gini ratio (X_5), and economic growth (X_6) have a significant influence on the number of crime cases in East Java. Based on these results, the following Poisson regression model can be formulated:

$$\ln(\hat{\mu}) = 5.120 + 3.196 \times 10^{-7}X_1 + 0.1005X_2 + 0.007684X_3 + 0.0001912X_4 + 3.303X_5 + 0.3191X_6 \quad (19)$$

3.4. Overdispersion Detection

The dispersion statistic was computed as:

$$\phi = \frac{8017.8}{31} = 258.639 \quad (20)$$

Because the value of ϕ is substantially greater than 1, the data clearly indicate a strong presence of overdispersion. This condition contradicts the core assumption of the Poisson model, which requires the mean and variance to be equal. Consequently, the Poisson model tends to underestimate the standard errors, leading to significance tests that cannot be trusted. For this reason, the Negative Binomial model provides a more suitable alternative.

3.5. Negative Binomial Regression Modeling

Negative Binomial regression was estimated using maximum likelihood with Newton–Raphson iteration. The results are given in Table 5.

Table 5. Estimation and Testing of the Negative Binomial Regression Model

Parameter	Estimation	Std. Error	Z_{Score}	P-Value	Significant
$\hat{\beta}_0$	4.521e+00	6.804e-01	6.644	3.06e-11	Significant
$\hat{\beta}_1$	1.158e-06	7.709e-07	1.502	0.133048	Not Significant
$\hat{\beta}_2$	1.368e-01	5.242e-02	2.610	0.009054	Significant
$\hat{\beta}_3$	5.870e-03	1.945e-03	3.019	0.002540	Significant
$\hat{\beta}_4$	1.331e-04	3.933e-05	3.384	0.000713	Significant
$\hat{\beta}_5$	8.440e-01	2.034e+00	-0.415	0.678208	Not Significant
$\hat{\beta}_6$	2.545e-01	7.737e-02	3.289	0.001004	Significant
Deviance = 38.661			df = 31		
AIC = 586.9489					

The simultaneous test aims to evaluate whether all predictor variables collectively influence the response variable. Based on Table 5, the deviance value obtained is $D(\hat{\beta}) = 38.661$, which exceeds the chi-square critical value $\chi^2_{(6;0.05)} = 12.592$. Therefore, H_0 is rejected, indicating that at least one predictor variable significantly affects the crime rate in East Java.

Partial testing is conducted to determine the significance of each individual predictor. Using a 5% significance level with $Z_{0.025} = 1.959$, the results show that $|Z|$ values for β_1 and β_5 are smaller than 1.959, while $|Z|$ values for $\beta_2, \beta_3, \beta_4$, and β_6 exceed 1.959. Thus, the number of beneficiary families (X_1) and the Gini ratio (X_5) have no significant effect, whereas the open unemployment rate (X_2), number of poor residents (X_3), population density (X_4), and economic growth (X_6) significantly influence the crime rate in East Java. The resulting negative binomial regression model can be expressed as follows:

$$\ln(\hat{\mu}) = 4.521 + 1.158 \times 10^{-6}X_1 + 0.1368X_2 + 0.00587X_3 + 0.0001331X_4 + 0.844X_5 + 0.2545X_6 \quad (21)$$

3.6. Spatial Diagnostics

The spatial dependence test was conducted to determine whether neighboring regions share similar characteristics. Using Moran's I statistic, the result yielded $Z_I = 3.548$. At a 5% significance level, $Z_{0.025} = 1.959$, leading to the rejection of H_0 . This indicates the presence of spatial dependence among the observed regions.

Spatial heterogeneity was examined using the Breusch–Pagan (BP) test to assess variation across locations. The test produced a BP value of 18.4829, which exceeds the chi-square critical value $\chi^2_{(6;0.05)} = 12.592$. Thus, H_0 was rejected, confirming spatial heterogeneity among the study areas.

With both spatial dependence and heterogeneity confirmed, the next analytical step involves applying the GWNBR model to account for spatial effects.

3.7. Model Comparison between Geographically Weighted Negative Binomial Regression(GWNBR) and Negative Binomial Regression

Based on the results obtained using R software at a 5% significance level, the F-statistic value was 3.2824. The critical region for rejecting H_0 occurs when $F > F_{(0.05;31;31)}$. Since $3.2824 > 1.822132$, H_0 is rejected. Therefore, it can be concluded that the GWNBR model provides a better fit than the standard Negative Binomial Regression model in explaining the distribution of crime cases in East Java.

Table 6. CV Value for Each Kernel Function

Kernel Function	Adaptive Gaussian	Adaptive Bisquare	Fixed Gaussian	Fixed Bisquare
CV	83,226,583	83,122,915	83,111,100	83,189,398
Optimal Bandwidth (h)	0.7894675	0.5012338	2.126352	3.339279

Based on the results in Table 6, the minimum CV value of 83,111,100 was obtained, indicating that the Fixed Gaussian kernel with an optimal bandwidth of 2.126352 is the best-performing function. The sensitivity of results to kernel function choice constitutes part of the robustness of the present analysis: all four kernel functions were evaluated, and the Fixed Gaussian kernel was selected on the basis of minimum cross-validation score. The consistency of model rankings across kernels provides confidence that the superiority of GWNBR over the global Negative Binomial model is not an artifact of a particular weighting scheme.

3.8. Simultaneous Test

Based on the results obtained using R software at a 5% significance level, the deviance value was $D(\hat{\beta}) = 18.0463$. The critical chi-square value at the same significance level is $\chi^2_{(6;0.05)} = 12.592$. Since $D(\hat{\beta}) = 18.0463 > 12.592$, H_0 is rejected. Therefore, it can be concluded that at least one parameter in the GWNBR model has a significant effect on the formed model.

3.9. Partial Test

The results of the parameter significance test indicate differences in parameter values across districts. This is due to the varying significance of variables produced by each weighting scheme used in the model. The significance test was conducted at a 5% significance level, with the critical region for rejecting H_0 determined by $|Z| > Z_{(\alpha/2)}$ or $|Z| > 1.959$. Table 7 presents the significant variables at each location.

Table 7. List of Significant Predictor Variables at Observation Locations

Location	Significant Variable
Batu City, Blitar Regency, Magetan Regency, Mojokerto Regency, Trenggalek Regency, Tuban Regency	X_4
Madiun City, Malang City, Ngawi Regency, Probolinggo Regency, Sidoarjo Regency, Situbondo Regency	X_4, X_5
Bojonegoro Regency, Bondowoso Regency, Kediri Regency, Mojokerto City, Pasuruan City, Lumajang Regency, Tulungagung Regency	X_2, X_4, X_5
Blitar City, Kediri City, Lamongan Regency, Nganjuk Regency, Pacitan Regency, Ponorogo Regency	X_1, X_2, X_4, X_5
Banyuwangi Regency, Jombang Regency, Probolinggo City, Sampang Regency, Sumenep Regency	X_2, X_4, X_5, X_6
Gresik Regency, Jember Regency, Madiun Regency, Malang Regency	X_1, X_2, X_3, X_4, X_5
Bangkalan Regency, Pasuruan Regency, Pamekasan Regency, Surabaya City	$X_1, X_2, X_3, X_4, X_5, X_6$

3.10. Model Interpretation

As an example of model interpretation, one of the regions where all variables in the model show significant effects is Surabaya City. The model for the number of crime cases in Surabaya City can be expressed as follows:

$$\hat{\mu} = \exp(0.338 + 0.0014X_1 + 0.0166X_2 + 0.0059X_3 + 0.0129X_4 + 0.0128X_5 + 0.0140X_6) \quad (22)$$

An increase of 1,000 beneficiary households (X_1) in Surabaya City is estimated to increase the number of crime cases by $\exp(0.001400) = 1.0014$ times from the previous level, assuming other variables remain constant. This indicates that a higher number of households dependent on social assistance may serve as an indicator of economic vulnerability, which in turn can drive the occurrence of criminal acts, particularly property crimes or petty theft.

An increase of 1 percent in the open unemployment rate (X_2) is estimated to raise the number of crime cases by $\exp(0.016625) = 1.01676$ times from the previous level. This positive effect suggests that the growing number of unemployed individuals in a large urban area such as Surabaya may heighten the risk of crime, in line with increased economic pressure and intense job competition.

An increase of 1,000 individuals in the poor population (X_3) is estimated to increase the number of crime cases by $\exp(0.005859) = 1.00588$ times from the previous level. This means that a larger low-income population corresponds to a higher potential for criminal activities, especially those driven by urgent economic needs.

An increase of one person per square kilometer (X_4) is estimated to raise the number of crime cases by $\exp(0.012911) = 1.01299$ times. High population density in urban areas such as Surabaya creates a complex social environment with higher potential for conflict and crime due to increased mobility and intense human interactions.

An increase of 0.01 points in the Gini ratio index (X_5) is estimated to increase the number of crime cases by $\exp(0.012790) = 1.01287$ times. This finding suggests that income inequality continues to contribute to higher crime rates. In Surabaya, the disparity between upper and lower economic classes may trigger feelings of social injustice that elevate the likelihood of property-related crimes.

An increase of 1 percent in economic growth (X_6) is estimated to raise the number of crime cases by $\exp(0.013961) = 1.01406$ times. The positive association between economic growth and crime in Surabaya warrants theoretical elaboration. Several mechanisms may explain this counterintuitive finding. First, rapid economic expansion in a metropolitan context generates increased flows of goods, people, and cash, creating more abundant and accessible criminal targets as predicted by routine activity theory. Second, growth-driven migration intensifies competition for housing, employment, and public services, potentially weakening community cohesion and informal social control as described by social disorganization theory. Third, economic growth in Surabaya has historically been concentrated in the formal sector and technology-intensive industries, bypassing low-skilled workers and thereby widening internal inequality even as aggregate output rises. This inequality amid growth may fuel feelings of relative deprivation that increase the propensity toward property crime. These mechanisms are broadly consistent with evidence from other rapidly urbanizing cities in developing economies [19].

3.11. Best Model Selection using AIC and Dispersion Values

Model selection was carried out using AIC and the dispersion value to evaluate overall fit and variance stability. A lower AIC indicates a more efficient model, while a dispersion value close to one reflects an appropriate handling of variability in count data.

Table 8 shows that the Poisson model performs poorly, producing extremely high dispersion and confirming severe overdispersion. The Negative Binomial model substantially improves the fit, yielding a dispersion value close to unity. The Geographically Weighted Negative Binomial Regression (GWNBR) model demonstrates the strongest performance, with the lowest AIC and the most stable dispersion estimate. These results indicate that incorporating spatial heterogeneity yields a markedly superior representation of crime patterns in East Java, making GWNBR the most suitable model among the three alternatives.

Table 8. Model Performance Comparison

Model	AIC	Dispersion
Poisson Regression	8373.365	258.639
Negative Binomial Regression	586.9489	1.247
GWNBR	406.6973	0.582

The results clearly show a progressive improvement from the Poisson model to the Negative Binomial and further to the Geographically Weighted Negative Binomial model. The very high dispersion value under the Poisson specification indicates that it cannot adequately represent the variability in crime counts. The Negative Binomial model significantly reduces dispersion, confirming that its additional variance parameter successfully addresses overdispersion. However, the GWNBR model offers the strongest performance overall, producing both the lowest AIC value and the most stable dispersion estimate. This demonstrates that accounting for spatial heterogeneity is crucial in modeling crime behavior across East Java, and that a geographically weighted approach yields a more accurate and context-sensitive representation of local crime patterns.

4. Conclusion

The findings of this research highlight that both spatial dependence and spatial heterogeneity play a substantial role in shaping the pattern of criminal incidents across districts and municipalities in East Java. The Poisson regression model initially applied was unable to accommodate the wide variation present in the data, as reflected by pronounced overdispersion. Although the Negative Binomial model provided an improvement by handling the excess variance, it still left spatial patterns unexplained. The diagnostic evaluations particularly Moran's I and the Breusch–Pagan test confirmed that spatial effects are inherent in the data and must be incorporated into the analytical framework. Consequently, the GWNBR model was estimated using a fixed Gaussian kernel with an optimal bandwidth of 2.126352. This model achieved the best overall fit, demonstrated by the lowest AIC value (406.6973), indicating that locally varying parameter structures offer a far more accurate representation of crime distribution than global count models.

The localized coefficients produced by the GWNBR model show that several socio-economic indicators namely the unemployment rate, the number of residents living below the poverty line, population density, and the rate of economic growth tend to increase crime levels in many areas. In contrast, the number of households receiving social assistance and the Gini index do not exhibit consistent effects across all locations. Surabaya serves as a clear example of how multiple variables can simultaneously exert significant influence, reflecting the complexity of social and economic conditions in large, densely populated metropolitan environments. These results underscore that crime determinants vary meaningfully from one region to another, and that a single global model would overlook important spatial nuances.

From a policy perspective, these insights suggest that crime-prevention strategies should be tailored to the conditions of individual districts rather than implemented uniformly across the province. Regions where unemployment strongly correlates with crime should prioritize programs that stimulate job creation, develop vocational skills, and expand employment opportunities. In East Java, this is particularly relevant for districts in the northern coastal corridor and Madura Island, where industrial employment opportunities remain limited and youth unemployment is above the provincial average. Existing programs such as the Kartu Prakerja (pre-employment card) and vocational training centers operated under the East Java Manpower Office could be expanded and better targeted toward these high-crime, high-unemployment districts. Areas where poverty acts as the dominant factor would benefit from strengthening social protection measures, improving access to welfare programs, and enhancing economic empowerment initiatives. In densely populated urban environments, urban-planning interventions such as improved spatial design, infrastructure management, and public-space utilization can help reduce pressures that contribute to heightened crime risk. The Surabaya metropolitan planning authority

could coordinate with district governments in the Gerbangkertosusila corridor to design inclusive urban growth strategies that reduce inequality amid expansion, consistent with the findings on the positive growth-crime nexus. Meanwhile, districts where rapid economic expansion appears linked to increases in crime need policies that promote inclusive growth and prevent socio-economic disparities from widening. Coordinated implementation of these geographically differentiated policies requires multi-stakeholder engagement, including local governments, community organizations, law enforcement agencies, and the private sector. Institutional barriers such as limited local fiscal capacity, fragmented coordination between district and provincial authorities, and low community participation in program design represent practical challenges that should be addressed in future policy planning.

Several limitations of this study merit acknowledgment. First, the analysis relies on a single cross-sectional year (2024), which precludes examination of temporal trends or causal identification of the socio-economic determinants of crime. Second, the use of aggregate district-level data may mask within-district spatial heterogeneity, a concern known as the modifiable areal unit problem (MAUP), whereby results can be sensitive to the chosen level of spatial aggregation. Third, potential omitted variables such as education levels, policing density, and social capital indicators were not included due to data limitations, and their exclusion may introduce bias in coefficient estimates. Fourth, the GWNBR model assumes that spatial non-stationarity is captured entirely through geographically varying coefficients, but does not model spatial autocorrelation in the error term explicitly, as spatial error or spatial lag models do.

Methodologically, future research may incorporate spatial-temporal models to observe changes in crime behavior over time, compare the performance of GWNBR with other spatial count models including zero-inflated variants, and investigate potential interactions among socio-economic factors such as the interplay between population density and unemployment. Expanding the set of explanatory variables such as education, policing capacity, public-service accessibility, or measures of social cohesion could further enrich understanding of local crime dynamics. Continued model validation, including alternative kernel functions and bandwidth strategies, is encouraged to ensure robustness. Overall, this study demonstrates that integrating spatial heterogeneity through GWNBR yields more precise and context-sensitive insights, enabling the development of targeted and evidence-based crime-prevention policies in East Java.

Acknowledgement

The authors extend their deep appreciation to Universitas Airlangga for providing academic support and research facilities essential to the completion of this study. Sincere thanks are also conveyed to Statistics Indonesia (BPS) of East Java Province for supplying the official data that formed the basis of the analysis. Their contributions were crucial in enabling the development of the spatial modeling and empirical findings presented in this research.

REFERENCES

1. B. C. Welsh and E. L. Piza, *Crime Prevention for Social Impact and Social Justice: The Ideals, the Challenges, the Progress*, The ANNALS of the American Academy of Political and Social Science, vol. 714, no. 1, pp. 8–21, 2024.
2. E. Itskovich, *Economic Inequality, Relative Deprivation, and Crime: An Individual-Level Examination*, Justice Quarterly, vol. 42, no. 4, pp. 637–658, 2024.
3. L. Sugiharti, R. Purwono, M. A. Esquivias, and H. Rohmawati, *The nexus between crime rates, poverty, and income inequality: a case study of Indonesia*, Economies, vol. 11, no. 2, 2023.
4. L. E. Cohen and M. Felson, *Social change and crime rate trends: a routine activity approach*, American Sociological Review, vol. 44, no. 4, pp. 588–608, 1979.
5. C. R. Shaw and H. D. McKay, *Juvenile Delinquency and Urban Areas*, University of Chicago Press, Chicago, 1942.
6. D. Anozhi and B. Novianda, *Socio-Economic and Property Crime Rate in Indonesia*, Economics Development Analysis Journal, vol. 12, no. 3, pp. 305–318, 2023.
7. A. Rowhani-Rahbar, D. A. Quistberg, E. R. Morgan, A. Hajat, and F. P. Rivara, *Income inequality and firearm homicide in the US: a county-level cohort study*, Injury Prevention, vol. 25, suppl. 1, pp. i25–i30, 2019.
8. J. P. Schleimer, V. A. Pear, C. D. McCort, A. B. Shev, A. D. Biasi, E. Tomsich, S. Buggs, H. S. Laqueur, and G. J. Wintemute, *Unemployment and crime in US cities during the coronavirus pandemic*, Journal of Urban Health, vol. 99, no. 1, pp. 82–91, 2022.

9. F. P. Anggaresa and Suryana, *Unraveling crime dynamics in Indonesia: exploring the impact of population density, poverty, average years of schooling, and open unemployment rate in Java in 2020*, *Welfare: Jurnal Ilmu Kesejahteraan Sosial*, vol. 12, no. 1, pp. 33–49, 2023.
10. C. Vilalta, E. Cadena, C. Garrocho, and G. Fondevila, *Beyond the immediate effects of income inequality on homicide rates: a reply to Daly's critique*, *Evolution and Human Behavior*, vol. 45, no. 4, 106597, 2024.
11. United Nations Office on Drugs and Crime, *International Classification of Crime for Statistical Purposes (ICCS): Implementation Manual*, UNODC, Vienna, 2024.
12. P. R. Estévez-Soto, *Crime and COVID-19: effect of changes in routine activities in Mexico City*, *Crime Science*, vol. 10, no. 1, pp. 1–17, 2021.
13. M. Pazzona, *Revisiting the income inequality-crime puzzle*, *World Development*, vol. 176, 106510, 2024.
14. Central Statistics Agency or Badan Pusat Statistik (BPS), *Number of Beneficiary Households and Food Social Assistance Budget by Province 2024*, 2025. [Online]. Available: <https://www.bps.go.id>
15. F. Jawadi, S. K. Mallick, A. I. Cheffou, and A. Augustine, *Does higher unemployment lead to greater criminality? Revisiting the debate over the business cycle*, *Journal of Economic Behavior & Organization*, vol. 182, pp. 448–471, 2021.
16. A. Amirusholihin, L. Rahadiantino, A. Nilasari, D. Y. Rakhmawati, and F. Fatoni, *How population density and welfare affect crime rates: a study in East Java Province, Indonesia*, *Revista de Gestão Social e Ambiental*, vol. 18, no. 8, e06224, 2024.
17. J. Saini and V. Srivastava, *Impact of population density and literacy levels on crime in India*, *Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–7, 2019.
18. S. Castell-Britton, *The relationship between overpopulation and crime rates in San Andres Island*, *Qeios*, 2024.
19. S. D. Purnomo, D. A. Supriyo, R. Rusito, T. Anindito, W. Hariadi, and D. Jati, *How economic indicators drive crime? Empirical study in a developing country, Indonesia*, *International Journal of Economics and Financial Issues*, vol. 13, no. 3, pp. 94–99, 2023.
20. J. M. Hilbe, *Negative Binomial Regression*, 2nd ed., Cambridge University Press, Cambridge, 2011.
21. A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data*, Cambridge University Press, Cambridge, 1998.
22. L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, 1988.
23. T. Saifudin, N. Chamidah, N. Azizah, F. S. Renianti, N. Carista, I. Y. Adyatma, and I. M. Tianda, *Geographically Weighted Negative Binomial Regression for Spatial Count Data: Methodology and Application*, *MethodsX*, vol. 14, Art. 103919, 2026.
24. A. R. da Silva and T. C. V. Rodrigues, *Geographically weighted negative binomial regression – incorporating overdispersion*, *Statistics and Computing*, vol. 24, no. 5, pp. 769–783, 2014.