JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 10 (ISSUE 1) JUNE 2026: 68-78
https://doi.org/10.21009/JSA.10106

ANALYSIS OF FACTORS EXPLAINING SENIOR HIGH SCHOOL DROPOUT RATE USING GEOGRAPHICALLY WEIGHTED REGRESSION

Yekti Widyaningsih[1*], Hana Adzania Nufaisah[2]

[1,2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia Kampus UI, Depok, 16424, Indonesia

Corresponding author's e-mail: * **yekti@sci.ui.ac.id**

Article History:
Received: September 04, 2025
Revised: February 28, 2026
Accepted: June 26, 2026
Published: June 30, 2026
Available online.

Keywords:
Dropout Rate;
Fixed Effect Model;
GWPR;
Spatial Heterogeneity.

A BSTRACT

East Nusa Tenggara (NTT) Province is one of the areas experiencing the problem of dropping out of high school education. Even though NTT Province has adequate educational facilities and teaching staff, the high school dropout rate in NTT Province always ranks top 9 in Indonesia for the 2019/2020 academic year to 2021/2022. Dropping out of school can be influenced by region (spatial) and does not occur at one time, so research is needed using panel-structured spatial data that accommodates spatial effects over time. Geographically Weighted Panel Regression (GWPR) is a local regression analysis method that considers the effect of spatial heterogeneity on panel-structured spatial data. This study aims to analyse the factors that explain the high school dropout rate in NTT Province in 2019-2021 using GWPR. The results showed that the GWPR model with the Fixed Exponential weighting function was the best model compared to other weighting functions based on R^2 and AIC. Population density, student-teacher ratio, regional minimum wage, open unemployment rate, student-to-school ratio, average length of schooling, and Smart Indonesia Program budget have a significant effect on explaining high school dropout rates in at least 21 regencies/cities in NTT Province. Grouping districts/cities based on variable significance using k-modes clustering produces 4 groups.

How to cite this article:
Widyaningsih, Y., Nufaisah, H. A., "ANALYSIS OF FACTORS EXPLAINING SENIOR HIGH SCHOOL DROPOUT RATE USING GEOGRAPHICALLY WEIGHTED REGRESSION", Journal Statistika dan Aplikasinya, vol. 10, iss. 1, pp. 68 - 78, June 2026

Journal homepage: *https://journal.unj.ac.id/unj/index.php/statistika*
Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Dropping out of school is a problem that occurs in various regions of Indonesia at all levels of education, including high school education. The province of East Nusa Tenggara (NTT) is one of the areas experiencing the problem of dropping out of high school education. Even though the NTT Province has adequate educational facilities and teaching staff, the high school dropout rate in the NTT Province always ranks in the top 9 in Indonesia in the 2019/2020 academic year to 2021/2022. In addition, in the 2019/2020 and 2021/2022 academic years, the high school dropout rate in NTT Province was higher than the high school dropout rate in Indonesia. If not addressed immediately, the problem of dropping out of school can have an impact on social, economic, poverty, and labor factors. The problem of dropping out of school can be influenced by the geographical conditions of each region [1]. In addition, it is known that the problem of dropping out of school does not only occur at one time. This shows that research is needed that accommodates regional (spatial) influences, over time, to analyze what factors explain the problem of dropping out of school in NTT Province. Therefore, this study uses spatial data with a panel structure, where the data contains location information in the form of coordinates for each regency/city in NTT Province during a certain period.

The method for explaining the effect of the independent variable on the dependent variable can use regression analysis. The use of spatial data in this study allows the appearance of spatial influences in the form of spatial heterogeneity. If its presence is ignored, it can give inaccurate estimates to the global regression model. Therefore, the problem of spatial heterogeneity in this study can be overcome by using Geographically Weighted Panel Regression (GWPR), which is a local panel regression model. GWPR can produce various parameter estimates at each observation point location [2]. Thus, this study aims to analyze the factors that explain the high school dropout rate in NTT Province in 2019-2021 using the GWPR.

2. METHODS

Material and Data

Dropout Concept

A condition where someone leaves school before the end of the study period can be referred to as dropping out of school [3]. The dropout rate of high school education shows the dropout rate for high school education with the following formula [4].

$$APS = \frac{JSPS}{JS} \times 100\% \tag{1}$$

where $JSPS$ is the number of high school dropouts and JS is the number of high school students in the previous year. Lower dropout rates reflect better educational conditions. On the other hand, the higher dropout rate reflects the poor and uneven educational conditions [5].

Research Data

The research data is secondary data consisting of observations in 22 regencies/cities in the Province of NTT from 2019 to 2021 and sourced from the official website of the Central Statistics Agency and the Ministry of Education, Culture, Research, and Technology in Indonesia. The variables used include one dependent variable in the form of high school dropout rate (APS) and seven independent variables that are thought to explain it, namely population density (KP), student-teacher ratio (RMG), regional minimum wage (UMR), open unemployment rate (TPT), student-to-school ratio (RMS), average length of schooling (RRLS), and Smart Indonesia Program budget (APIP).

Research Method

Panel data regression analysis is a regression analysis based on panel data, which is useful for identifying the relationship between the independent variables and the dependent variable. In general, the panel data regression model equation can be written as follows [6].

$$y_{it} = \alpha + \sum_{k=1}^{p} \beta_k x_{itk} + u_{it} \quad (2)$$

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}, i = 1, 2, \dots, N ; t = 1, 2, \dots, T \quad (3)$$

where y_{it} is the dependent variable at the individual i and time t , α is the intercept coefficient, β_k is the slope vector to k, x_{itk} is the vector of the independent variable k for individual i and time t , p is the number of independent variables, u_{it} is the error component of individual i and time t , μ_i is the specific effect of the individual i, λ_t is the specific effect of time t, ε_{it} is the error at the individual i and time t , N is the number of individual units, and T is the number of time units.

Approaches for estimating panel data regression models are divided into the Common Effect Model (CEM), Random Effect Model (REM), and Fixed Effect Model (FEM). CEM is a model with the assumption that there are no differences in variation between individual units and units of time. REM is a model that assumes the effects of individuals or time are random variables [6]. The next approach is FEM, which is used in this study. The consideration of choosing a panel data regression model between FEM or REM is whether the panel data has a large number of individual units N and a small number of time units T that is short panels with individual units in the data are not the result of random sampling, then the right model is FEM [7].

Fixed Effect Model

FEM assumes that the slope coefficient is constant, and the specific effect of the individual unit or time unit is a parameter that can be estimated from an intercept. The following are some forms of FEM [6].

a. Fixed Effect Model with Individual-Specific Effects

The FEM equation with individual-specific effects can be written as follows.

$$y_{it} = \alpha + \mu_i + \sum_{k=1}^{p} \beta_k x_{itk} + \varepsilon_{it} \quad (4)$$

To estimate the parameters in the model, it can be done with the within transformation and followed by parameter estimation using OLS. The equation after the within transformation is as follows.

$$\tilde{y}_{it} = \sum_{k=1}^{p} \beta_k \tilde{x}_{itk} + \tilde{\varepsilon}_{it} \quad (5)$$

$$\tilde{y}_{it} = y_{it} - \underline{y}_{i.} ; \tilde{x}_{itk} = x_{itk} - \underline{x}_{i.k} ; \tilde{\varepsilon}_{it} = \varepsilon_{it} - \underline{\varepsilon}_{i.} \quad (6)$$

b. Fixed Effect Model with Time-Specific Effects

The FEM equation with time-specific effects can be written as follows.

$$y_{it} = \alpha + \sum_{k=1}^{p} \beta_k x_{itk} + \lambda_t + \varepsilon_{it} \quad (7)$$

Estimating the parameters in the model can be done with the within transformation and followed by parameter estimation using OLS. The equation after the within transformation is as follows.

$$\breve{y}_{it} = \sum_{k=1}^{p} \beta_k \breve{x}_{itk} + \breve{\varepsilon}_{it} \quad (8)$$

$$\check{y}_{it} = y_{it} - \underline{y}_{.t} \; ; \check{x}_{itk} = x_{itk} - \underline{x}_{tk} \; ; \check{\varepsilon}_{it} = \varepsilon_{it} - \underline{\varepsilon}_{t} \tag{9}$$

c. Fixed Effect Model with Individual and Time-Specific Effects

The FEM equation with individual and time-specific effects can be written as follows [6].

$$y_{it} = \alpha + \sum_{k=1}^{p} \beta_k \, x_{itk} + \mu_i + \lambda_t + \varepsilon_{it} \tag{10}$$

To estimate the parameters in the model, it can be done with the within transformation and then proceed with parameter estimation using OLS. The model equation after the within transformation is as follows.

$$y_{it}^{*} = \sum_{k=1}^{p} \beta_k \, x_{itk}^{*} + \varepsilon_{it}^{*} \tag{11}$$

$$y_{it}^{*} = y_{it} - \underline{y}_{i.} - \underline{y}_{.t} + \underline{y}_{..} \; ; x_{itk}^{*} = x_{itk} - \underline{x}_{i.k} - \underline{x}_{tk} + \underline{x}_{.k} \; ; \varepsilon_{it}^{*} = \varepsilon_{it} - \underline{\varepsilon}_{i.} - \underline{\varepsilon}_{t} + \underline{\varepsilon}_{..} \tag{12}$$

There are specific effect tests as follows.

a. Individual-Specific Effect Test

H_0: There is no significant individual-specific effect

H_1: There is a significant individual-specific effect

The test statistic used is $F = \frac{(RSS_{null} - RSS_{individual})/(N-1)}{(RSS_{individual})/(NT-N-p)}$ where RSS_{null} is the residual sum of squares under the null hypothesis. $RSS_{individual}$ is the residual sum of squares of the FEM estimates with individual-specific effects. The decision rule is if the $p - value < \alpha$, then H_0 is rejected.

b. Time-Specific Effect Test

H_0: There is no significant time-specific effect

H_1: There is a significant time-specific effect

The test statistic used is $F = \frac{(RSS_{null} - RSS_{time})/(T-1)}{(RSS_{time})/(NT-T-p)}$ where RSS_{time} is the residual sum of squares of FEM estimates with time-specific effects. The decision rule is if the $p - value < \alpha$, then H_0 is rejected. [8]

Then, for the panel data regression model assumption testing, this can be done by testing non-multicollinearity. Multicollinearity means that there is a linear relationship between the independent variables in the regression model, which can be detected by the Variance Inflation Factor (VIF) value. If the VIF value > 10, there is multicollinearity [7]. The normality test aims to test whether the residuals of the model are correctly normally distributed or not. To test the assumption of normality can use the Lilliefors Test [9]. The non-autocorrelation assumption test investigates whether a model displays a correlation between residuals. The presence of autocorrelation in the residuals can be detected by using the Durbin-Watson test [7]. The homoscedasticity test aims to test whether the assumption of the same variance of each residual has been fulfilled [10]. The test is carried out by forming a scatter plot between the estimated value of the dependent variable and the residual. If the scatter plot has points that spread randomly, then the homoscedasticity assumption is fulfilled. Meanwhile, if it is the other way around, then there is an indication of heteroscedasticity [11].

Spatial Heterogeneity

Spatial heterogeneity is a spatial aspect that shows characteristic differences between one location and another, which can be detected by the Breusch-Pagan test. The Breusch-Pagan test hypothesis is as follows [12].

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_N^2 = \sigma^2$ (No spatial heterogeneity)

H_1: Minimal terdapat $\sigma_i^2 \neq \sigma^2, i = 1,2, \ldots N$ (Spatial heterogeneity occurs)

The test statistic is $BP = \left(\frac{1}{2}\right) f'Z(Z'Z)^{-1}Z'f$ where f is a vector $(f_1, f_2, \ldots, f_N)'$ which has elements $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$, Z is a matrix of size $N \times (p+1)$ which contains vectors that have been standardized for each observation, e_i is the residual on observation i, σ^2 is the residual variance, N is the number of observations, p is the number of independent variables. If $BP > \chi^2_{\alpha,p}$ or $p - value < \alpha$ then H_0 is rejected, indicating the occurrence of spatial heterogeneity.

Geographically Weighted Panel Regression (GWPR) Models

The Geographically Weighted Panel Regression (GWPR) model is a model that combines the panel data regression model with Geographically Weighted Regression (GWR) [2]. The GWPR model used in this study is the GWPR fixed effect model, which is a combination model between the fixed effect panel data regression model and the GWR model. The within transformation is used in model building by subtracting the fixed effect model from a model that has been averaged over time. The following is the equation for the fixed effect model [13].

$$y_{it} = \beta_0(u_i, v_i) + \mu_i + \sum_{j=1}^{p} \beta_j(u_i, v_i) x_{itj} + \varepsilon_{it} \tag{13}$$

where y_{it} is the value of the dependent variable for location i and time t, x_{itj} is the value of the independent variable j for location i and time t, μ_i is the specific unobserved effect of location i, $\beta_0(u_i, v_i)$ is the intercept coefficient at location i, $\beta_j(u_i, v_i)$ is the regression coefficient of the independent variable j for location i and ε_{it} is the residual at location i and time t. Then, if it is continued with the within transformation, the following model is obtained [13].

$$\dddot{y}_{it} = \sum_{j=1}^{p} \beta_j(u_i, v_i) \dddot{x}_{itj} + \dddot{\varepsilon}_{it} \tag{14}$$

$$\dddot{y}_{it} = y_{it} - \underline{y}_{i.}\,;\ \dddot{x}_{itj} = x_{itj} - \underline{x}_{i.j}\,;\ \dddot{\varepsilon}_{it} = \varepsilon_{it} - \underline{\varepsilon}_{i.} \tag{15}$$

where $\dddot{y}_{it}$ is the value of the average corrected dependent variable for the location i and time t, $\dddot{x}_{itj}$ is the value of the average corrected independent variable j for location i and time t, (u_i, v_i) is the coordinates for the location i, $\beta_j(u_i, v_i)$ is the regression coefficient of the independent variable, average corrected j for location i, p is the number of independent variables, and $\dddot{\varepsilon}_{it}$ is the corrected average error for location i and time t.

The weighting elements used to estimate the parameters of the GWPR model can be in the form of a spatial weighting matrix. A spatial weighting matrix for location i is $W(u_i, v_i) = diag(w_{11}(u_i, v_i), \cdots, w_{N1}(u_i, v_i), \cdots, w_{1T}(u_i, v_i), \cdots, w_{NT}(u_i, v_i))$ where $w_{NT}(u_i, v_i)$ is the weighted value at the N-th observation location and T-th time. The weight value will be obtained by calculating using the kernel weighting function. The following is the definition of the kernel weighting function [14].

a. *Boxcar:* $w_{ij}(Boxcar) = \{1, for\ d_{ij} < b\ 0,\ else$ (16)

b. *Bisquare:* $w_{ij}(Bisquare) = \{\left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, for\ d_{ij} < b\ 0,\ else$ (17)

c. *Gaussian:* $w_{ij}(Gaussian) = exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$ (18)

d. *Tricube:* $w_{ij}(Tricube) = \{\left(1 - \left(\frac{d_{ij}}{b}\right)^3\right)^3, for\ d_{ij} < b\ 0,\ else$ (19)

e. *Exponential:* $w_{ij}(Exponential) = exp\left(-\frac{d_{ij}}{b}\right)$ (20)

where w_{ij} is the weight value for the location i relative to location j, b is the optimum bandwidth which used in the weighting function. The size of the bandwidth can describe the distance limit for the closest observation, which has a big effect on model formation at the i −th observation location [15]. The bandwidth used in this research is fixed and adaptive bandwidth. Optimum bandwidth can be obtained through the criteria for the smallest CV score where $CV = \sum_{i=1}^{N}(\underline{y_i} - \underline{\hat{y}_{\neq i}}(b))^2$ where $\underline{y_i}$ is the average value of the dependent variable over time at location i and $\underline{\hat{y}_{\neq i}}(b)$ is the estimated value of $\underline{y_i}$ without including the location i [2].

To obtain the estimated value of the GWPR model parameters, it is carried out by minimizing the sum of squared errors and equating it to zero. The parameter estimation result of the GWPR model is $\hat{\beta}(u_i, v_i) = \left[X^{\cdots\prime}W(u_i, v_i)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_i, v_i)Y^{\cdots}$ with $\hat{\beta}'(u_i, v_i) = (\hat{\beta}_1(u_i, v_i), \hat{\beta}_2(u_i, v_i), \dots, \hat{\beta}_p(u_i, v_i))$ [2]. After obtaining the GWPR model parameter estimates, it can be continued by determining the estimated value of $Y^{\cdots}$. If $\widehat{Y^{\cdots}} = (\widehat{y^{\cdots}}_{11}, \widehat{y^{\cdots}}_{21}, \dots, \widehat{y^{\cdots}}_{N1}, \widehat{y^{\cdots}}_{1T}, \widehat{y^{\cdots}}_{2T}, \dots, \widehat{y^{\cdots}}_{NT})'$, then the equation $\widehat{Y^{\cdots}}$ can be written as a result of the projection $Y^{\cdots}$ as follows $\widehat{Y^{\cdots}} = LY^{\cdots}$ with [2]:

$$
\begin{aligned}
&L \\
&= \left[x^{\cdots\prime}_{11}\left[X^{\cdots\prime}W(u_1, v_1)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_1, v_1)\; x^{\cdots\prime}_{21}\left[X^{\cdots\prime}W(u_2, v_2)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_2, v_2)\right. \\
&\vdots\; x^{\cdots\prime}_{N1}\left[X^{\cdots\prime}W(u_N, v_N)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_N, v_N) \\
&\vdots\; x^{\cdots\prime}_{1T}\left[X^{\cdots\prime}W(u_1, v_1)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_1, v_1)\; x^{\cdots\prime}_{2T}\left[X^{\cdots\prime}W(u_2, v_2)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_2, v_2) \\
&\left.\vdots\; x^{\cdots\prime}_{NT}\left[X^{\cdots\prime}W(u_N, v_N)X^{\cdots}\right]^{-1}X^{\cdots\prime}W(u_N, v_N)\right]
\end{aligned}
\tag{21}
$$

The significance test is used to identify parameters that significantly affect the dependent variable in the model. The hypothesis used is [16].

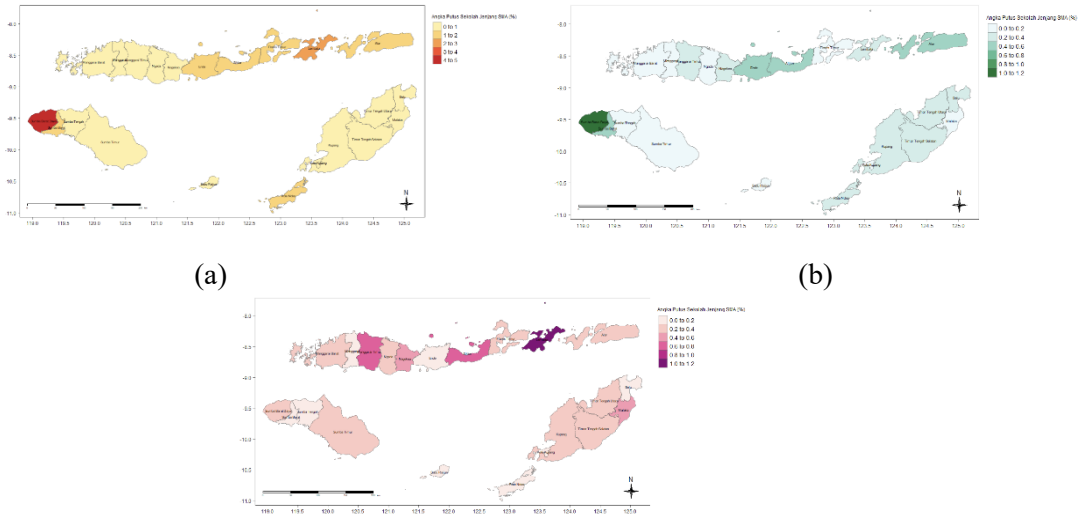
H_0: $\beta_j(u_i, v_i) = 0$, untuk $j = 1, 2, \cdots, p$ dan $i = 1, 2, \cdots, N$

H_1: Terdapat minimal satu $\beta_j(u_i, v_i) \neq 0$, untuk $j = 1, 2, \cdots, p$ dan $i = 1, 2, \cdots, N$

The test statistic used is $T_j(u_i, v_i) = \frac{\hat{\beta}_j(u_i, v_i)}{\hat{\sigma}\sqrt{c_{jj}}}$, $j = 1, 2, \dots, p$ where $T_j(u_i, v_i)$ is the value of T for the independent variable j at location i, $\hat{\beta}_j(u_i, v_i)$ is the parameter estimate for the independent variable j at location i , c_{jj} is the diagonal element j in matrix $C_i C_i'$ with $C_i = (X^{\cdots\prime}W(u_i, v_i)X^{\cdots})^{-1}X^{\cdots\prime}W(u_i, v_i)$, and $\hat{\sigma} = \sqrt{\frac{RSS_{GWPR}}{\delta_1}}$ with $\delta_i = tr((I - L)'(I - L))^i$ for $i = 1, 2$ and I is an identity matrix of size $NT \times NT$. If $|T_j(u_i, v_i)| > t_{\frac{\alpha}{2}; df = \frac{\delta_1^2}{\delta_2}}$ or $p - value < \alpha$, then H_0 is rejected, which means that the independent variable parameter j at the location i has a significant effect on the variable dependent [2].

2. RESULTS

The descriptive statistics show that the high school dropout rate (APS) in East Nusa Tenggara Province during the 2019–2021 period varies across regencies/cities. The data consist of 66 observations obtained from 22 regencies/cities over three years. APS values range from 0% to 4.210%, with the lowest value recorded in Sabu Raijua Regency in 2020 and the highest value observed in Southwest Sumba Regency in 2019. The median APS value is 0.395, while the mean value is higher at 0.570, indicating an asymmetric distribution in which a small number of observations with relatively high dropout rates increase the average value. Furthermore, the coefficient of variation of 115.697 indicates a high degree of dispersion in APS values across the observations. Based on this data summary, the spatial distribution of APS across regencies/cities in NTT Province is illustrated in Figure 1.



(a) (b)

(c)

(a) APS Distribution in 2019, (b) APS Distribution in 2020, (c) APS Distribution in 2021
Figure 1. Map of APS Distribution in the Province of NTT

The maps in **Figure 1** (a), **Figure 1** (b), and **Figure 1** (c) respectively depict the distribution of APS by regency/city in NTT Province in 2019, 2020, and 2021. The darker the color at a location, the higher the APS. APS has a different color in each regency/city. Regencies/cities that have close APS values tend to cluster in certain of the NTT provinces.

Fixed Effect Panel Data parts Regression Model

Fixed effect panel data regression modelling was carried out as an initial analysis of dropout data. First, a multicollinearity test is performed using the VIF value. The result is that there is no multicollinearity. Therefore, the analysis was continued by using all variables by testing individual and time-specific effects. Based on the results of testing individual-specific effects and time-specific effects, it was found that there were only significant individual-specific effects. The following is the estimated result of a fixed effect model with individual-specific effects.

$$\widehat{APS}_{it} = 0.00917264\,\widehat{KP}_{it} + 0.00964202\,\widehat{RMG}_{it} - 8.31004584\,\widehat{UMR}_{it} - 0.02796911\,\widehat{TPT}_{it} - 0.00263528\,\widehat{RMS}_{it} + 1.62597002\,\widehat{RRLS}_{it} + 0.00039147\,\widehat{APIP}_{it}$$

with $i = 1, \ldots, 22$ and $t = 2019, 2020, 2021$. At a significance level of $\alpha = 0.05$, three variables significantly influence APS, namely UMR, RRLS, and APIP. Then, the value of R^2 in the model is 0.64421, which means that the independent variables in the model can explain the APS of 64.421%, and the remaining 35.579% is explained by other variables outside the study.

The analysis is continued by testing the assumption of normality, with $\alpha = 0.05$ obtained that the residuals are normally distributed. In the non-autocorrelation test, with $\alpha = 0.05$ obtained that there is no autocorrelation in the residuals, and in the homoscedasticity test, obtained that the variance of the residuals is not constant. The assumption of homoscedasticity that is not met can indicate the presence of spatial heterogeneity, so the analysis is continued with spatial heterogeneity testing using the Breusch-Pagan test. With $\alpha = 0.05$, obtained $p - value = 0.03191 < \alpha = 0.05$, conclude there is spatial heterogeneity.

Geographically Weighted Panel Regression Model (GWPR) Modeling

In the previous analysis, it was found that there was spatial heterogeneity, so the analysis continued with the formation of the GWPR model using various weighting functions.

Table 1. R^2 and the AIC Model

Model	R^2	AIC
Global Regression Models	0.64421	52.527
GWPR (Fixed Boxcar) Models	0.9064013	-25.00167
GWPR (Fixed Bisquare) Models	0.941587	-54.61223
GWPR (Fixed Gaussian) Models	0.9997704	-404.8798
GWPR (Fixed Tricube) Models	0.9330335	-46.57913
GWPR (Fixed Exponential) Models	0.9999852	-585.1535
GWPR (Adaptive Boxcar) Models	0.7350993	26.45254
GWPR (Adaptive Bisquare) Models	0.9632865	-80.65238
GWPR (Adaptive Gaussian) Models	0.8160078	6.918491
GWPR (Adaptive Tricube) Models	0.9591134	-73.69161
GWPR (Adaptive Exponential) Models	0.8561571	-7.225432

Table 1 shows that the GWPR model with the Fixed Exponential kernel weighting function has the largest R^2 value and the smallest AIC, so further analysis is carried out using the GWPR (Fixed Exponential) model. Based on parameter estimation and parameter significance testing carried out with RStudio software, here is an example of the GWPR (Fixed Exponential) model in Kupang Regency ($i = 6$) which was for $t = 2019, 2020, 2021$.

$$\widehat{APS}^{\cdots}_{6t} = -0.001922\, KP^{\cdots}_{6t} + 0.094373\, RMG^{\cdots}_{6t} - 1.925408\, UMR^{\cdots}_{6t} - 0.057241\, TPT^{\cdots}_{6t} - 0.596723\, RRLS^{\cdots}_{6t} - 0.002414\, RMS^{\cdots}_{6t} - 0.000005\, APIP^{\cdots}_{6t}$$

3.3 Regency/City Grouping Based on the Significance of 7 Variables Using K-Modes Clustering

The first stage in k-modes clustering is to determine the number of k clusters you want to form, in this case, $k = 2, 3, 4$. The optimal number of clusters was determined using the silhouette coefficient criteria. The highest silhouette coefficient value, namely 0.8636364, was obtained when the number of clusters was 4. **Table 1** and **Figure 2** show the results of the grouping.

Table 2. Grouping of Significant Independent Variables from K-Modes Clustering Results

Cluster	Significant Variables	Regency/City
1	KP, RMG, UMR, TPT, RRLS, RMS	Kupang
2	KP, RMG, UMR, RRLS, RMS, APIP	Nagekeo
3	KP, RMG, UMR, TPT, RMS, APIP	Alor
4	KP, RMG, UMR, TPT, RMS, RRLS, APIP	Malaka, Belu, Timor Tengah Utara, Timor Tengah Selatan, Kupang City, Rote Ndao, Sabu Raijua, Sumba Timur, Sumba Tengah, Kabupaten Sumba Barat, Sumba Barat Daya, Manggarai Barat, Manggarai, Manggarai Timur, Ngada, Ende, Sikka, Flores Timur, Lembata.

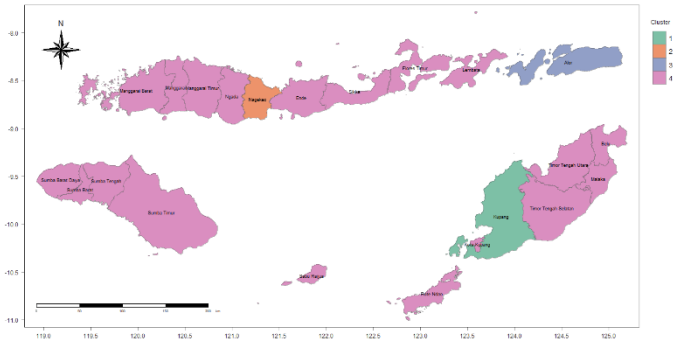


Figure 2. Grouping Map of Significant Independent Variables from K-Modes Clustering Results

4. DISCUSSIONS

Based on the results and discussion of the analysis, it was found that the GWPR model with the Fixed Exponential kernel weighting function was the best model. Population density, student-teacher ratio, regional minimum wage, open unemployment rate, student-to-school ratio, average length of schooling, and Smart Indonesia Program budget have a significant effect on explaining high school dropout rates in at least 21 regencies/cities in NTT Province. Grouping regencies/cities based on variable significance using k-modes clustering produces 4 groups. These results are expected to be a consideration for the government in determining effective policies and programs to address the problem of dropping out of school, which is adapted to the conditions in each district/city of NTT Province.

Research on school dropouts that has been conducted includes research by Sanusi et al. using spatial regression to analyze the factors that influence the dropout rate of high school education in South Sulawesi Province [20]. Research by Temu et al., which uses binary logistic regression to analyze the factors that influence high school dropouts in East Nusa Tenggara Province in 2016 [21]. Research by Bidari & Budiantara, which uses nonparametric truncated spline regression to determine the factors that influence the percentage of school dropouts in East Java [22]. Research using GWPR provides more location-specific results and can provide a higher R^2.

Implications of The Research

Population density, student-teacher ratio, regional minimum wage, and student-to-school ratio have a significant effect on the dropout rate of high school education in all districts/cities of NTT Province. Therefore, it is hoped that the government can pay more attention to population equality, equal distribution of the number of students and teachers, increasing the regional minimum wage, and equalizing the number of students and schools spread across the districts/cities of NTT Province, which is expected to be one solution to eradicating the problem of dropping out of high school education in NTT Province.

The discussion explains the meaning of the results of the research or the illustration. The discussion should be able to answer the meaning of the result, why it happened, how it happened, and why it was different or not significantly different.
In the discussion, it is necessary to make a confrontation or confirmation of a related study, as in the literature review in the introduction (Randolph, 2009). The author explains why it is the same or different from other studies. The author should be able to compare their research results with two or three previous studies.

5. CONCLUSION

Based on the results and discussion of the analysis of the data on high school dropouts in NTT Province in 2019-2021 and the factors that can explain them, the following conclusions were obtained.

A general description of the high school dropout rate in East Nusa Tenggara Province in 2019-2021 can be obtained through descriptive statistics, namely by calculating the number of observations, minimum values, median values, average values, maximum values, and coefficients of variation as well as maps. Based on the calculation results, it was found that the number of observations was 66, with the lowest number being 0% and the highest being 4.210%. Then, it was found that the average value was greater than the median value, and the coefficient of variation was 115.697. The maps show differences and changes in the dropout rate of high school education in districts/cities in NTT Province from 2019 to 2021.

Based on the general overview analysis, the dropout rate of high school education in districts/cities in NTT Province in 2019-2021 has varying values, where the highest number is in Southwest Sumba Regency in 2019 and the lowest number is in Sabu Raijua Regency in 2020. Then, it was found that there is a diversity in the dropout rate values of high school education in NTT Province, where districts/cities that have close dropout rates tend to cluster in certain parts of NTT Province.

Based on the analysis using Geographically Weighted Panel Regression (GWPR) to determine the factors that can explain the dropout rate of high school education in NTT Province in 2019-2021, the

GWPR model with the Fixed Exponential kernel weighting function is the best model compared to the global regression model and the GWPR model with other kernel weighting functions. Population density, student-teacher ratio, regional minimum wage, and student-to-school ratio have a significant effect on the dropout rate of high school education in each district/city in NTT Province. Meanwhile, the open unemployment rate, average length of schooling, and the Smart Indonesia Program budget have a significant effect on the dropout rate of high school education in only 21 districts/cities in NTT Province. Most variables affect the dropout rate of high school education, with varying relationships according to the conditions of each district/city. Grouping of districts/cities based on variables that significantly influence the dropout rate of high school education using k-modes clustering resulted in 4 groups, where the first, second, and third groups had 1 member, and the fourth group had 19 members of the district/city.

The results of this study are expected to be a consideration for the government in determining effective policies and programs to overcome the problem of dropping out of school, which are adjusted to the conditions in each district/city of NTT Province. Population density, student-teacher ratio, regional minimum wage, and student-to-school ratio have a significant effect on the dropout rate of high school education in all districts/cities of NTT Province. Therefore, it is hoped that the government can pay more attention to population equality, equal distribution of the number of students and teachers, increasing the regional minimum wage, and equalizing the number of students and schools spread across the districts/cities of NTT Province, which is expected to be one solution to eradicating the problem of dropping out of high school education in NTT Province. Further research can be conducted in smaller observation areas, such as sub-districts. In this way, programs to eradicate the problem of school dropouts can be designed to be more appropriately targeted to the conditions of each region.

6. REFERENCES

[1] R. Nasir, S. Annas, and M. Nusrang, "Pemodelan dengan Spatial Autoregressive (SAR) pada Angka Putus Sekolah Bagi Anak Usia Wajib Belajar di Provinsi Sulawesi Selatan". *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 3, no. 1, pp. 44-50, 2021. doi:10.35580/variansiunm9358.

[2] A. Rusgiyono, and A. Prahutama, "Geographically Weighted Panel Regression with Fixed Effect for Modeling the Number of Infant Mortality in Central Java, Indonesia". *Media Statistika*, vol. 14, no. 1, pp. 10-20, Jun. 2021. doi: 10.14710/medstat.14.1.10-20.

[3] N. K. A. S. Cahyani, N. L. P. Suciptawati, and I. K. G. Sukarsa, "Identifikasi Faktor yang Memengaruhi Anak Putus Sekolah di Kabupaten Badung". *E-Jurnal Matematika*, vol. 8, no. 4, pp. 289-297, Nov. 2019. doi:10.24843/MTK.2019.v08.i04.p267.

[4] Kemendikbudristek, *Statistik Persekolahan SMA 2021/2022*. Tangerang Selatan: Pusdatin Kemendikbudristek, 2022.

[5] R. A. Sirait, "Pengaruh Jarak ke Sekolah terhadap Angka Partisipasi dan Putus Sekolah SMP di Indonesia". *J. budg.*, vol. 4, no. 1, pp. 24-42, Dec. 2022. [Online]. Available from https://ejurnal.dpr.go.id/index.php/jurnalbudget/article/view/24.

[6] B. H. Baltagi, *Econometric Analysis of Panel Data*, 6th ed., New York: John Wiley & Sons, 2021.

[7] D. N. Gujarati, and D. C. Porter, *Basic Econometrics*, 5th ed., New York: McGraw-Hill Education, 2009.

[8] R. M. Kunst, "Econometric Methods fo Panel Data—Part II". *Consultado el*, vol. 27, pp. 1-9, April. 2009. [Online]. Available from https://homepage.univie.ac.at/robert.kunst/panels2e.pdf .

[9] I. Tamara, D. Ispriyanti, and A. Prahutama, "Pembentukan Model Spasial Data Panel *Fixed Effect* Menggunakan GUI Matlab (Studi Kasus : Kemiskinan di Jawa Tengah)". *Jurnal Gaussian*, vol. 5, no. 3, pp. 417-426, Aug. 2016. doi:10.14710/j.gauss.5.3.417-426.

[10] R. Celik, "A New Test To Detect Monotonic and Non Monotonic Types For Heteroscedasticity". *Journal of Applied Statistics*, vol. 44 no. 2, pp. 342-361, Jan. 2017. doi:10.1080/02664763.2016.1169258. Publishers, 1988.

[11] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis,* 5th ed., New Jersey: John Wiley & Sons, 2012.

[12] L. Anselin, *Spatial Econometrics: Methods and Models*. Dodrect: Kluwer Academic

[13] S. M. Meutuah, H. Yasin, and D. A. I. Maruddani, "Pemodelan *Fixed Effect Geographically Weighted Panel Regression* untuk Indeks Pembangunan Manusia di Jawa Tengah". *Jurnal Gaussian*, vol. 6, no. 2, pp. 241-250, Apr. 2017. doi: 10.14710/j.gauss.6.2.241-250.

[14] J. E. H. Percival, *et al.*, "Exploratory spatial data analysis with gwpcorMapper: an interactive mapping tool for geographically weighted correlation and partial correlation". *Journal of Geovisualization and Spatial Analysis*, vol. 6, no. 1, pp. 1-18, May. 2022. doi: 10.1007/s41651-022-00111-3.

[15] S. Georganos, *et al.*, "Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression". *Journal of Arid Environments*, vol. 146, pp. 64–74, May. 2017. doi:10.1016/j.jaridenv.2017.06.004.

[16] J. Wang, K. Chen, and X. Song, "Differences Among Influencing Factors of China's Provincial Energy Intensity: Empirical Analysis from a Geographically Weighted Regression Model". *Polish Journal of Environmental Studies,* vol. 29, no.4, pp. 2901-2916, Apr. 2020. doi:10.15244/pjoes/113097.

[17] A. S. Fotheringham, C. Brundson, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. England: John Wiley and Sons, Ltd, 2002.

[18] J. Z. Huang, *Clustering Categorical Data with k-Modes*. Hong Kong: IGI Global, 2009.

[19] J. Han, M. Kamber, and J. Pei. *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann, 2012.

[20] Sanusi, W., Ihsan, H., & Syam, N. H. (2018). Model Regresi Spasial dan Aplikasinya dalam Menganalisis Angka Putus Sekolah Usia Wajib Belajar di Provinsi Sulawesi Selatan. Journal of Mathematics, Computations, and Statistics, 1(2), 183-192.

[21] Temu, C. C., Tolok, M. S., Azmi, P. V., & Marsisno, W. (2019). Faktor-Faktor yang Memengaruhi Putus Sekolah Usia SMA di Provinsi NTT Tahun 2016. Seminar Nasional Official Statistics 2019, 2019(1), 583-592.

[22] Bidari, D. R., & Budiantara, I. N. (2020). Pemodelan Faktor yang Mempengaruhi Presentase Anak Putus Sekolah di Jawa Timur Menggunakan Regresi Nonparametrik Spline Truncated. Jurnal Sains dan Seni ITS, 1(2), 2337-3520.