Jurnal Varian
Vol. 8, No. 1, October 2024, pp 97–112
e-ISSN: 2581-2017

Modeling the Number of High School Dropouts in Indonesia Using GWGPR

Nur Azizah, Junaidi, Nurul Fiskia Gamayanti, Hartayuni Sain, Fadjriyani
Universitas Tadulako, Palu, Indonesia

Article Info

Article history:

Received : 07-14-2024
Revised : 09-02-2024
Accepted : 10-18-2024

Keywords:

Dropout;
GWGPR;
Overdispersion;
Senior High School.

ABSTRACT

In 2022, the high school dropout rate is the highest compared to other levels of education in Indonesia. Seeing the urgency of the 12-year Compulsory Education program, completing education up to the high school level is an important thing that needs to be considered. Thus, it is necessary to know the factors that influence the dropout rate in the hope that this problem can be reduced. This study aims to model the high school dropout rate using geographically weighted generalized poisson regression (GWGPR) based on the factors that influence it. GWGPR is used if the response variable is overdispersed and depends on the location observed. The results of this study indicate that each province has a different regression model. The GWGPR model with the adaptive tricube kernel weighting function is the best model because it has the smallest AIC value compared to other weighting functions. In Central Sulawesi Province, the GWGPR model with the adaptive tricube kernel weighting function formed is $\hat{\mu}_{26} = \exp(8,1267 - 0,1267X_4 + 0,0344X_5 + 0,0957X_6 + 0,1173X_7)$. With the significant variables are the average length of schooling, the percentage of the population aged 7-17 years who receive PIP, the open unemployment rate, and the percentage of children who do not live with parents.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i1.4248

Corresponding Author:

Nur Azizah,
Department of Statistics, Universitas Tadulako, Palu, Indonesia,
Email: nurazizah.qi28@gmail.com

How to Cite:
Azizah, N., Junaidi, J., Gamayanti, N. F., Sain, H., & Fadjriyani, F. (2024). Modeling the Number of High School Dropouts in Indonesia Using GWGPR. *Jurnal Varian*, *8*(1), 97-112.

A. INTRODUCTION

Until now, the development of the education sector has been a challenge in Indonesia. The government consistently strives to improve the country's participation and quality of education. One of the steps that has been taken is the increase in the duration of compulsory education from 9 years to 12 years. This action is part of the government's efforts to achieve the education development goals of increasing the average years of schooling and reducing the dropout rate (Limbong & Setiadi, 2021). In 2022, the high school dropout rate is the highest compared to other education levels, which is 1,38% or as many as 10.091 high school students dropped out of school in the 2022/2023 academic year (BPS, 2022). Considering the urgency of the 12 Years Compulsory Education program, completing studies up to the senior high school level is an important matter that needs to be considered. Therefore, it is necessary to know the factors that affect the dropout rate in the hope that the problem can be reduced. Regression is one of the statistical methods used to model the relationship between predictor variables and response variables. Regression aims to find a functional relationship between these variables and predict the value of the response variable based on the value of the given predictor variable (Septianto et al., 2023).

The number of school dropouts in Indonesia is a critical indicator that requires serious attention. Statistical analysis of this data often employs Poisson regression due to its nature as count data, which is assumed to follow a Poisson distribution. Poisson regression, assuming equidispersion (where the mean and variance are equal), has become the standard approach. However, recent studies have shown that this assumption is often violated in real-world contexts.

Data frequently exhibit overdispersion (where the variance exceeds the mean) or underdispersion (where the variance is less than the mean), indicating that the traditional Poisson regression model may not always be appropriate. This inadequacy can lead to biased parameter estimates and errors in statistical inference.

The knowledge gap arises when traditional Poisson regression models are applied universally without considering the presence of overdispersion or underdispersion, leading to less accurate analytical results. Generalized Poisson Regression (GPR) has been developed to bridge this gap as an alternative that can handle variance that deviates from the equidispersion assumption. GPR allows for more flexible and accurate modeling, accommodating both overdispersion and underdispersion, thereby providing more reliable outcomes in analyzing school dropout data. Generalized Poisson Regression modeling produces a global regression model for all observation locations (Tyas et al., 2023).

However, further challenges emerge in spatial data analysis, where variability across locations can affect the model's results. In this context, Geographically Weighted Generalized Poisson Regression (GWGPR) offers a solution by integrating spatial elements into the GPR model. GWGPR extends GPR by incorporating a weighting function based on Euclidean distance between observation locations, allowing for more specific analysis tailored to the geographical characteristics of the studied regions (Mahmuda and Harini, 2014).

Based on research conducted by Hakim (2020) on the factors causing children to drop out of school using logistic regression, it was found that the education of the head of the household, ownership of KIP/PIP, number of household members, working children, poverty, and area of residence had a significant effect on children dropping out of school. Although it has the same case study, the difference between this research and previous studies is that the method used is logistic regression on binary data in Aceh Province, which does not consider the problem of overdispersion or spatial variability. In contrast, this study will use the Geographically Weighted Generalized Poisson Regression (GWGPR) method to model the factors that influence high school dropouts throughout Indonesia so that it can handle the problem of overdispersion and accommodate spatial effects.

In addition, other previous studies conducted by Sabtika et al. (2021) showed the effectiveness of the GWGPR model in cases of postpartum mortality. As a result, this model has the lowest Akaike's Information Criterion (AIC) value compared to other models. This shows that GWGPR outperforms global models such as Poisson regression. Another study conducted by Safire & Purhadi (2023) examined the factors that influence the number of diabetes mellitus cases in East Java using GWGPR and GWNBR (Geographically Weighted Negative Binomial Regression). The results showed that the GWGPR model had a smaller corrected Akaike's Information Criterion (AICc) value than the GWNBR model. This shows that the GWGPR method is more appropriate. However, these two previous studies did not explore the effects of different weighting functions in improving model accuracy. In fact, choosing the right weighting function can significantly affect model results. Therefore, the gap between this research and the two previous research is that this study will evaluate three adaptive weighting functions (adaptive Gaussian, adaptive bisquare, and adaptive tricube). By exploring various weighting functions, this study aims to determine the weighting function that provides the best results in modeling the factors causing school dropout.

Adatunaung et al. (2023) focused on the performance of various kernel functions in the Geographically Weighted Regression (GWR) model to determine the factors affecting the Human Development Index in South Sulawesi Province. They found that the adaptive tricube kernel weighting function outperformed the adaptive Gaussian and bisquare functions because it had the highest R^2 and the lowest AIC values. This research emphasized the importance of selecting the appropriate weighting function, but these findings have not yet been directly applied to the GWGPR model.

Based on the description above, this study aims to fill these gaps by not only applying GWGPR to the analysis of school dropout data in Indonesia but also by carefully evaluating the performance of different weighting functions within this model. In this study, the parameter estimation process will be conducted using three different adaptive kernel weighting functions: Gaussian, bisquare, and tricube, to determine which weighting function is the best. Therefore, this research is expected to provide a new perspective on how different weighting functions can enhance the effectiveness of GWGPR in capturing spatial variability and addressing the issues of overdispersion and underdispersion in count data. Additionally, the results of this study are also expected to provide insights into the factors that are suspected to influence the number of school dropouts in Indonesia, which can be used as a basis for government consideration in formulating educational policies in Indonesia.

B. RESEARCH METHOD

The data used in this study are secondary data obtained from KEMENDIKBUD, BPS, and KEMENPPPA. The variable used is the number of high school dropouts in Indonesia in 2022 as the response variable. There are 7 predictor variables used, namely the percentage of poor people (X_1), The percentage of the population aged 10-17 years who work (X_2), The percentage of the population aged 0-17 years who are neglected (X_3), the average length of schooling (X_4), the percentage of the population aged 7-17 years who get PIP (X_5), the open unemployment rate (X_6), and the percentage of children who do not live with parents (X_7).

Data analysis in this study used RStudio and SAS Studio software. The following are the stages of analysis carried out in this study:

1. Input data used as research variables.
2. Conducting multicollinearity test based on VIF value.
   The symptom of multicollinearity can be seen from the Variance Inflation Factor (VIF) value. If the VIF value < 10, it is stated that there is no multicollinearity (Leonita et al., 2023). The VIF value can be formulated in the following equation:

$$VIF_j = \frac{1}{1 - R_p^2}; p = 1, 2, \ldots, k \tag{1}$$

   which R_p^2 is the coefficient of determination between the pth predictor variable and other predictor variables.
3. Conducting the Poisson distribution test.
   The Kolmogorov-Smirnov test can be used to assess whether the observed data follows a Poisson distribution. The test hypothesis is as follows (Tyas et al., 2023):
   $H_0 : F(y) = F_0(y)$ (The data follows a Poisson distribution).
   $H_1 : F(y) \neq F_0(y)$ (The data not follows a Poisson distribution).

   Statistic test:

$$D = max\left|\left(F(y) - F_0(y)\right)\right| \tag{2}$$

   which:
   $F(y)$: Cumulative distribution function.
   $F_0(y)$: Empirical value of the sample cumulative distribution.

   If $D_{hit} > D_{(1-\alpha,\ n)}$ or $p - value < \alpha$ then H_0 rejected, or in other words, the data does not follow the Poisson distribution.
4. Test the assumption of equidispersion.
   Equidispersion can be detected by determining the value of dispersion (θ), which is the deviance value divided by the degree of freedom (db). Data is equidispersion if the dispersion value (θ) is equal to 1, if the dispersion value (θ) is less than 1 it means underdispersion, and if the dispersion value (θ) is more than 1 it means overdispersion (Winata, 2023).

$$\theta = \frac{Deviance}{db} \tag{3}$$

   which $db = n - p, p$ is the number of parameters and n is the number of observations. The deviance value can be obtained using the following equation:

$$Deviance = 2\sum_{i=1}^{n}\left(y_i\, ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)\right) \tag{4}$$

   which:
   y_i : Actual value of the i-th observed response variable.
   $\hat{y}_i$: The estimated value of the i-th observed response variable.
   n : Number of observations.
   i : $1,\ 2, \ldots,\ n.$

DOI: https://doi.org/10.30812/varian.v8i1.4248

5. Modeling Generalized Poisson Regression.
The GPR model is similar to the Poisson regression model, which is as follows (Sari et al., 2023):

$$\hat{\mu}_i = exp\left(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}\right) \tag{5}$$

which:
μ_i : The average number of events that occur in a given time interval.
β_0 : Regression model intercept value.
$\beta_1,\ \beta_2, \ldots,\ \beta_k$: Regression coefficient of predictor variable 1 to k.
$X_{i1},\ X_{i2}, \ldots, X_{ik}\&$: The value of the 1st to k-th predictor variable of the i-th observation.

6. Conducting spatial heterogeneity test using Breusch-Pagan test.
Spatial heterogeneity can be tested using the Breusch-Pagan (BP) test which has the following hypothesis (Lumbantoruan et al., 2023):
H_0 : $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_n^2 = \sigma^2$ (There is no spatial heterogeneity).
H_1 : There is at least one $\sigma_i^2 \neq \sigma^2; i = 1, 2, \ldots,\ n$ (There is spatial heterogeneity).

Statistic test:

$$BP = \left(\frac{1}{2}\right) f^T Z \left(Z^T Z\right)^{-1} Z^T f\ \sim\ \chi^2_{p-1} \tag{6}$$

which:
f : A $1 \times n$ vector with $f_i = \frac{e_i^2}{\hat{\sigma}^2} - 1$.
e_i : Residual at the i-th observation.
i : $1, 2, \ldots, n.$
σ^2 : Variance of the residuals.
Z : Vector of normalized response variables for each observation.

The decision to reject H_0 is obtained if the value of the Breusch-Pagan test statistic is greater than the critical value $\chi^2{}_{(\alpha,\ p-1)}$ or the p-value is less than the significant level α.

7. Determining the optimum bandwidth using cross validation.
The selection of the optimum bandwidth is used to adjust the variance and bias of the parameter estimates. If the bandwidth is large, it will cause a large bias because the model is too smooth. If the bandwidth is too small, it will cause a small bias because the model is too coarse. The method used to determine the optimum bandwidth is Cross Validation (CV).

$$CV\left(h\right) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}\left(h_i\right)\right)^2 \tag{7}$$

which:
n : Number of observations.
i : $1,\ 2,\ \ldots,\ n.$
y_i : The i-th observation.
$\hat{y}_{\neq i}\left(h_i\right)$: The estimated value of y_i conditional on the observation of location $\left(u_i,\ v_i\right)$

8. Calculating the Euclidean distance.
The distance between location id and location j is obtained from the euclidean distance which can be calculated using the equation (Hong et al., 2021).

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{8}$$

which:
u_i : Latitude of the i-th location.
v_i : Longitude of the i-th location.

9. Determining the weight matrix.
Spatial weights calculated using adaptive kernel functions produce different bandwidth values for each observation location. The adaptive kernel spatial weight function is categorized into adaptive Gaussian, adaptive squared, and adaptive tricube kernel functions (Al-Hasani et al., 2021).
   a. Adaptive Gaussian Kernel

$$W_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right] \tag{9}$$

   b. Adaptive Bisquare Kernel

$$W_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h_i}\right)^2\right)^2 & , d_{ij} \leq h_i \\ & 0,\ d_{ij} > h_i \end{cases} \tag{10}$$

   c. Adaptive Tricube Kernel

$$W_{ij} = \begin{cases} \left(1-\left(\frac{d_{ij}}{h_i}\right)^3\right) & , d_{ij} \leq h_i \\ & 0,\ d_{ij} > h_i \end{cases} \tag{11}$$

   which:
   d_{ij} : Euclidean distance between the i-th location and the j-th location.
   h_i : Bandwidth of the i-th location.

10. Implementing Geographically Weighted Generalized Poisson Regression modeling which includes:
   a. Estimating the parameters of the GWGPR model.
   In the GWGPR model, the method used to estimate the model parameters is the Maximum Likelihood Estimation (MLE) method, with the GWGPR model likelihood function as follows (Tyas et al., 2023):

$$L\left(\ \beta\left(u_i, v_i\right)\right) = \prod_{i=1}^{n} f\left(y_i\right) = \prod_{i=1}^{n}\left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i}\frac{\left(1+\theta y_i\right)^{y_i-1}}{y_i!} exp\left(-\frac{\mu_i\left(1+\theta y_i\right)}{1+\theta\mu_i}\right) \tag{12}$$

   The likelihood function in Equation 12 is then converted into natural logarithm form (ln-likelihood) with the geographical weights shown in Equation 13.

$$\begin{aligned} &\ln L^*\left(\beta\left(u_i, v_i\right)\right) \\ &\quad= \sum_{i=1}^{n} w_{ij}\Bigg[y_i\left(x_i^T\beta(u_i,v_i) - \ln(1+\theta e^{x_i^T\beta(u_i,v_i)})\right) \\ &\quad+\left(y_i-1\right)\ln\left(1+\theta y_1\right) - \ln y_i! - \frac{e^{x_i^T\beta(u_i,v_i)(1+\theta y_i)}}{1+\theta e^{x_i^T\beta(u_i,v_i)}}\Bigg] \end{aligned} \tag{13}$$

   The process of obtaining parameter estimators of the GWGPR model is by deriving Equation 13 for each parameter and then equating it to zero. However, the results cannot be done analytically because the results obtained are not close form, so it is necessary to do iteration, namely Newton-Raphson iteration with the function as follows.

$$\beta_{(m+1)}\left(u_i v_i\right) = \beta_m\left(u_i v_i\right) - H^{-1}{}_{(m)}\left(\beta_m\left(u_i v_i\right)\right) g_{(m)}\left(\beta_m\left(u_i v_i\right)\right) \tag{14}$$

   b. Testing the GWGPR model parameters both simultaneously and partially.
   Simultaneous testing is done using the Likelihood Ratio Test (LRT) method with the following hypothesis (Purhadi et al., 2021):
   H_0 : $\beta_1\left(u_i,\ v_i\right) = \beta_2\left(u_i,\ v_i\right) = \ldots = \beta_k\left(u_i,\ v_i\right) = 0$ (All predictdo not affect no effect on the response variable).
   H_1 : There is at least one $\beta_p\left(u_i,\ v_i\right) \neq 0$ (There is at least one predictor variable that affects the response variable).
   with $i = 1,\ 2,\ \ldots,\ n$ and $p = 1,\ 2,\ \ldots,\ k$

Statistic test:

$$\begin{aligned} G &= -2ln\Lambda = -2\,ln\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right) \\ &= 2\left(\ln L\left(\hat{\Omega}\right) - \ln L\left(\hat{\omega}\right)\right) \end{aligned} \tag{15}$$

which:

$L\left(\hat{\omega}\right)$: Likelihood function without involving predictor variables.

$L\left(\hat{\Omega}\right)$: Likelihood function involving predictor variables.

Rejection criteria:

Reject H_0 if $G > \chi^2_{(\alpha,\ n-p-1)}$ or the p-value is less than the significant level α, which means that at least one predictor variable is significant to the response variable.

Partial test is used to find out which predictor variables have a significant effect on the response variabel with the following hypothesis (Widyaningsih & Budiawan, 2023):

H_0 : $\beta_p\left(u_i,\ v_i\right) = 0$ (The p-th variable has no significant effect).

H_1 : There is at least one $\beta_p\left(u_i,\ v_i\right) \neq 0$ (The p-th variable has significant effect).

with $i = 1,\ 2,\ \ldots,\ nandp = 1,\ 2,\ \ldots,\ k$

Statistic test:

$$W = \left(\frac{\hat{\beta}_p\left(u_i,\ v_i\right)}{se\left(\hat{\beta}_p\left(u_i,\ v_i\right)\right)}\right)^2 \tag{16}$$

which:

$\hat{\beta}_p\left(u_i,\ v_i\right)$: Estimated value for parameter β_p in i-th location.

$se\left(\hat{\beta}_p\left(u_i,\ v_i\right)\right)$: Estimated standard error β_p in i-th locatin.

Rejection criteria:

Reject H_0 if the test statistic W is greater than the critical value $W > \chi^2_{(\alpha,\ db=1)}$ or the p-value is less than the significant level α.

c. Selecting the best model by looking at the smallest AIC value.
The calculation of the Aikake Information Criterion (AIC) value is as follows (Mahama et al., 2020).

$$AIC = -2lnL\left(\hat{\beta}\right) + 2p \tag{17}$$

which:

$L\left(\hat{\beta}\right)$: The maximum value of the likelihood function of the model parameters.

p : Number of parameters in the model.

The data analysis steps above are presented as a flowchart as shown in Figure 1.

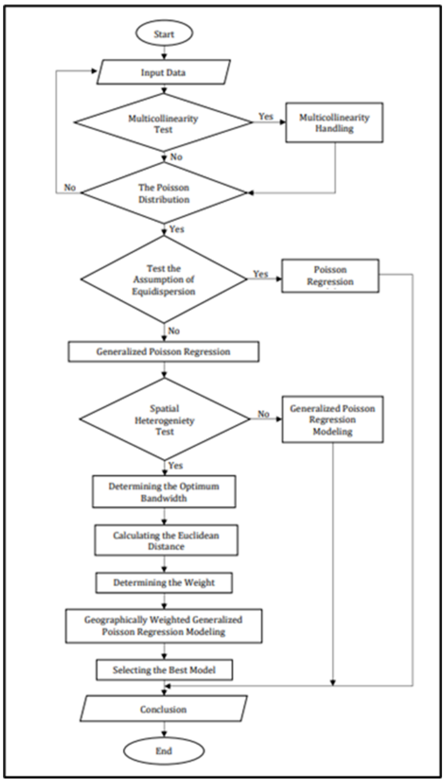

Figure 1. Research Data Analysis Flow Chart

C. RESULT AND DISCUSSION

1. Multicollinearity Test

Multicollinearity testing is used to identify whether the regression model's predictor variables correlate. Table 1 presents the VIF value for each predictor variable.

Table 1. Multicollinearity Test

Predictor Variables	VIF Value
X_1	3,659
X_2	2,795
X_3	2,778
X_4	1,688
X_5	1,620
X_6	2,191
X_7	1,976

Based on Table 1, it can be seen that all predictor variables, X_1 to X_7 have VIF values less than 10, so it can be concluded that there is no multicollinearity or no correlation between predictor variables and other predictor variables.

2. Poisson Distribution Test

Data distribution testing to determine whether the data on the number of high school dropouts is Poisson distributed or not. The method used in this test is the Kolmogorov-Smirnov test with the following hypothesis:

H_0 : Data on the number of high school dropouts is Poisson distributed.
H_1 : Data on the number of high school dropouts is not Poisson distributed.

Table 2. Poisson Distribution Test

Kolmogorov sminorv Test	p-value
0,147	0,458

The Kolmogorov smirnov test results in Table 2 show that the p-value is 0,458 greater than α (0,05), so it is decided that H_0 fails to be rejected. So it can be concluded that the data on the number of high school dropouts follows the Poisson distribution.

3. Equidispersion Assumption Test

One way to determine whether or not overdispersion occurs is to use the dispersion value, which can be seen in the following table.

Table 3. Equidispersion Test

Deviance	df	Dispersi
8758,1	26	336,85

Based on Table 3, the deviance value is 8758.1 and if divided by the degree of freedom, the dispersion value will be 336.85, where the value is more than 1, then the Poisson regression model of the number of high school dropouts in Indonesia experiences a case of overdispersion. Because the model experiences a case of overdispersion, then one of the alternative methods that can be used is Generalized Poisson Regression.

4. GPR Modeling

The Generalized Poisson Regression model is a statistical model implemented to analyze the relationship between variables that experience overdispersion. The parameters estimation value β is obtained as follows:

Table 4. GPR Parameter Estimation Value

Parameters	Estimate
β_0	7,748
β_1	0,236
β_2	-0,130
β_3	0,613
β_4	-0,974
β_5	-0,116
β_6	0,903
β_7	0,731

Based on Table 4, it can be seen that the Generalized Poisson Regression model formed is as follows:

$$\mu = \exp\left(7,748 + 0,236X_1 - 0,130X_2 + 0,613X_3 - 0,974X_4 - 0,116X_5 + 0,903X_6 + 0,731X_7\right)$$

5. Spatial Heterogeneity Test

Spatial heterogeneity testing is conducted to determine whether the data to be spatial modeling contains spatial heterogeneity. Geographically Weighted Regression analysis will be appropriate if there is diversity between provinces. The effect of spatial heterogeneity can be known by using the Bruesch-Pagan test statistic with the following hypothesis:

H_0 : There is no spatial heterogeneity.

H_1 : There is spatial heterogeneity.

Table 5. Bruesch-Pagan Test

Breusch-Pagan Test	p-value
17,571	0,014

DOI: https://doi.org/10.30812/varian.v8i1.4248

Based on Table 5, the p-value is 0,014 smaller than α (0,05) which means H_0 is rejected, it can be concluded that there is spatial heterogeneity or differences in variation between provinces, which allows the GWGPR model to be applied in this study.

6. Determination of Optimum Bandwidth

The first step before performing GWGPR modeling is determining each location's weight matrix. The weighting functions used are adaptive gaussian kernel, adaptive bisquare kernel, and adaptive tricube kernel where each weighting requires an optimum bandwidth value. Determination of the optimum bandwidth (h) value using cross validation (CV) criteria. The adaptive kernel weighting produces a bandwidth value that differs for each location. The optimum bandwidth value using CV can be found in Table 6.

Table 6. Optimum Bandwidth

Location	Adaptive Gaussian Kernel	Adaptive Bisquare Kernel	Adaptive Tricube Kernel
Aceh	25,597	42,541	42,543
North Sumatra	23,170	39,986	39,987
West Sumatra	21,591	38,222	38,223
⋮	⋮	⋮	⋮
West Papua	28,249	36,281	36,281
Papua	34,318	42,550	42,550

7. Euclidean Distance Calculation

After getting the optimum bandwidth value, the next step is to determine the weight matrix W at each i-th location by first calculating the Euclidean distance (d_{ij}) to all locations. Suppose a sample of the Euclidean distance between two points, namely Aceh and North Sumatra Provinces, is given as follows, and the calculation results are presented in Table 7.

$$
\begin{aligned}
d_{(1,\,2)} &= \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \\
&= \sqrt{(4,41533 - 2,36304)^2 + (96,9956 - 99,2161)^2} \\
&= \sqrt{(2,05229)^2 + (-2,2205)^2} \\
&= \sqrt{4,211894244 + 4,93062025} \\
&= \sqrt{4,211894244 + 4,93062025} \\
&= 3,0237
\end{aligned}
$$

Table 7. Euclidean distance between provinces

Location	Aceh	North Sumatra	West Sumatra	...	West Papua	Papua
Aceh	0	3,0237	6,2172	...	36,2864	42,5557
North Sumatra	3,0237	0	3,3136	...	33,8107	39,9998
West Sumatra	6,2172	3,3136	0	...	32,2212	38,2354
⋮	⋮	⋮	⋮	⋮	⋮	⋮
West Papua	36,2864	33,8107	32,2212	...	0	6,4909
Papua	42,5557	39,9998	38,2354	...	6,4909	0

8. Weighting Matrix

a. Adaptive Gaussian Kernel Weights Matrix

The following is an example of calculating the weights between Aceh and North Sumatra Provinces with the adaptive Gaussian kernel function. The calculation results are presented in Table 8.

$$
\begin{aligned}
W_{(1,\,2)} &= \exp\left[-\tfrac{1}{2}\left(\tfrac{d_{ij}}{h_i}\right)^2\right] \\
&= \exp\left[-\tfrac{1}{2}\left(\tfrac{3{,}0237}{25{,}59729}\right)^2\right] \\
&= 0,993
\end{aligned}
$$

Table 8. Adaptive Gaussian Kernel Function Weight Matrix

Location	Aceh	North Sumatra	West Sumatra	...	West Papua	Papua
Aceh	1	0,993	0,971	...	0,366	0,251
North Sumatra	0,991	1	0,990	...	0,345	0,225
West Sumatra	0,959	0,988	1	...	0,328	0,208
⋮	⋮	⋮	⋮	⋮	⋮	⋮
West Papua	0,438	0,489	0,522	...	1	0,974
Papua	0,463	0,507	0,538	...	0,982	1

b. Adaptive Bisquare Kernel Weights Matrix

The following is an example of calculating the weights between Aceh and North Sumatra Provinces with the adaptive square kernel function. The calculation results are presented in Table 9.

$$
\begin{aligned}
W_{(1,\,2)} &= \left(1-\left(\tfrac{d_{ij}}{h_i}\right)^2\right)^2 \\
&= \left(1-\left(\tfrac{3{,}0237}{42{,}54138}\right)^2\right)^2 \\
&= 0,990
\end{aligned}
$$

Table 9. Adaptive Bisquare Kernel Function Weight Matrix

Location	Aceh	North Sumatra	West Sumatra	...	West Papua	Papua
Aceh	1	0,990	0,958	...	0,074	0
North Sumatra	0,989	1	0,986	...	0,081	0
West Sumatra	0,948	0,985	1	...	0,084	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
West Papua	0	0,017	0,045	...	1	0,937
Papua	0	0,013	0,037	...	0,954	1

c. Adaptive Tricube Kernel Weights Matrix

The following is an example of calculating the weights between Aceh and North Sumatra Provinces with the adaptive tricube kernel function. The calculation results are presented in Table 10.

$$
\begin{aligned}
W_{(1,\,2)} &= \left(1-\left(\tfrac{d_{ij}}{h_i}\right)^3\right)^3 \\
&= \left(1-\left(\tfrac{3{,}0237}{42{,}54271}\right)^3\right)^3 \\
&= 0,999
\end{aligned}
$$

Table 10. Adaptive Tricube Kernel Function Weight Matrix

Location	Aceh	North Sumatra	West Sumatra	...	West Papua	Papua
Aceh	1	0,999	0,991	...	0,055	0
North Sumatra	0,999	1	0,998	...	0,062	0
West Sumatra	0,987	0,998	1	...	0,064	0

Location	Aceh	North Sumatra	West Sumatra	...	West Papua	Papua
⋮	⋮	⋮	⋮	⋮	⋮	⋮
West Papua	0	0,007	0,027	...	1	0,983
Papua	0	0,005	0,021	...	0,989	1

9. GWGPR Model

a. Simultaneous Testing

The parameters that have been estimated are then tested to see whether the parameters have a significant effect simultaneously on the variable number of high school dropouts or not. The simultaneous test is done with the likelihood ratio test G. The hypothesis used is as follows:

H_0 : All predictor variables do not affect the response variable.

H_1 : There is at least one predictor variable that affects the response variable.

Table 11. Likelihood Ratio Test

Weighting Function	$\ln L(\hat{\Omega})$	$\ln L(\hat{\omega})$	G
Adaptive Gaussian Kernel	-243,666	-1253,061	2018,79
Adaptive Bisquare Kernel	-224,079	-1240,6	1993,041
Adaptive Tricube Kernel	-224,108	-1239,85	1991,483

Based on Table 11, the value of G with adaptive gaussian kernel weighting function is 2018.79, the value of G with adaptive bisquare kernel weighting function is 1993.041, the value of G with adaptive tricube kernel weighting function is 1991.483, and the value of $\chi^2_{(0,05;34-7-1)}$ is 38.885 or the value of G from the three weighting functions is greater than $\chi^2_{(0,05;34-7-1)}$ then H_0 is rejected, so it can be concluded that the predictor variables simultaneously significantly affect the variable number of dropouts, whether using the adaptive Gaussian kernel, adaptive square kernel, or adaptive tricube kernel weighting function.

b. Partial Testing

Partial testing is used to determine the variables that significantly affect the number of high school dropouts with the GWGPR model in each observation location. The test statistic used is the Wald test. H_0 is rejected or the variable affects the number of high school dropouts if $W > \chi^2_{0,05;\ 1}(3,841)$. The following are the variables that have a significant effect in each province.

1. Partial Testing of GWGPR Adaptive Gaussian Kernel

Table 12. Significant Variables in Each Province with Adaptive Gaussian Kernel

Affected Variables	Provinces
X_6	Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Riau Islands
X_6, X_7	Lampung, Bangka Belitung Islands, Special Capital Region of Jakarta, West Java, Central Java, Special Region of Yogyakarta, East Java, Banten, Bali, West Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan
X_5, X_6, X_7	East Nusa Tenggara, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Papua

We will present partial parameter testing at the 26th research location, Central Sulawesi Province, as an example.

Table 13. Parameter Estimation of GWGPR Adaptive Gaussian Kernel of Central Sulawesi Province

Parameter	Estimate	Wald
β_0	7,7111	170,4929*
β_1	-0,0126	0,7058
β_2	0,0005	0,0011
β_3	0,0217	1,2108

β_4	-0,0771	1,9890
β_5	0,0252	4,7723*
β_6	0,1065	9,4599*
β_7	0,0766	7,0075*

Based on Table 13, it can be seen that there are four parameters, namely β_0, β_5, β_6 and β_7, which have a test statistic value of $W > \chi^2_{0,05;\ 1}(3,841)$ then H_0 is rejected, so it can be concluded that at the 5% significance level, the parameters β_0, β_5, β_6 and β_7 have a significant effect on the model. Parameters that do not significantly affect the model, namely β_1, β_2, β_3 and β_4 need to be removed from the model so that the final GWGPR model with adaptive gaussian kernel weighting function in Central Sulawesi Province is formed:

$$\hat{\mu}_{26} = \exp\left(7,7111 + 0,0252X_5 + 0,1065X_6 + 0,0766X_7\right)$$

2. Partial Testing of GWGPR Adaptive Bisquare Kernel

Table 14. Significant Variables in Each Province with Adaptive Bisquare Kernel

Affected Variables	Provinces
X_6	Aceh, North Sumatra, West Sumatra, Riau
X_6, X_7	Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Daerah Khusus Ibukota Jakarta, Jawa Barat, Jawa Tengah, Daerah Istimewa Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara
X_3, X_5, X_7	West Papua, Papua
X_5, X_6, X_7	East Nusa Tenggara, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi
X_3, X_5, X_6, X_7	North Sulawesi, Maluku, North Maluku

We will present partial parameter testing at the 26th research location, Central Sulawesi Province, as an example.

Table 15. Parameter Estimation of GWGPR Adaptive Bisquare Kernel of Central Sulawesi Province

Parameter	Estimate	Wald
β_0	8,0970	122,243*
β_1	-0,0261	1,975
β_2	-0,0205	0,817
β_3	0,0433	2,311
β_4	-0,1250	3,533
β_5	0,0341	6,126*
β_6	0,0980	4,848*
β_7	0,1169	10,892*

Based on Table 15, it can be seen that there are four parameters, namely β_0, β_5, $\beta_6 and \beta_7$, which have a test statistic value of $W > \chi^2_{0,05;\ 1}(3,841)$ then H_0 is rejected, so it can be concluded that at the 5% significance level the parameters β_0, β_5, β_6 and β_7 have a significant effect on the model. Parameters that do not have a significant effect on the model, namely β_1, β_2, β_3 and β_4, need to be removed from the model so that the final GWGPR model with adaptive bisquare kernel weighting function in Central Sulawesi Province is formed:

$$\hat{\mu}_{26} = \exp\left(8,0970 + 0,0341X_5 + 0,0980X_6 + 0,1169X_7\right)$$

3. Partial Testing of GWGPR Adaptive Tricube Kernel

Table 16. Significant Variables in Each Province with Adaptive Tricube Kernel

Affected Variable	Provinces
X_6, X_7	Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Riau Islands, Special Capital Region of Jakarta, West Java, Central Java, Special Region of Yogyakarta, East Java, Banten, Bali, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan
X_4, X_6, X_7	West Nusa Tenggara
X_4, X_5, X_6, X_7	East Nusa Tenggara, Central Sulawesi, South Sulawesi, Southeast Sulawesi, West Sulawesi
X_3, X_5, X_7	West Papua, Papua
X_3, X_5, X_6, X_7	North Sulawesi, Gorontalo, Maluku, North Maluku

We will present partial parameter testing at the 26th research location, Central Sulawesi Province, as an example.

Table 17. Parameter Estimation of GWGPR Adaptive Tricube Kernel of Central Sulawesi Province

Parameter	Estimate	Wald
β_0	8,1267	135,918*
β_1	-0,0259	2,096
β_2	-0,0226	1,029
β_3	0,0450	2,538
β_4	-0,1267	4,060*
β_5	0,0344	6,893*
β_6	0,0957	4,996*
β_7	0,1173	11,606*

Based on Table 17, it can be seen that there are five parameters, namely β_0, β_4, β_5, β_6 and β_7, which have a test statistic value of $W > \chi^2_{0,05;\ 1}(3,841)$ then H_0 is rejected, so it can be concluded that at the 5% significance level the parameters β_0, β_4, β_5, β_6 and β_7 Parameters that do not have a significant effect on the model, namely β_1, β_2, and β_3, need to be removed from the model so that the final GWGPR model with adaptive bisquare kernel weighting function in Central Sulawesi Province is formed:

$$\hat{\mu}_{26} = \exp\left(8,1267 - 0,1267X_4 + 0,0344X_5 + 0,0957X_6 + 0,1173X_7\right)$$

10. Best Model Selection

Table 18. Parameter Estimation of GWGPR Adaptive Tricube Kernel of Central Sulawesi Province

Model	AIC
GPR	1297,348
GWGPR Adaptive Gaussian Kernel	566,605
GWGPR Adaptive Bisquare Kernel	565,253
GWGPR Adaptive Tricube Kernel	565,200

Based on Table 18 shows that the model with the smallest AIC value is the GWGPR model with the adaptive tricube kernel weighting function, so it can be concluded that the GWGPR model using the adaptive tricube kernel weighting function is better used to model the data on the number of high school dropouts in Indonesia in 2022 compared to the adaptive gaussian kernel and adaptive bisquare kernel weighting functions. In addition, from the table above, it can also be seen that the AIC value of the GWGPR model is smaller than the AIC value of the GPR model. This result is consistent with the results of a study conducted by Sabtika et al. (2021), which also showed that the local model, GWGPR, is superior in capturing spatial heterogeneity compared to the global model or GPR. In addition, research by Nisa et al. (2022) and Adatunaung et al. (2023) also showed that the adaptive tricube kernel weighting function, performs better in regions with non-uniform spatial distribution by providing a more flexible weighting scheme that adapts to the local data density.

11. Interpretation of the Best GWGPR Model Results

Based on the results obtained, it is known that the GWGPR model with adaptive tricube kernel weighting function is better at analyzing the number of high school dropouts in Indonesia. The GWGPR adaptive tricube kernel model formed in Central Sulawesi Province is as follows:

$$\hat{\mu}_{26} = \exp\left(8,1267 - 0,1267X_4 + 0,0344X_5 + 0,0957X_6 + 0,1173X_7\right)$$

With the significant variables being the average years of schooling (X_4), the percentage of the population aged 7-17 who received PIP (X_5), the open unemployment rate (X_6) nd the percentage of children who do not live with their parents (X_7). Meanwhile, based on the model, it can be explained that:

1. The estimated value for parameter β_0 is 8,1267, which means that the number of high school dropouts will remain as $\exp(8,1267) = 3383,6152 \approx 3384$ students without being influenced by other variables.
2. The estimated value for parameter β_4 is 0,1267 which means that for every additional 1 year of average years of schooling, it will be inversely proportional to the number of high school dropouts by $\exp(0,1267) = 1,1351 \approx 1$ student.
3. The estimated value for parameter β_5 is 0,0344, which means that for every additional 1 percent of the population aged 7-17 who received PIP, it will be proportional to the increase in the number of high school dropouts by $\exp(0,0344) = 1,0350 \approx 1$ student.
4. The estimated value for parameter β_6 is 0,0957, which means that for every 1 percent increase in the open unemployment rate, the number of high school dropouts will increase by $\exp(0,0957) = 1,1004 \approx 1$ student.
5. The estimated value for parameter β_7 is 0,1173, which means that for every additional 1 percent of children who do not live with their parents, the number of high school dropouts will increase by $\exp(0,1173\ ) = 1,1245 \approx 1$ student.

D. CONCLUSION AND SUGGESTION

The best GWGPR model based on the smallest AIC value is the Adaptive Tricube Kernel GWGPR model. The resulting model varies from one province to another, so it will produce 34 models. One of them is the GWGPR model with adaptive tricube kernel weighting function in Central Sulawesi Province:

$$\hat{\mu}_{26} = \exp\left(8,1267 - 0,1267X_4 + 0,0344X_5 + 0,0957X_6 + 0,1173X_7\right)$$

With significant variables being the average years of schooling (X_4), the percentage of the population aged 7-17 who received PIP (X_5), the open unemployment rate (X_6), and the percentage of children who do not live with their parents (X_7).

In future research, it is recommended that other distance metrics, such as Manhattan, be used and that research samples be used at smaller levels, such as districts or cities and villages so that they can sharpen spatial analysis in an area.

REFERENCES

Adatunaung, A. F., Hatidja, D., & Weku, W. C. D. (2023). Performa Kernel pada Model Geographically Weighted Regression untuk Menentukan Faktor-faktor Yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Sulawesi Selatan. *Jurnal Ilmiah Sains*, *23*(2), 140–148. https://doi.org/10.35799/jis.v23i2.48867

Al-Hasani, G., Asaduzzaman, M., & Soliman, A.-H. (2021). Geographically weighted Poisson regression models with different kernels: Application to road traffic accident data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *7*(2), 166–181. https://doi.org/10.1080/23737484.2020.1869628

BPS. (2022). *Statistik Pendidikan 2022*. Badan Pusat Statistika.

Hakim, A. (2020). Faktor Penyebab Anak Putus Sekolah. *Jurnal Pendidikan*, *21*(2), 122–132. https://doi.org/10.33830/jp.v21i2.907.2020

Hong, Z., Mei, C., Wang, H., & Du, W. (2021). Spatiotemporal effects of climate factors on childhood hand, foot, and mouth disease: A case study using mixed geographically and temporally weighted regression models. *International Journal of Geographical Information Science*, *35*(8), 1611–1633. https://doi.org/10.1080/13658816.2021.1882681

Leonita, L., Jeby, N. H., Sevien, & Ohyver, M. (2023). Predicting Cancer Death Rate and Determining the Major Cause of Cancer using Ridge Regression. *Procedia Computer Science*, *227*, 599–605. https://doi.org/10.1016/j.procs.2023.10.563

Limbong, F. S., & Setiadi, Y. (2021). Faktor-Faktor yang Mempengaruhi Putus Sekolah Jenjang SMA/Sederajat di Provinsi DKI Jakarta Tahun 2019. *Prosiding Seminar Pendidikan Matematika dan Matematika*, *3*, 1–8. https://doi.org/10.21831/pspmm.v3i0.144

Lumbantoruan, J. F. A. K., Ohyver, M., & Moniaga, J. V. (2023). Developing a Poverty Model in Papua using Geographically Weighted Regression. *Procedia Computer Science*, *227*, 243–252. https://doi.org/10.1016/j.procs.2023.10.522

Mahama, A., Awuni, J. A., Mabe, F. N., & Azumah, S. B. (2020). Modelling adoption intensity of improved soybean production technologies in Ghana - a Generalized Poisson approach. *Heliyon*, *6*(3), e03543. https://doi.org/10.1016/j.heliyon.2020.e03543

Nisa, R. F., Sugito, S., & Hakim, A. R. (2022). Pemodelan Mixed Geographically Weighted Regression dengan Adaptive Bandwidth untuk Angka Harapan Hidup (Studi Kasus : Angka Harapan Hidup di Jawa Tengah). *Jurnal Gaussian*, *11*(1), 67–76. https://doi.org/10.14710/j.gauss.v11i1.33998

Purhadi, Sari, D. N., Aini, Q., & Irhamah. (2021). Geographically weighted bivariate zero inflated generalized Poisson regression model and its application. *Heliyon*, *7*(7), e07491. https://doi.org/10.1016/j.heliyon.2021.e07491

Sabtika, W., Prahutama, A., & Yasin, H. (2021). Pemodelan Geographically Weighted Generalized Poisson Regression (GWPR) Pada Kasus Kematian Ibu Nifas di Jawa Tengah. *Jurnal Gaussian*, *10*(2), 259–268. https://doi.org/10.14710/j.gauss.10.2.259-268

Safire, E. D., & Purhadi. (2023). Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Diabetes Melitus di Jawa Timur Menggunakan GWGPR dan GWNBR. *Inferensi*, *6*(1), 1–14.

Sari, R. N., Nabilla, U., & Sari, R. P. (2023). Poisson Regression Modeling Generalized in Maternal Mortality Cases in Aceh Tamiang Regency. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *17*(1), 0235–0244. https://doi.org/10.30598/barekengvol17iss1pp0235-0244

Septianto, T., Pramadhana, D., & Haromainy, M. M. A. (2023, August 26). *Matematika Komputasi untuk Vokasi*. Thalibul Ilmi Publishing & Education.

Tyas, S. W., Gunardi, & Puspitasari, L. A. (2023). 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*, *10*, 102002. https://doi.org/10.1016/j.mex.2023.102002

Widyaningsih, Y., & Budiawan, Z. A. (2023). Poisson Regression Models to Analyze Factors That Influence the Number of Tuberculosis Cases in Java. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *17*(3), 1761–1772. https://doi.org/10.30598/barekengvol17iss3pp1761-1772

Winata, H. M. (2023). Mengatasi Overdispersi Dengan Regresi Binomial Negatif pada Angka Kematian Ibu di Kota Bandung. *Jurnal Gaussian*, *11*(4), 616–622. https://doi.org/10.14710/j.gauss.11.4.616-622

[This page intentionally left blank.]