BAREKENG: Journal of Mathematics and Its Applications
September 2025 Volume 19 Issue 3 Page 1469-1484
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss3pp1469-1484

GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION AND GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION MODELING ON PROPERTY CRIME CASES IN CENTRAL JAVA

Prizka Rismawati Arum[1], Rahmad Putra Gautama[2*], M. Al Haris[3]

[1,2,3]Departement of Statistic, Faculty of Science and Agricultural Technology, Universitas Muhammadiyah Semarang
Jln. Kedungmundu No.18, Kedungmundu, Kota Semarang, Jawa Tengah 50273 Indonesia

*Corresponding author's e-mail: * **rahmadgautama15@gmail.com***

ABSTRACT

Article History:

Received: 29th August 2024
Revised: 21st November 2024
Accepted: 3rd January 2025
Published: 1st July 2025

Keywords:

Adaptive Bisquare Kernel; Geographically Weighted Generalized Poisson Regression; Geographically Weighted Negative Binomial Regression; Overdispersion; Property Crime

Property crime in Indonesia remains one of the most prevalent categories of crime across various regions of the country. This category encompasses a range of criminal acts, including theft, illegal appropriation of goods, robbery, motor vehicle theft, arson, and property damage. One of the commonly used regression analysis methods is Poisson regression. The assumption violation of overdispersion in Poisson regression is often found in property crime data in Central Java. This study also considers spatial aspects, depicting local regional characteristics and the integration of local and global variables. Therefore, this study employs Geographically Weighted Generalized Poisson Regression (GWGPR) and Geographically Weighted Negative Binomial Regression (GWNBR) methods with Adaptive Bisquare Kernel weighting. The aim of this research is to develop a model for each district/city in Central Java using Adaptive Bisquare Kernel weighting, thus providing a more accurate representation of the factors influencing property crime in each region. The AIC value criterion of 411.3652 indicates that the GWNBR method is the most suitable for modeling the number of property crime cases in each district/city in Central Java compared to Poisson regression, negative binomial regression, and GWGPR methods.

How to cite this article:
P. R. Arum, R. P. Gautama and M. A. Haris., "GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION AND GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION MODELING ON PROPERTY CRIME CASES IN CENTRAL JAVA," *BAREKENG: J. Math. & App.*, vol. 19, no. 3, pp. 1469-1484, September, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Criminal acts are not merely violations of the law but also attacks on societal norms that undermine the foundation of social order. In science, criminologists offer a startling perspective, depicting deviant behavior not just as a violation but as a real threat that must be unraveled through a structural understanding of society. His analysis highlights this phenomenon as a result of inequality in the distribution of power, authority, and wealth, further exacerbated by economic and political changes. Essential factors such as education level, job opportunities, and the characteristics of offenders also shape crime patterns. Equally important, population density, police patrol effectiveness, environmental conditions, and the intensity of neighborhood watch programs play a significant role in shaping the crime landscape. Changes in the times contribute to the complexity of societal life. Rapid technological advancements trigger urbanization processes in the context of modern society, necessitating change [1].

Property crime encompasses various criminal acts such as theft, unlawful appropriation of goods, depriving others of their rights, motor vehicle theft, arson, and property destruction. These offenses often involve elements of threat or violence, especially in cases of robbery or extortion. Property crime includes criminal acts committed with the intent to enrich the perpetrator. In the context of property, crimes can be divided into two main categories: damaging and stealing property. People with economic difficulties tend to be involved in such crimes. In Central Java, factors such as economic crises, income inequality, and economic injustice are major drivers of crime. Crimes that emphasizes that any criminal act aimed at enriching the perpetrator can be categorized as property crime [2]. Besides economic factors, education also plays a significant role. Low education is a characteristic of impoverished communities, thus increasing the likelihood of committing crimes to meet daily needs and sustain their existence [3]. According to the subcultural delinquency theory, criminal behavior is associated with the distinctive cultural patterns of a given community. This factor can contribute to the occurrence of crime, as the causes of crime can be influenced by the location. In Central Java, urban areas are more prone to crime compared to rural areas [4].

The research that is widely used for this model usually uses health data, for example to determine the distribution of Tuberculosis disease. Of course, this aspect is different in terms of the dimensions of the variables studied and the continuity of the case study being observed. Because there is a rare similarity between criminal cases and Tuberculosis, the methods used can be the same for both cases.

A common analytical method used to identify the factors contributing to property crime is regression analysis. Regression analysis is a technique employed to model the relationship between response variables and predictor variables, using continuous data in this context [5]. Poisson regression analysis is the most suitable regression model to model discrete data on property crime rates due to its discrete (count) data structure [6]. Poisson regression is a type of nonlinear regression analysis that uses a response variable following a Poisson distribution. This distribution represents the number of events or successes occurring within a specific period or area for a random variable.

In Poisson regression studies, it is common to encounter assumption violations where the variance is either smaller than the mean (underdispersion) or greater than the mean (overdispersion). If Poisson data contains overdispersion, then important information can be lost because the dispersion parameter is not modeled. This can result in less accurate and unreliable analysis results, affecting the conclusions of the study [7]. In dealing with overdispersed data, researchers generally use the Negative Binomial regression model. However, there are alternative methods such as Generalized Poisson Regression (GPR) that can be considered. The GPR model produces parameters that are global, applicable to all locations assuming uniform characteristics in each location [8]. In the case of property crime, the characteristics of locations vary, so this assumption is not appropriate. Spatial data analysis is therefore more appropriate, as this method is able to address variation and heterogeneity between locations more effectively, offering a deeper understanding of location-specific differences.

The Geographically Weighted Negative Binomial Regression (GWNBR) method is an extension of Negative Binomial regression designed to address overdispersion in Poisson regression by incorporating spatial effects. Negative Binomial regression is crucial because it effectively manages overdispersion, a situation where the data variance exceeds the mean—a condition that Poisson regression often struggles to handle properly [9].

In addition, Generalized Poisson Regression (GPR) and Geographically Weighted Generalized Poisson Regression (GWGPR) methods were also used in this study. GPR is a regression method that can handle both underdispersion and overdispersion, with more flexibility than standard Poisson regression. GWGPR is an extension of GPR that incorporates spatial effects, thus identifying spatial variations in the data and providing a more accurate model for data influenced by geographical factors [10].

In the GWGPR and GWNBR models, the Maximum Likelihood Estimation (MLE) method is used to estimate the parameters of these models by assigning weights to the observation locations. Maximum Likelihood Estimation (MLE) is a very important method in the context of GWGPR and GWNBR models for several interrelated reasons. This method was chosen mainly because of its optimal ability to estimate spatial parameters. MLE can well accommodate spatial weighting, handle non-linear relationships between variables, and is effective in estimating spatially varying parameters. This weighting matrix is formed from a weighting function where the function depends on the distance between observation locations or the size of the environment in other words called Bandwidth or weighting [11].

The purpose of this weighting is to interpret the parameters at each observation location. To determine the weights at various locations, this study utilized the Adaptive Bisquare Kernel function. The value of the Adaptive Bisquare Kernel function can be calculated using Adaptive Kernel Adaptive Bandwidth, which varies at each observation point. This research aims to identify the variables influencing the number of property crimes in the Central Java region, using the Adaptive Bisquare Kernel weighting function [12]. To analyze the dependent variables on the number of property crimes in the Central Java region in 2023, the author wants to know the weighting function that produces the most appropriate model parameter estimates by comparing the two methods.

2. RESEARCH METHODS

2.1 Dataset

This research utilizes secondary data as the basis for analysis. The focus of the research is on the response variable, property crime, which is taken from the 2023 Central Java Regional Police Criminal Investigation Department data covering each regency/city in Central Java. Information on the factors influencing property crime was obtained through general crime division data and district/city crime data issued by the Central Java Regional Police. Sources from the Central Bureau of Statistics website provide a wealth of information on the characteristics and dynamics of each district/city in Central Java. It includes elements such as economic growth, social structure, and other variables that may influence property crime rates. The regions that are the focus of this research cover the entire area of Central Java Province. With this approach, it is hoped that the research will be able to provide holistic insights into the factors that influence property crime at the district/city level in Central Java. Variable used $\boldsymbol{X_1}$ variable is the number of poor people (JPM), $\boldsymbol{X_2}$ open unemployment rate (TPT), $\boldsymbol{X_3}$ school enrollment rate (APS), $\boldsymbol{X_4}$ gini ratio (GR), $\boldsymbol{X_5}$ population density (KP) and $\boldsymbol{X_6}$ average net income of independent workers (PBPB). The selection of these variables collectively covers various theoretical perspectives on crime, including economic factors (JPM, TPT, PBPB, GR), social factors (APS), and environmental factors (KP). This multi-dimensional approach is in line with modern criminological research that emphasizes the importance of considering various contributing factors to understand and analyze crime patterns. By considering socioeconomic and demographic factors that have proven relevant in previous studies, this selection of variables provides a comprehensive framework for analyzing property crime in Central Java.

2.2 Multicollinearity

An effective regression model is characterized by the absence of multicollinearity. To detect the presence of multicollinearity, the Variance Inflation Factor (VIF) score is used. The test equation for VIF is as follows [13]:

$$\boldsymbol{VIF_k = \frac{1}{1 - {R_k}^2}} \tag{1}$$

$\boldsymbol{{R_k}^2}$ = Determination Coefficient Value

Multicollinearity can be identified using the Variance Inflation Factor (VIF). A VIF value below 10 suggests there is no significant correlation between predictor variables, while a VIF value above 10 indicates that a predictor variable is correlated with, or shows a relationship to, other predictor variables [14].

2.3 Poisson Regression

Poisson regression modeling is a type of Generalized Linear Model (GLM), which extends linear regression models. In Poisson regression, Equation 1 illustrates the relationship between the response variable (Y) and the predictor variable (X) as follows [15]:

$$\boldsymbol{E}((\boldsymbol{Y}_i|\boldsymbol{X}_i) = \mu_i = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \cdots + \beta_n \boldsymbol{x}_n + \varepsilon_1 \quad (2)$$

To link $E((\boldsymbol{Y}_i|\boldsymbol{X}_i)$ with the linear function ${x_i}^T\beta$ so that the relationship between the two becomes precise, the link function $link\ \eta_i = \log({x_i}^T\beta)$ or $\eta_i = \log(\mu_i) = {x_i}^T\beta$ is used. Thus, the Poisson model formed is [16]:

$$\boldsymbol{E}(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \mu_i = \boldsymbol{exp}\left({\boldsymbol{x}_i}^{\boldsymbol{T}}\boldsymbol{\beta}\right), \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{n} \quad (3)$$

$\beta =$ unknown parameters and must be estimated using the notation $\beta^T = [\beta_0\ \beta_1\ \beta_2 \dots \beta_n]$

$x_i =$ independent variable which has the notation ${x_i}^T = [1\ x_{1i} x_{2i} \dots x_{ni}]$

So that:

$$\boldsymbol{\mu_i} = \boldsymbol{exp}(\boldsymbol{\beta_0} + \boldsymbol{\beta_1 x_{1i}} + \boldsymbol{\beta_2 x_{2i}} + \dots + \boldsymbol{\beta_n x_{ni}}) \quad (4)$$

2.4 Overdispersion

When Poisson regression is applied under overdispersion conditions, it can lead to increased variability in the dependent variable [17]. To detect overdispersion, the Pearson Chi-Square value per degree of freedom is used, calculated using the following formula [18]:

$$\boldsymbol{dispersion\ ratio\ \beta} = \frac{\boldsymbol{\chi^2}}{\boldsymbol{n} - \boldsymbol{k} - \mathbf{1}} \quad (5)$$

With $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{Var(\mu_i)}$

The decision to be made is equidispersion if the dispersion ratio $= 0$, overdispersion if the dispersion ratio > 1, and underdispersion if the dispersion ratio < 0 [19].

2.5 Spatial Heterogeneity

The Breusch-Pagan test is employed to detect the presence of spatial heterogeneity. The hypothesis for the test is as follows [20]:

H_0 : Absence of spatial heterogeneity properties

H_1 : Spatial heterogeneity exists in the data

Values of the Breusch-Pagan Test

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f \sim X_P^2 \quad (6)$$

The decision is to reject H_0 and accept H_1 if BP $> x^2_{(p+1)}$. Thus, the conclusion that can be drawn when rejecting H_0 is that there is spatial heterogeneity [21].

2.6 Generalized Poisson Regression (GPR)

The Geographically Poisson Regression (GPR) model is a type of model that is suitable for count data when there is an over/under dispersion problem. GPR model model has two parameters, namely the μ parameter as well as an additional parameter known as θ, which serves as a dispersion parameter. known as θ, which serves as the dispersion parameter. distribution Generalized Poisson (GP) distribution can be explained as follows [22]:

$$f(y;\mu;\theta) = \left(\frac{\mu}{1+\theta\mu}\right)^y \frac{(1+\theta y)^{y-1}}{y!} \exp\left(-\frac{\mu(1+\theta y)}{1+\theta\mu}\right), y = 0,1,2, \dots \quad (7)$$

In the GPR model, the mean and variance are given by $E(y) = \mu$ and $(y) = \mu(1 + \theta\mu)^2$. If θ equals zero, the GPR model simplifies to an ordinary Poisson regression model. Conversely, if θ > 0, the GPR model indicates overdispersion, while if θ < 0, it indicates underdispersion. The GPR model has the same form as the Poisson regression model, but it can handle overdispersion and underdispersion cases more effectively.

$$\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \tag{8}$$

2.7 Negative Binomial Regression (NBR)

Negative binomial regression models are effective for modeling discrete data subject to overdispersion, as the negative binomial distribution is an extension of the Poisson-Gamma distribution that includes a dispersion parameter θ. This distribution allows handling additional variability beyond that anticipated by a simple Poisson model, making it more suitable for data with greater variability. In this context, the regression model formed on the Poisson-Gamma mixture distribution is expressed in the form of parameters $\mu = \alpha\beta$ and $\theta = 1$ α so that the mean and variance are obtained in the form as below:

$$E(Y) = \mu \text{ and } V[Y] = \mu + \theta\mu 2 = \mu(1 + \theta\mu)$$

The probability mass function for the negative binomial regression model is given by [23]:

$$f(y, \mu, \theta) = \frac{\Gamma(y + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y + 1)} \left(\frac{1}{1 + \theta\mu}\right)^{\theta^{-1}} \left(\frac{\theta\mu}{1 + \theta\mu}\right)^{y} \tag{9}$$

The negative binomial model integrates the Poisson distribution with the Gamma distribution. Below is an explanation of the negative binomial regression model:

$$\mu_i = \exp\left(\beta_0(u_i, v_i) + \left(\beta_1(u_i, v_i)x_{i1} + \cdots + \beta_p(u_i, v_i)x_{ip}\right)\right) \tag{10}$$

The probability function for the negative binomial distribution is as follows:

$$f(y, \mu, \theta) = \exp\left\{ln\left(\frac{(\Gamma(y + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y + 1)}\right) + \frac{1}{\theta} ln\left(\frac{1}{1} + \theta\mu\right) + y\, ln\left(\frac{\theta\mu}{1 + \theta\mu}\right)\right\} \tag{11}$$

2.8 Spatial Weighting

In spatial analysis, parameter estimates at a particular location point (u, v) are formed by a local weighting matrix (W), which is created through the application of a kernel function.

$$W(u_i, v_i) = \begin{bmatrix} W(u_1, v_1) & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & W(u_i, v_i) \end{bmatrix}$$

The spatial weights used in this study are fixed bisquare kernel and adaptive bisquare kernel [24]:

$$W_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b_i}\right)^2\right]^2 & ; \quad for\ d_{ij} < b \\ 0; & others \end{cases} \tag{12}$$

Description:

b : A non-negative parameter called bandwidth

b_i : Bandwidth at the i-th location

d_{ij} : Euclidean distance between location i and location j.

The value of d_{ij} is obtained from the formula:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{13}$$

d_{ij} represents the Euclidean distance between locations (u_i, v_i) and b_i denotes the optimal bandwidth value at the i-th location. Bandwidth acts as a smoothing parameter. A small bandwidth value can lead to increased variability and an undersmoothed model, while a large bandwidth value may introduce high bias and an oversmoothed model. The optimal bandwidth value is determined using Cross Validation (CV) [10]:

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(\boldsymbol{b})]^2 \tag{14}$$

Where $\hat{y}_{\neq i}(\boldsymbol{b})$ represents the estimated value of y_i within radius b, with the observation at point i excluded from the estimation process.

2.8 Geographically Weighted Generalized Poisson Regression (GWGPR)

The GWGPR model is an extension of the Generalized Poisson Regression (GPR). The probability distribution function of GWGPR for each location is expressed as follows [25]:

$$f(y_i|\mu_i, \theta) = \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \frac{(1+\theta y_i)^{y_i - 1}}{y_i!} exp\left(\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right),\ y_i = 0,1,2, \ldots \tag{15}$$

Where,

y_i : observed value of the i-th response

x_{ik} : observation value of the k-th predictor variable at the observation location (u_i, v_i)

$\beta_j(u_i, v_i)$: regression coefficient of the j-th predictor variable for each location (u_i, v_i)

$\theta(u_i, v_i)$: disperse parameter for each location (u_i, v_i)

μ_i : $e^{x_i^T \beta(u_i, v_i)}, \theta_i = \theta(u_i, v_i)$

The form of the GWGPR equation is as follows:

$$\mu_i = exp\left(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \cdots + \beta_p(u_i, v_i)\right)x_{ip} \tag{16}$$

The parameters of the GWGPR model are estimated using Maximum Likelihood Estimation (MLE). To obtain these parameter estimates, a numerical approach is needed, specifically the Newton-Raphson iteration [25].

2.9 Geographically Weighted Negative Binomial Regression (GWNBR)

Geographically Weighted Negative Binomial Regression (GWNBR) is highly effective for estimating data with spatial heterogeneity, particularly for count data that exhibit overdispersion. The GWNBR model can be expressed as follows [26]:

$$\mu_i = e^{x_i^T \beta(u_i, v_i)} = e^{(\beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \cdots + \beta_p(u_i, v_i)x_{ip})} \tag{17}$$

The negative binomial distribution function for each location can be expressed using the following equation:

$$f(y_i|\beta_j(u_i, v_i)\theta_i) = \frac{\Gamma\left(y_i + \frac{1}{\theta_i}\right)}{\Gamma\left(\frac{1}{\theta_i}\right)\Gamma(y_i + 1)} + \left(\frac{1}{1+\theta_i\mu_i}\right)^{\frac{1}{\theta_i}} + \left(\frac{\theta_i\mu_i}{1+\theta_i\mu_i}\right)^{y_i}, y_i = 0,1,2, \ldots \tag{18}$$

Parameters of the GWNBR model are estimated using the Maximum Likelihood Estimation (MLE) method. To obtain the parameter estimates for the GWNBR model, additional numerical techniques are needed. Specifically, the Newton-Raphson iteration is employed in this process [26].

2.10 Parameter Testing for GPR, NBR, GWGPR and GWNBR Models

Estimation of GWGPR and GWNBR model parameters can use Maximum Likelihood Estimation (MLE).

1. Simultaneous Test

   Simultaneous significance test is conducted using the Maximum Likelihood Ratio Test (MLRT), with the following hypotheses:

   $H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots \beta_p(u_i, v_i) = 0$

   H_1 : there is at least one $\beta_k(u_i, v_i) \neq 0, k = 1,2, \dots, p$

   The test statistics used are as follows:

$$D(\hat{\beta}) = -2\, ln(\Delta) = -2\, ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) \tag{19}$$

   Where,

   $L\omega$: Likelihood function that does not involve predictor variables

   $L\Omega$: Likelihood function involving predictor variables

   Reject H_0 if the value of $D\hat{\beta} > X^2_{(\alpha,k)}$ indicating that at least one parameter in the GWGPR model has a significant impact on the response variable.

2. Partial Test

   Partial testing is conducted to identify which parameters significantly impact the response variable at each location, using the following hypotheses:

   $H_0 :\ \beta_k(u_i, v_i) = 0$

   H_1 : there is at least one $\beta_k \neq 0, k = 1,2, \dots, p$

   The test statistics used are as follows:

$$Z_{count} = \frac{\beta_k(u_i, v_i)}{se(\beta_k(u_i, v_i))} \tag{20}$$

   Reject H_0 if the value of $|Z_{count}| > Z_{\frac{\alpha}{2}}$ indicating that the parameter has a significant influence on the response variable at each location.

Parameter estimation for the Poisson Regression, NB, GPR model is done using the Maximum Likelihood Estimation (MLE) method.

1. Simultaneous Test

   The negative binomial regression model fit test is performed using the deviance test. The testing process is as follows:

   $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_p = 0$

   H_1: there is at least one $\beta_k \neq 0, k = 1,2, \dots, p$.

   Test Statistics:

$$D(\boldsymbol{\hat{\beta}}) = -2ln\Delta = -2\, ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\omega})}\right) \tag{21}$$

   Reject H_0 if the value of $D(\boldsymbol{\beta}) > \boldsymbol{x}(\alpha; p)\mathbf{2}$, meaning that at least one variable has an influence on the model.

2. Partial Test

   This step is important to ensure that each predictor variable that is considered influential does contributes significantly to the response variable, and does not just function as a nuisance variable in the model.

   $H_0: \beta_k = 0$

   $H_1: \beta_k \neq 0, k = 1,2, \dots, p$.

Test Statistics:

$$Z_{count} = \frac{\beta_j}{SE(\beta_j)} \tag{22}$$

Reject H_0 if the value of $|Z_{count}| > Z_{\frac{\alpha}{2}}$, meaning that the variable has a significant effect on the response variable in the negative binomial regression model.

2.11 Best Model Selection

In evaluating the goodness of the model, Fotheringham et al. (2002) suggested using the AIC criterion. The mathematical expression to calculate the AIC value is as follows [27]:

$$AIC = D(h) + 2K(h) \tag{23}$$

Where:

$$D(h) = \sum_{j=1}^{n} \frac{y_j \log \hat{y}_j \left(\boldsymbol{\beta}(u_j, v_j), h\right)}{y_j} + \left(y_j - \left(\hat{y}_j \boldsymbol{\beta}(u_j, v_j), h\right)\right)$$

$K(h)$ = trace $(S(h))$

$D(h)$ = is the deviant of the model with bandwidth (h)

$K(h)$ = number of parameters in the model with bandwidth (h)

3. RESULTS AND DISCUSSION

3.1 Multicollinearity

The goal of assessing multicollinearity is to determine if any predictor variables are highly correlated with one another. To identify multicollinearity, the VIF (Variance Inflation Factor) value can be used as a criterion.

Table 1. VIF Value of Each Predictor Variable

Variable	VIF
X_1	2.785
X_2	1.398
X_3	1.690
X_4	1.986
X_5	1.544
X_6	2.046

Multicollinearity occurs if the VIF value on each predictor variable has a value of more than 10. Based on Table 1 above, all variables are free from multicollinearity or there is no linear relationship between predictor variables in the data on the number of property crime cases in Central Java [28].

3.2 Poisson Regression

Poisson regression is a non-linear regression model applied to count data where the response variable follows a Poisson distribution. Table 2 shows Poisson regression model applied to the property crime data in Central Java.

Table 2. Poisson Regression Model Results

Parameter	Estimate	Std. Error	Z Value	Pr(>|z|)
(Intercept)	1.035	0.2445	4.232	2.31×10^{-5} ***
$\boldsymbol{X_1}$	0.02363	0.005371	4.399	1.09×10^{-5} ***
$\boldsymbol{X_2}$	0.04934	0.008015	6.156	7.45×10^{-10} ***
$\boldsymbol{X_3}$	−0.0170	0.002114	−8.042	2×10^{-16} ***

X_4	7.364	0.4789	15.376	2×10^{-16} ***
X_5	-2.916×10^{-5}	5.818×10^{-6}	-5.011	5.41×10^{-7} ***
X_6	1.489×10^{-6}	4.373×10^{-8}	34.062	2×10^{-16} ***

Signif. codes: 0'***' 0.001'**' 0.01'*' 0.05'.' ''1

Simultaneous and Partial testing is carried out with the following hypotheses:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$

H_1 : There is at least one $\beta_k \neq 0 ; k = 1,2, \dots , n$

The decision to reject H_0 if the deviance value or $D(\hat{\beta}) > X^2_{k;\alpha}$. From the value of chi square table, the deviance value is 1447.6 and the X^2 table value is 37.916. Therefore, the value of $D(\hat{\beta}) > X^2_{k;\alpha}$ (1447.6 > 37.916), so the conclusion is that there is at least one dependent variable that affects the response variable.

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$

H_1 : there is at least one $\beta_k \neq 0 ; k = 1,2, \dots , n$

After getting the calculated Z value, a comparison is made with the Z table value. With a significance level of 10%, the value of $Z_{\frac{\alpha}{2}} = Z_{0.05} = 1.65$. From **Table 2**, it can be concluded that all variables X_1, namely the number of poor people, X_2 open unemployment rate, X_3 school enrollment rate, X_4 gini ratio X_5 population density, and X_6 average net income of free workers partially affect the model because they have a value $|Z_{count}| \geq Z_{\frac{\alpha}{2}}$.

3.3 Overdispersion

Poisson regression has a characteristic where the mean value is equal to the variance value, known as equidispersion. Equidispersion can be assessed by the dispersion coefficient range from 0 to 1.

The dispersion ratio value obtained is:

$$\alpha = \frac{Poisson\ Deviance}{n-k-1} = \frac{1447,6}{35-6-1} = 53,614$$

From the results of the dispersion ratio value, it is concluded that the value is greater than 1, which means that the data on the number of property crime regression models is overdispersed.

3.4 Negative Binomial Regression

Negative binomial regression can also be used to handle overdispersion problems that may occur in Poisson regression. **Table 3** shows a negative binomial regression model that is applied to property crime data in Central Java in 2023.

Table 3. Negative Binomial Regression Model Results

Parameter	Estimate	Std. Error	Z Value	Pr(>|z|)
(Intercept)	2.388	1.633	1.463	0.14356
X_1	0.01772	0.03377	0.525	0.59973
X_2	0.06137	0.05253	1.168	0.24273
X_3	-0.007293	0.01315	-0.555	0.57923
X_4	3.817	3.217	1.187	0.23541
X_5	-2.093×10^{-5}	3.853×10^{-5}	-0.543	0.58704
X_6	1.015×10^{-5}	3.393×10^{-7}	2.990	0.00279 **

Signif. codes: 0'***' 0.001'**' 0.01'*' 0.05'.' ''1

Simultaneous and Partial testing is carried out with the following hypotheses:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$

H_1 : there is at least one $\beta_k \neq 0 ; k = 1,2, \dots , n$

The deviance value of the negative binomial regression model is 36.014. With a 10% significance level, the table value $X^2_{(k;\alpha)} = 37.916$ is obtained, so it is decided to accept H_0, which means that there are one or more predictor variables that have a significant effect on the response variable.

Based on Table 3 on the value of $\left|Z_{hitung}\right|$ which is compared with $Z_{\frac{\alpha}{2}}$ which is $Z_{\frac{0,1}{2}} = 1{,}65$. The table shows that there is a parameter β_6 which has a value of $Z_{\frac{0,1}{2}} = 1{,}65$ so reject H_0 which means that the variable that has a significant effect on the number of property crimes in Central Java is the average net income of independent workers (X_6). The Z value provides a firmer and more direct interpretation of the magnitude of the standardization effect. When we use the Z statistic, we can directly compare its value with the critical value of the standard normal distribution. This is particularly helpful in the context of this study, where with a 10% significance level, the critical value can be directly compared to the calculated Z value, providing clarity in statistical decision-making.

3.5 Generalized Poisson Regression

The Generalized Poisson Regression (GPR) model is used when there is a case of over/under dispersion in the Poisson regression model. Modeling results Poisson regression modeling results obtained the ratio of the deviance value to the degree of freedom is greater than 1. than 1, meaning that the data on the number of property crimes in Central Java in 2023 experiencing overdispersion, so an analysis is carried out using the GPR method.

Table 4. Parameter Estimation of GPR Model

Parameter	Estimate	Z value
β_0	4.815×10^{-10}	-1.089609
β_1	1.421×10^{-9}	-1154890
β_2	1.635×10^{-9}	-2579503878
β_3	2.617×10^{-8}	-1157245
β_4	7.655×10^{-11}	-2.168007
β_5	-1.688×10^{-5}	-0.1300896
β_6	3.446×10^{-7}	7621669508 **
Deviance	974.2087	
AIC	411.5927	

Signif. codes: 0'***' 0.001'**' 0.01'*' 0.05'.' ''1

Simultaneous and Partial testing is carried out with the following hypotheses:

$H_0 : \beta_1 = \beta_2 = \beta_k = 0$

H_1 : There is at least one $\beta_k \neq 0, k = 1,2, \ldots, k$

Based on many previous studies, the following values were used to determine an interpretation of the model. The test results obtained with a significance level of 10% (α = 0.1) obtained the value of $x^2_{(0,1;6)} = 37.916$ which means smaller than the deviance value = 974.2087, so reject H_0 which means there is at least one predictor variable that has a significant effect on the response variable. So, it can be continued with partial testing.

$H_0 : \beta_k = 0$

H_1 : There is at least one $\beta_k \neq 0, k = 1,2, \ldots, p$

Based on Table 4 in the column it can be seen that the value of $|Z_{count}|$ there is a parameter β_6 which has a value of $|Z_{count}| > 1{,}65$ so reject H_0 which means that the variable that has a significant effect on the number of property crime cases in Central Java is the average net income of free workers (X_6).

3.6 Spatial Heterogeneity

Spatial heterogeneity testing is conducted to identify the presence of diversity in spatial data, which is a characteristic of spatial heterogeneity. The test statistic used to test spatial heterogeneity is Breusch-Pagan (BP), with the following hypothesis:

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_{34}^2 = \sigma^2$ (variance between locations is equal)

H_1 : there is at least one $\sigma_i^2 \neq \sigma^2$ (variance between locations is different)

The Breusch-Pagan test statistic was found to be 10.934 and the p-value was 0.09045. An α of 10% is used so that the conclusion is obtained to reject H_0 or the variance between locations is different. This means that there are differences in characteristics between one observation point and another.

3.7 Geographically Weighted Generalized Poisson Regression (GWGPR)

Geographically Weighted Generalized Poisson Regression (GWGPR) represents an advanced form of Generalized Poisson Regression (GPR). To assess significance, the Maximum Likelihood Ratio Test (MLRT) is employed. Both simultaneous and partial tests are conducted based on the following hypotheses:

H_0: $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots \beta_k(u_i, v_i) = 0$

H_1 : there is at least one $\beta_k(u_i, v_i) \neq 0, k = 1,2, \dots, k$

The results of the calculation of deviance obtained test results with a significance level of 10% (α = 0.1) and $x^2_{(0,1:6)} = 37.916$. This means reject H_0, that is, there is at least one predictor variable that has a significant effect on the response variable.

H_0 : $\beta_k(u_i, v_i) = 0$

H_1 : there is at least one $\beta_k \neq 0, k = 1,2, \dots, p$

Based on the test results, the parameters are different for each observation location of Regency/City in Central Java. A comparison was used to determine each variable that affects each district/city in Central Java by looking at the calculated Z value compared to the Z table. The value of $|Z_{hitung}|$ which is compared with $Z_{(0,1:2)}$ is $Z_{(0,1:2)} = 1,65$. Reject H_0 if the value of $|Z_{hitung}| > 1,65$.

The grouping of districts/cities in Central Java based on significant variables is presented in Figure 1.

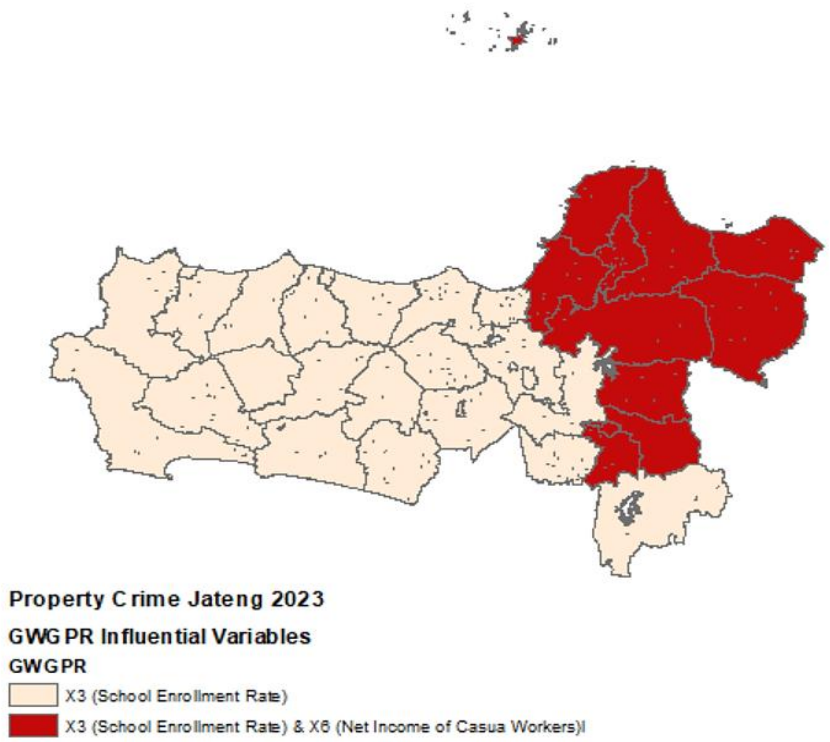

Figure 1. Grouping Based on Significant Variables with GWGPR Model.

From the Figure 1 above, it can be seen that 12 regencies / cities in Central Java, namely Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, and Surakarta City have 2 influential variables each marked in red and X_6 influences all regencies / cities in Central Java Province shown in reddish brown. This shows that there are allegations of spatial aspects in the data on property crime rates in Central Java Province.

3.8 Geographically Weighted Negative Binomial Regression (GWNBR)

Geographically Weighted Negative Binomial Regression (GWNBR) is an extension of negative binomial regression. Simultaneous significance test using Maximum Likelihood Ratio Test (MLRT).

Simultaneous and Partial testing is carried out with the following hypotheses:

H_0 : $\beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \cdots \beta_k(u_i, v_i) = 0$

H_1 : There is at least one $\beta_k(u_i, v_i) \neq 0, k = 1,2, \dots, k$

The results of the test with a significance level of 10% (α = 0.1) and $x^2_{(0,1:6)} = 37.916$ where the Deviance > $x^2_{(0,1:6)} = 37.916$. This means reject H_0 , that is, there is at least one predictor variable that has a significant effect on the response variable. So that it can be continued with partial testing.

H_0 : $\beta_k(u_i, v_i) = 0$

H_1 : There is at least one $\beta_k \neq 0, k = 1,2, \dots, k$

Based on the test results, the parameters are different for each observation location of in Central Java. The value of $Z_{(0,1:2)}$ which is compared with $Z_{(\alpha:2)}$ is $Z_{(0,1:2)} = 1,65$. Reject H_0 if the value of $|Z_{count}| >$ 1.65.

The grouping of districts/cities in Central Java based on significant variables is presented in Figure 2.

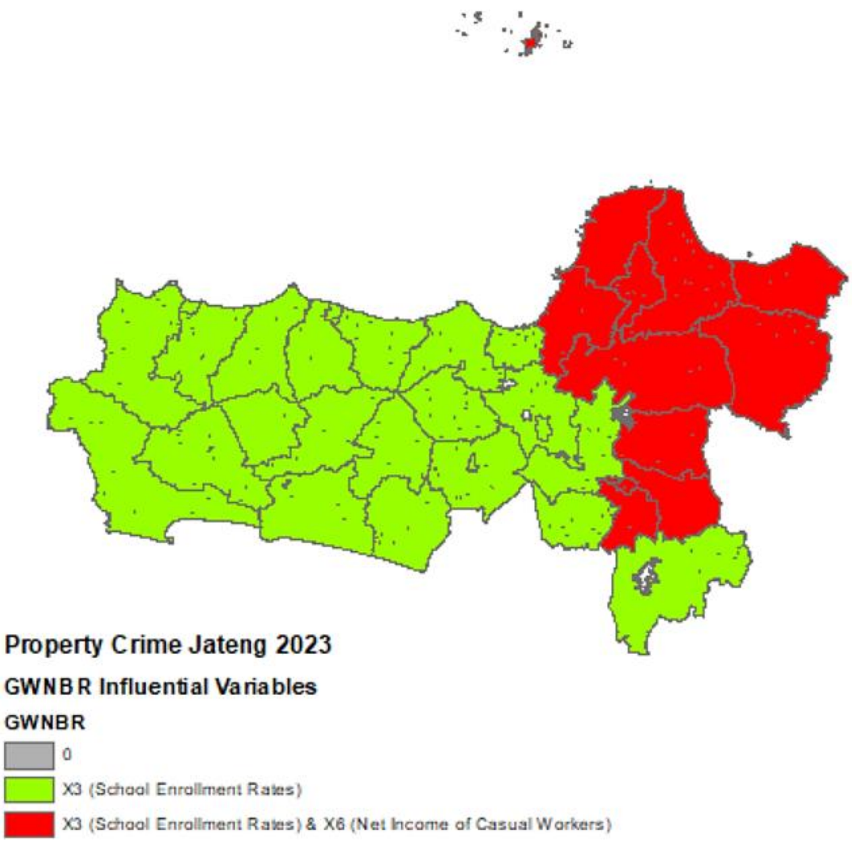

Figure 2. Grouping Based on Significant Variables with the GWNBR Model

Based on Figure 2, it can be seen that 12 regencies / cities in Central Java, namely Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, and Surakarta City have 2 influential variables each marked in red and X_6 influences all regencies / cities in Central Java Province shown in green. This shows that there are allegations of spatial aspects in the data on property crime rates in Central Java Province.

3.9 Best Model Selection

The best model is selected by comparing the AIC values of each model and selecting the model with the smallest AIC value. Where the smallest AIC value indicates that the better the GWGPR and GWNBR models. The AIC results of the GWGPR and GWNBR models as well as the Poisson and Negative Binomial Models.

Table 5. Comparison of AIC Values

Model	AIC
Poisson Regression	1708.1
NBR	424.97
GPR	411.5927
GWGPR	411.5927
GWNBR	**411.3652**

Table 5 shows that the best model in modeling the number of property crime cases in Central Java Province in 2023 with the smallest AIC value of 411.3652 is the GWNBR model with Adaptive Bisquare Kernel weighting.

4. CONCLUSIONS

According to the test results for GWGPR and GWNBR modeling, two distinct groupings were identified. The clustering analysis reveals that two variables significantly impact property crime rates across all districts and cities in Central Java:

1. The percentage of net income of free workers (X_6)
2. The school enrollment rate (X_3),

which affects only 12 districts and cities.

Conversely, the percentage of poverty rate (X_1), open unemployment rate (X_2), Gini ratio (X_4), and population density (X_5) do not significantly influence the number of property crime cases in any district or city in Central Java.

Additionally, the school enrollment rate (X_3) is not significant in 23 districts and cities within the province. The AIC criterion indicates that the GWNBR method is the most appropriate for modeling property crime cases in each district and city of Central Java, with a value of 411.3652, compared to Poisson regression, negative binomial regression, and GWGPR.

REFERENCES

[1] Y. Yigzaw, A. Mekuriaw, dan T. Amsalu, “ANALYZING PHYSICAL AND SOCIO-ECONOMIC FACTORS FOR PROPERTY CRIME INCIDENT IN ADDIS ABABA, ETHIOPIA,” *Heliyon*, vol. 9, no. 2, hal. e13282, 2023, doi: https://doi.org/10.1016/j.heliyon.2023.e13282.

[2] B. J. Ficarra, *INCARCERATED CONGENITAL DIAPHRAGMATIC HERNIA*, vol. 92, no. 3. 1956. doi: https://doi.org/10.1016/S0002-9610(56)80126-8.

[3] Lance Lochner, "Education, Work and Crime: Theory and Evidence," New York, University of Rochester, 1999, pp. 2-3.

[4] S. M. T. Situmaeng, *BUKU AJAR KRIMINOLOGI*. 2021.

[5] B. A. Wisudaningsi, I. Arofah, dan K. A. Belang, “PENGARUH KUALITAS PELAYANAN DAN KUALITAS PRODUK TERHADAP KEPUASAN KONSUMEN DENGAN MENGGUNAKAN METODE ANALISIS REGRESI LINEAR BERGANDA,” *Statmat J. Stat. Dan Mat.*, vol. 1, no. 1, hal. 103–117, 2019, doi: https://doi.org/10.32493/sm.v1i1.2377.

[6] S. W. Tyas, Gunardi, dan L. A. Puspitasari, “GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION MODEL WITH THE BEST KERNEL FUNCTION IN THE CASE OF THE NUMBER OF POSTPARTUM MATERNAL MORTALITY IN EAST JAVA,” *MethodsX*, vol. 10, no. October 2022, hal. 102002, 2023, doi: https://doi.org/10.1016/j.mex.2023.102002.

[7] N. M. R. Keswari, I. W. Sumarjaya, dan N. L. P. Suciptawati, “PERBANDINGAN REGRESI BINOMIAL NEGATIF DAN REGRESI GENERALISASI POISSON DALAM MENGATASI OVERDISPERSI (STUDI KASUS: JUMLAH TENAGA KERJA USAHA PENCETAK GENTENG DI BR. DUKUH, DESA PEJATEN),” *E-Jurnal Mat.*, vol. 3, no. 3, hal. 107, 2014, doi: https://doi.org/10.24843/mtk.2014.v03.i03.p072.

[8] D. Rahmadini, I. N. Manfaati, dan P. A. Rismawati, “PEMODELAN BIVARIATE GENERALIZED POISSON REGRESSION PADA KASUS ANGKA KEMATIAN DI PROVINSI JAWA TENGAH,” *Pros. Semin. Nas. UNIMUS*, vol. 6, hal. 401–410, 2023, [Daring]. Tersedia pada: https://prosiding.unimus.ac.id/index.php/semnas/article/view/1482

[9] W. Pratama dan S. P. Wulandari, “PEMETAAN DAN PEMODELAN JUMLAH KASUS PENYAKIT TUBERCULOSIS (TBC) DI PROVINSI JAWA BARAT DENGAN PENDEKATAN GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION,” *J. Sains dan Seni ITS*, vol. 4, no. 1, hal. 37–42, 2015.

[10] M. Adryanta dan P. Purhadi, “ANALISIS METODE GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION UNTUK PEMODELAN FAKTOR YANG MEMPENGARUHI JUMLAH KEMATIAN ANAK DI PROVINSI JAWA TIMUR,” *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: https://doi.org/10.12962/j23373520.v8i2.43562.

[11] Desriwendi, Abdul Hoyyi, dan Triastuti Wuryandari, “PEMODELAN GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION (GWLR) DENGAN FUNGSI PEMBOBOT FIXED GAUSSIAN KERNEL DAN ADAPTIVE GAUSSIAN KERNEL (STUDI KASUS : LAJU PERTUMBUHAN PENDUDUK PROVINSI JAWA TENGAH),” *J. Gaussian*, vol. 4, no. 2, hal. 193–204, 2015.

[12] D. L. Sogen, P. R. Arum, dan R. Wasono, “PEMODELAN GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) DENGAN PEMBOBOT FIXED BISQUARE KERNEL DAN ADAPTIVE BISQUARE KERNEL PADA KASUS DBD DI RIAU GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) MODELING WITH FIXED BISQUARE KERNEL AND ADAPTIVE BISQUA,” *Semin. Nas. Publ. Hasil-Hasil Penelit. dan Pengabdi. Masy.*, vol. 6, no. 18 Oktober 2023, hal. 371–387, 2023.

[13] E. Supriyadi, S. Mariani, dan Sugiman, “PERBANDINGAN METODE PARTIAL LEAST SQUARE (PLS) DAN,” *Unnes J. Math.*, vol. 6, no. 2, hal. 117–128, 2017.

[14] P. R. Arum dan S. Alfian, “PEMODELAN PERTUMBUHAN EKONOMI DI JAWA BARAT MENGGUNAKAN METODE GEOGRAPHICALLY WEIGHTED PANEL REGRESSION,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, hal. 219–227, 2022, doi: https://doi.org/10.36456/jstat.vol15.no2.a5506.

[15] K. A. Yulianingsih, K. G. Gde, dan L. P. Suciptawati, “PENERAPAN REGRESI POISSON UNTUK MENGETAHUI FAKTOR-FAKTOR YANG MEMENGARUHI JUMLAH SISWA SMA/SMK YANG TIDAK LULUS UN DI BALI,” *e-Jurnal Mat.*, vol. 1, no. Agustus, hal. 59–63, 2012.

[16] W. Sabtika, A. Prahutama, dan H. Yasin, “PEMODELAN GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION (GWGPR) PADA KASUS KEMATIAN IBU NIFAS DI JAWA TENGAH,” *J. Gaussian*, vol. 10, no. 2, hal. 259–268, 2021, doi: https://doi.org/10.14710/j.gauss.v10i2.30946.

[17] A. Rahayu, “MODEL-MODEL REGRESI UNTUK MENGATASI MASALAH OVERDIPERSI PADA REGRESI POISSON,” *J. Peqguruang Conf. Ser.*, vol. 2, no. 1, hal. 1, 2021, doi: https://doi.org/10.35329/jp.v2i1.1866.

[18] K. Antonio dan J. Beirlant, “ACTUARIAL STATISTICS WITH GENERALIZED LINEAR MIXED MODELS,” *Insur. Math. Econ.*, vol. 40, no. 1, hal. 58–76, 2007, doi: https://doi.org/10.1016/j.insmatheco.2006.02.013.

[19] M. Beisemann, “A FLEXIBLE APPROACH TO MODELLING OVER-, UNDER- AND EQUIDISPERSED COUNT DATA IN IRT: THE TWO-PARAMETER CONWAY–MAXWELL–POISSON MODEL,” *Br. J. Math. Stat. Psychol.*, vol. 75, no. 3, hal. 411–443, 2022, doi: https://doi.org/10.1111/bmsp.12273.

[20] A. G. Halunga, C. D. Orme, dan T. Yamagata, *A HETEROSKEDASTICITY ROBUST BREUSCH–PAGAN TEST FOR CONTEMPORANEOUS CORRELATION IN DYNAMIC PANEL DATA MODELS*, vol. 198, no. 2. 2017. doi: https://doi.org/10.1016/j.jeconom.2016.12.005.

[21] P. Rismawati Arum, R. Putra Gautama, I. Fitriani, dan F. Naza Nurvahyani, “IDENTIFYING FACTORS THAT INFLUENCE LIFE EXPECTANCY IN CENTRAL JAVA USING SPATIAL REGRESSION MODELS,” *J Stat.*, vol. 16, no. 2, hal. 605–613, 2023.

[22] M. Al Haris dan P. R. Arum, “NEGATIVE BINOMIAL REGRESSION AND GENERALIZED POISSON REGRESSION MODELS ON THE NUMBER OF TRAFFIC ACCIDENTS IN CENTRAL JAVA,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 2, hal. 471–482, 2022, doi: https://doi.org/10.30598/barekengvol16iss2pp471-482.

[23] C. Raschke dan W. H. Greene, “ERRATUM TO FUNCTIONAL FORMS FOR THE NEGATIVE BINOMIAL MODEL FOR COUNT DATA [ECONOMICS LETTERS 99, (2008), 585-590],” *Econ. Lett.*, vol. 107, no. 2, hal. 313, 2010, doi: https://doi.org/10.1016/j.econlet.2009.09.026.

[24] Y. Wibawati, “MAXIMUM LIKELIHOOD ESTIMATION MODEL LINEAR DAN LOG-LINEAR DALAM REGRESI POISSON,” no. 3, hal. 181–188, 2009.

[25] E. U. L. Fitri, “PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH KASUS TUBERKULOSIS DI JAWA TIMUR MENGGUNAKAN METODE GEOGRAPHICALLY WEIGHTED GENERALIZED POISSON REGRESSION DAN GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION,” 2017.

[26] B. W. Y. Priambodo dan I. Irhamah, “PEMETAAN JUMLAH PROPERTY CRIME DI PROVINSI JAWA TIMUR MENGGUNAKAN METODE GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION (GWNBR) DAN GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR),” *Inferensi*, vol. 2, no. 2, hal. 53, 2019, doi: https://doi.org/10.12962/j27213862.v2i2.6818.

[27] T. Nakaya, A. S. Fotheringham, C. Brunsdon, dan M. Charlton, “GEOGRAPHICALLY WEIGHTED POISSON REGRESSION FOR DISEASE ASSOCIATION MAPPING,” *Stat. Med.*, vol. 24, no. 17, hal. 2695–2717, 2005, doi: https://doi.org/10.1002/sim.2129.

[28] R. P. Gautama, “PENGARUH PENGELUARAN PERKAPITA, JUMLAH PENDUDUK MISKIN, TINGKAT PENGANGGURAN TERBUKA (TPT), DAN TINGKAT PARTISIPASI ANGKATAN KERJA (TPAK) TERHADAP

PENDAPATAN PAJAK PROVINSI JAWA TENGAH," *Emerg. Stat. Data Sci. J.*, vol. 2, no. 1, hal. 97–106, 2024, doi: https://doi.org/10.20885/esds.vol2.iss.1.art10.