JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3.6784

SPATIAL MODELING OF SCHOOL DROPOUT RATES IN UNDERDEVELOPED AREAS OF PAPUA USING GEOGRAPHICALLY WEIGHTED REGRESSION

Harun Al Azies[1,2*], Hani Brilianti Rochmanto[3], Fenny Fitriani[3]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[2]Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, 50131, Semarang, Indonesia
[3]Department of Statistics, Faculty of Science and Technology, Universitas PGRI Adi Buana, Surabaya 60234, Indonesia

Keywords:
Disadvantaged Regions, Fixed Gaussian Kernel, Geographically Weighted Regression, School Dropout, Multiple Linear Regression.

Corespondent Email:
harun.alazies@dsn.dinus.ac.id

Abstrak. Makalah ini mengidentifikasi faktor–faktor yang diduga mempengaruhi kasus putus sekolah pada daerah tertinggal di Provinsi Papua serta menyelidiki adanya pengaruh geografis. Tujuan dari penelitian ini adalah adalah mendapatkan estimasi dan statistik uji parameter model daerah tertinggal di Provinsi Papua dengan pendekatan Geographically Weighted Regression (GWR), dan mengetahui faktor-faktor yang mempengaruhi angka putus sekolah daerah tertinggal sehingga dapat digunakan sebagai referensi pemerintah dalam menentukan arah kebijakan untuk menanggulangi masalah putus sekolah pada daerah tertinggal. Hasil penelitian menunjukkan angka putus sekolah pada daerah tertinggal di Papua tertinggi pada jenjang SMP, dan ada indikasi bahwa kasus angka putus sekolah menyebar secara spasial karena terdapat heterogenitas antar lokasi pengamatan yang artinya jika suatu daerah memiliki kasus putus sekolah yang tingi atau sebaliknya, ada kemungkinan daerah sekitarnya memiliki beban yang sama, sehingga menggunakan pemodelan regresi spasial dengan fungsi Fixed Gaussian Kernel. Hasil pengelompokan dengan GWR menghasilkan dua kelompok berdasarkan variabel signifikan. Adapun variabel yang signifikan yaitu rasio murid-guru jenjang pendidikan SMP, rasio murid-rombongan belajar jenjang pendidikan SMP, dan angka putus sekolah (APTs) jenjang SD.

Abstract. *This study examines the factors hypothesized to contribute to school dropout rates in disadvantaged regions of Papua Province and explores potential geographical influences. The primary aims are to derive parameter estimates and statistical tests for the model of underdeveloped regions in Papua using Geographically Weighted Regression (GWR) and to determine the factors influencing school dropout rates in these areas, providing a basis for governmental policy development to mitigate school dropout issues in disadvantaged regions. Findings reveal that the highest dropout rates occur at the junior high school level, with indications of spatial clustering in dropout cases due to heterogeneity among observation sites. This suggests that regions with elevated dropout rates, or conversely low rates, are likely to have neighboring areas with comparable patterns, necessitating the use of spatial regression modeling with a Fixed Gaussian Kernel function. GWR analysis resulted in two clusters based on significant variables, which include the student-teacher ratio at the junior high school level, the student-classroom* *ratio at the junior high school level, and the elementary school dropout rate (APTs).*

1. INTRODUCTION

Education serves as a vital instrument for unlocking and nurturing individual potential, fostering competitive capabilities [1]. Consequently, ensuring access to high-quality education is imperative. UNICEF emphasizes universal access to quality education [2], a stance reinforced by the World Conference on Education, which advocates for all children globally to access and complete quality education [3]. This aligns with global efforts over recent decades to enhance investments in educational infrastructure [4]. However, school dropout remains a significant global challenge in education [5].

School dropout is a widespread and recurring issue across nations and educational levels, constituting a substantial problem for many countries [6]. Global data highlight persistently elevated dropout rates, with the 2017 International Commission on Financing's Global Education Opportunity report indicating that approximately 250 million children and adolescents dropped out of school in 2016 [1]. Dropout is a complex issue, extending beyond academic underperformance or behavioral challenges. It represents a dynamic process influenced by social and environmental contexts, including family background, socio-academic environments, demographic factors, and other variables [7].

A critical educational challenge is mitigating high dropout rates, particularly at the junior high school level, which poses significant personal, familial, and societal implications worldwide [8]. Previous research indicates that secondary and upper secondary school dropout rates globally range from 17% to 50%, with higher rates in developing countries [9], [10]. In Indonesia, dropout rates remain notably high. At the elementary level, Indonesia ranks sixth among six countries with the highest dropout numbers (2 million students). At the junior high school level, Indonesia and Myanmar rank fifth among six countries with the highest dropout rates (1.9 million) [11].

High dropout rates constitute a major challenge in Indonesia. Despite significant enrollment increases—110% at the elementary level and 101% at the junior high school level in 2015 [12]—approximately 2.4 million elementary and junior high school students failed to complete their education. Consequently, Indonesia ranks 56th out of 127 countries in global dropout rankings [12]. Papua Province exhibits particularly high dropout rates, driven by limited educational access. According to the Central Statistics Agency, Papua records the second-highest school dropout rate (APTs) in Indonesia. Moreover, most areas in Papua are classified as disadvantaged regions under Presidential Regulation No. 131 of 2015 on Underdeveloped Regions for 2015–2019.

Prior studies on high dropout rates in Papua [13], who utilized nonparametric spline regression to identify economic factors, infrastructure, and elementary school dropout rates as significant contributors to junior high school dropout rates in Papua. Survival analysis was applied to assess the risk of adolescent dropouts in Papua [14]. However, no studies have specifically investigated factors influencing dropout rates in Papua's disadvantaged regions. Given that dropout factors vary by region based on local characteristics, identifying these factors while accounting for geographical influences is essential. Geographically Weighted Regression (GWR), a spatial modeling approach, incorporates geographical factors as independent variables affecting the response variable [15]. GWR extends global regression models, building on nonparametric regression principles [16]. Applications of GWR include studies on poverty in East Java's disadvantaged regions [16], infant mortality and stunting in Indonesia [17], and the Human Development Index in Papua using geographically weighted ridge regression [18]. Additionally, GWR with Principal Components Analysis was employed to examine the impact of environmental degradation on economic growth in Indonesia [19].

This study introduces a novel spatial approach to identify factors contributing to high dropout rates in Papua's disadvantaged regions. It is the first to model school dropout rates in these regions using Geographically Weighted Regression. The objectives are to estimate model parameters and test statistics for disadvantaged regions in Papua using GWR and to identify factors influencing dropout rates, providing evidence-based insights for government policies to address school dropout challenges in disadvantaged areas.

2. TINJAUAN PUSTAKA

Isi bagian tinjauan pustaka ditulis ringkas, dan hanya teori yang benar-benar digunakan sebagai dasar penelitian. (*The contents of the literature review section are written briefly, and only theory is actually used as a basis for research.*)

3. METHODS

3.1. Data Sources and Research Variables

This study utilizes secondary data from the Central Bureau of Statistics (BPS) of Papua Province, specifically the 2020 Papua Province Education Indicators Dataset, which provides continuous data and information on educational indicator trends. The data are derived from the processing of primary data collected by BPS Papua through the annual National Socio-Economic Survey (Susenas), supplemented by secondary data from relevant institutions. The unit of observation for modeling factors influencing school dropout rates in Papua's disadvantaged regions is each regency/city classified as underdeveloped. According to Presidential Regulation No. 131 of 2015 on the Designation of Underdeveloped Regions for 2015 – 2019, 26 out of 29 regencies in Papua Province are categorized as disadvantaged. These regions are detailed in Table 1 below.

Table 1. Underdeveloped Regencies in Papua Province

No.	Regency/City	No.	Regency/City
1	Merauke	14	Sarmi
2	Jayawijaya	15	Keerom
3	Nabire	16	Waropen
4	Kepulauan Yapen	17	Supiori
5	Biak Numfor	18	Mamberamo Raya
6	Paniai	19	Nduga
7	Puncak Jaya	20	Lanny Jaya
8	Boven Digoel	21	Mamberamo Tengah
9	Mappi	22	Yalimo
10	Asmat	23	Puncak
11	Yahukimo	24	Dogiyai
12	Pegunungan Bintang	25	Intan Jaya
13	Tolikara	26	Deiyai

(Source: Presidential Regulation No. 131 of 2015)

This study employs eight predictor variables and one response variable. The variables utilized in this research are detailed as follows.

Table 2. Research Variables

Variable Type	Indicator
Response Variable	Junior High School Dropout Rate (APTs) (Y)
Predictor Variables	Student-Teacher Ratio at Junior High School Level (X1)
	Student-Classroom Ratio at Junior High School Level (X2)
	Classroom-Class Ratio at Junior High School Level (X3)
	Percentage of Qualified Teachers at Junior High School Level (X4)
	Elementary School Dropout Rate (APTs) (X5)
	Percentage of Non-Food Expenditure for Middle-Income Groups (X6)
	Percentage of Non-Food Expenditure for High-Income Groups (X7)
	Distribution of Households with Floor Area ≤19 m² (X8)

(Source: Central Bureau of Statistics of Papua Province, 2020)

3.2. Analysis Stages

The analytical steps employed to address the research objectives are outlined as follows:

1. Conducting descriptive analysis of school dropout cases in disadvantaged regions of Papua Province.
2. Modeling the School Dropout Rate (APTs) using classical linear regression with the Ordinary Least Square (OLS) method [20].
3. Implementing the spatial modeling stages using Geographically Weighted Regression, which include:
   a. Determining the optimal bandwidth based on the Cross Validation (CV) criterion [21], with calculations performed to achieve the minimum CV value [22].

   b. Estimating GWR model parameters using the optimal bandwidth.
4. Comparing the results of OLS and GWR models based on the coefficient of determination (R^2) and the sum of squared errors (SSE) [23].
   a. Assessing model fit to evaluate the influence of geographical factors on school dropout occurrences.
   b. Testing the significance of parameters individually.
5. Interpreting and drawing conclusions from the obtained results.

3.3. Concept of Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is an extension of the regression model in which parameters are estimated at each geographical location [24], resulting in distinct regression parameter values for each spatial point [25], [26]. GWR builds upon the global regression framework, with its foundational concept derived from nonparametric regression [27]. In the GWR model, the response variable (y) is predicted using predictor variables, with regression coefficients that vary depending on the location of the observed data. The GWR model can be expressed as follows [15], [28].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \qquad (1)$$

where:

y_i : The observed value of the response variable at location *(i)*

(u_i, v_i) : The coordinates (longitude, latitude) of location *(i)*

$\beta_k(u_i, v_i)$: The regression coefficient for the *(k)*-th predictor variable at location *(i)*

The weighting function in the GWR model is critical, as it represents the spatial relationships between observed data points. A kernel function is employed to estimate parameters in the GWR model when the distance function (w_j) is continuous and monotonically decreasing [25]. The weighting functions derived from kernel functions include the Gaussian Distance Function, Exponential Function, Bisquare Function, and Tricube Kernel Function [29]. In this study, the weighting function utilized is the Gaussian distance function [15].

$$w_j(u_i, v_i) = exp\left[-\tfrac{1}{2}\left(d_{ij}/h\right)^2\right] \qquad (2)$$

where $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$ represents the Euclidean distance between location (u_i, v_i) and location (u_j, v_j), and (h) is a non-negative parameter, commonly referred to as the smoothing parameter (bandwidth).

The bandwidth can be conceptualized as the radius of a circle, where points within this radius are considered to have an influence [30]. In constructing a GWR model, the bandwidth plays a critical role, as it affects the model's accuracy by balancing variance and bias [31]. Several methods exist for selecting the optimal bandwidth, one of which is the Cross Validation (CV) method [15], mathematically defined as follows:

$$CV(h) = \sum_{i=1}^{n}(y_i - \hat{y}_{\neq i}(h))^2 \qquad (3)$$

Let $\hat{y}_{\neq i}(h)$ be the estimated value at location (u_i, v_i), where the observation at that location is excluded from the estimation process. To obtain the optimal value of h, it is selected such that it yields the minimum cross-validation (CV) value. Subsequently, hypothesis testing for the Geographically Weighted Regression (GWR) model involves two main aspects: testing the model fit and testing the model parameters. The hypothesis test for model fit in GWR is formulated as follows:

- H_0: $\beta_k(u_i, v_i) = \beta_k$ on every $k = 0,1,2,\cdots,p$, and $i = 1,2,\cdots,n$
  (There is no significant difference between the global regression model and the GWR model.)
- H_1 : At least one $\beta_k(u_i, v_i) \neq \beta_k$, $k = 0,1,2,\cdots,p$
  (There is significant difference between the global regression model and the GWR model.)

The test statistic for this hypothesis is defined as follows:

$$F_{count} = \frac{RSS(H_1)/\left(\frac{\delta_1^2}{\delta_2}\right)}{RSS(H_0)/(n-p-1)} \qquad (4)$$

The significance test for model parameters at each location is conducted through partial (individual) parameter testing. The hypotheses are stated as follows:

- $H_0: \beta_k(u_i, v_i) = 0$
- $H_1: \beta_k(u_i, v_i) \neq 0$ with $k = 1,2, \cdots, p$

The test statistic used is:

$$T_{hit} = \frac{\hat{\beta}_k(u_i,v_i)}{\hat{\sigma}\sqrt{c_{kk}}} \quad (5)$$

4. RESULT AND DISCUSSION

4.1. General Overview of School Dropout Cases in Underdeveloped Regions of Papua

The school dropout rate by educational level is defined as the percentage of students who, in the current academic year, have discontinued their studies before completing a particular level of education, relative to the number of students who were enrolled at the same level in the previous academic year.

According to data released by the Statistics Indonesia (BPS) of Papua Province (Figure 1), the lower secondary education level (junior high school or equivalent) exhibited the highest school dropout rate compared to other educational levels. In 2020, approximately 44 to 45 out of every 1,000 individuals dropped out at the junior high school level. At the primary school (or equivalent) level, 24 to 25 out of every 1,000 students discontinued their education before completion. In contrast, the dropout rate at the senior high school (or equivalent) level was the lowest, with only 7 to 8 out of every 1,000 students dropping out, which is around three students lower than the national average

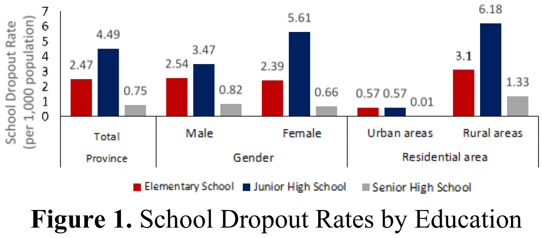

Figure 1. School Dropout Rates by Education Level and Characteristics in Papua Province, 2020 (Source: BPS, Susenas)

Figure 2. Distribution of Junior High School (SMP) Dropout Rates in Underdeveloped Areas of Papua Province, 2020 (Source: BPS Papua, processed data)

Figure 2 illustrates that among the 26 underdeveloped regions in Papua Province, the highest dropout rate at the junior high school level in 2020 was found in Nduga Regency, with a dropout rate of 29.52 per 1,000 population. This means that approximately 29 to 30 out of every 1,000 individuals dropped out of junior high school or equivalent education in Nduga. Meanwhile, 11 regions recorded a dropout rate of zero, indicating that none of the 1,000 students enrolled in junior high school or equivalent education dropped out. These regions include Asmat, Jayawijaya, Yapen Islands, Puncak Jaya, Boven Digoel, Tolikara, Sarmi, Waropen, Mamberamo Raya, Mamberamo Tengah, and Dogiyai Regencies.

4.2. Linear Regression Model

The results and discussion of the linear regression model can be used to examine the relationship between school dropout rates in each underdeveloped regency in Papua Province and the factors suspected to influence them. Simultaneous parameter testing involves testing all parameters in the regression model collectively. The hypotheses are as follows:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$

H_1: At least one $\beta_k \neq 0$ $(k = 1,2,..8)$

Table 3. Results of Simultaneous Significance Test for the Regression Model

Degree of freedom (df)	$F_{statistic}$	$F_{(0,1;8;17)}$	P_{-value}	R^2
8;17	4.347	1.0226	0.0052	67.16%

The simultaneous test of the regression model using the Ordinary Least Squares (OLS) method yielded an F-statistic of 4.347, which is greater than the critical value $F_{(0.1;8,17)}$=2.061, and a p-value smaller than the significance level (α) of 10%. Therefore, the null hypothesis H_0 is rejected (Table 3). The OLS regression model produces parameters that are statistically significant at the 10% significance level. This indicates that at least one predictor variable significantly influences the response variable, which is the school dropout rate in underdeveloped regions of Papua.

The coefficient of determination (R^2) is 67.16%, meaning that the regression model explains 67.16% of the variability in the dropout rates in these regions, while the remaining 32.84% is explained by other variables outside the model.

Next, the results of the partial parameter tests of the regression model are presented as follows.

Table 4. Results of Partial Significance Tests for the Regression Model

Variable	Coefficient	t_{value}	P_{value}
Student-teacher ratio at junior high school level (X1)	-1.197	-2.31	0.033*
Student-study group (rombel) ratio at junior high level (X2)	1.156	2.08	0.053*
Class-study group (rombel) ratio at junior high level (X3)	-25.4	-1.64	0.118
Percentage of qualified teachers at junior high level (X4)	0.196	0.96	0.350
School dropout rate at primary school level (X5)	1.440	3.97	0.001*
Percentage of non-food expenditure for middle-income group (X6)	-0.281	-1.14	0.271
Percentage of non-food expenditure for high-income group (X7)	0.273	1.17	0.257
Distribution of households with floor area ≤ 19 m² (X8)	0.0953	1.36	0.191

*Significant at the 10% significance level ($\alpha = 0.10$)

Based on the results of the partial significance test presented in Table 4, it can be concluded that the variables that have a significant influence on school dropout cases in underdeveloped regions of Papua are: the student-teacher ratio at the junior high school level (X1), the student–study group (rombel/rombongan belajar) ratio at the junior high school level (X2), and the school dropout rate at the primary school level (X5). The resulting regression model is as follows:

$$y = -9{,}7\ 1{,}197\,X1 + 1{,}156\,X2 - 25{,}4\,X3 + 0{,}196\,X4 + 1{,}440\,X5 - 0{,}281\,X6 + 0{,}273\,X7 + 0{,}0953\,X8 \quad (6)$$

4.3. Results of the Geographically Weighted Regression (GWR) Model for School Dropout Cases

The Geographically Weighted Regression (GWR) model extends traditional regression by estimating parameters at each geographical location, resulting in spatially varying regression coefficients. The steps for developing the GWR model are outlined below.

4.3.1. Heteroskedasticity Test

This diagnostic test assesses the presence of spatial heterogeneity, which is crucial for determining the appropriate spatial model for analyzing school dropout rates. The Breusch-Pagan test [32] is employed to evaluate the homogeneity of variance in the residuals. A robust GWR model is indicated by the presence of heteroskedasticity. The test results are presented in Table 5 below.

Table 5. Results of the Heteroskedasticity Test for the GWR Model

Degree of Freedom (df)	Breusch-Pagan Score	P-value
8	20.775	0.0078

From Table 5, the p-value of 0.00777 is less than the significance level (α) of 10%, leading to the rejection of the null hypothesis (H_0). This indicates that the variance of residuals in the model is non-homogeneous. Consequently, the GWR approach is suitable for addressing spatial heterogeneity in the linear regression model.

4.3.2. Selection of Optimal Weighting

The GWR model employs weighting based on the geographical location of each regency/city. The optimal weighting is determined using the Cross Validation (CV) criterion, with results for each weighting method shown in Table 6.

Table 6. Selection of Optimal Weighting Based on Cross Validation (CV) Score

Weighting Method	Cross Validation (CV) Score
Fixed Gaussian	33.16338
Fixed Bi-Square	34.70170

A weighting method is considered optimal when it yields the lowest CV score. Table 6 indicates that the Fixed Gaussian weighting is optimal for GWR modeling, as it has a lower CV score compared to other weighting methods.

4.3.3. Simultaneous Testing of the GWR Model

The GWR model is tested to determine the influence of location-specific factors in the disadvantaged regions of Papua Province on school dropout occurrences. The hypotheses formulated for constructing the GWR model are as follows.

- $H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \beta_3(u_i, v_i) = \beta_4(u_i, v_i) = \beta_5(u_i, v_i) = \beta_6(u_i, v_i) = \beta_7(u_i, v_i) = \beta_8(u_i, v_i) = \beta_k$
- H_1: minimal ada satu $\beta_k(u_i, v_i) \neq \beta_k$

Table 7. Results of GWR Model Testing

Model Estimation	SSE	(df)	F-Statistic	P-value
GWR Model	8.0744	16.462	0.5098	0.5127
Linear Regression	8.2090	9		

At a significance level (α) of 10%, the null hypothesis (H_0) is rejected because the F-statistic (0.5098) is less than the critical value $F_{(0.1;\,9;\,16.462)} = 2.0553$

Thus, it is concluded that there is no significant difference between the GWR and linear regression models. However, the GWR model demonstrates greater efficiency than the OLS regression model, as evidenced by its lower Sum of Squared Errors (SSE) value compared to that of the OLS model

4.3.4. Selection of the Best Model

The optimal regression model is selected based on the goodness-of-fit criteria, specifically the coefficient of determination (R^2) and the sum of squared errors (SSE). A higher R^2 value indicates a better model fit compared to other models, while a lower SSE further supports model quality. Table 8 presents the goodness-of-fit measures for the linear regression and GWR models.

Table 8. Selection of the Best Model

Model	R^2	SSE
Linear Regression	67.16%	8.2090
GWR	67.70%	8.0744

Table 8 indicates that the overall R^2 value for the GWR model is higher than that of the linear regression model. This suggests that the GWR model is more suitable for modeling school dropout cases in disadvantaged regions of Papua Province in 2020.

4.3.5. Mapping of Disadvantaged Regions

The GWR model is better suited for modeling school dropout cases in Papua's disadvantaged regions, as discussed in Table 8, compared to the OLS model, due to the presence of spatial heterogeneity, which indicates a non-stationary spatial process with varying variance across observation regions. The parameters generated by the GWR model are location-specific, tailored to the data observed in each region. Factors influencing school dropout cases in Papua's disadvantaged regions vary spatially, reflecting the dominant challenges in each area. Different influencing factors necessitate tailored policies for each region. Therefore, partial testing of each parameter in every observed region is conducted. Based on partial testing of the GWR model parameters using the t-statistic, the significance of the model parameters is evaluated with the following hypotheses.

- $H_0: \beta k\,(ui, vi) = 0$
- $H_1: \beta k(ui, vi) \neq \beta k;\ i = 1,2,\ldots,26, k = 1,2,3..,8$

At a significance level (α) of 10%, the critical t-value $t_{(0.1;\,17)}$ is 2.110. The following presents the grouping of regions based on significant variables.

Table 9. Distribution of Regions Based on Influential Variables

Significant Variables	Regency/City
1. Student-Teacher Ratio at Junior High School Level (X1) 2. Student-Classroom Ratio at Junior High School Level (X2) 3. Elementary School Dropout Rate (APTs) (X5)	Nabire, Sarmi, Mamberamo Tengah, Kepulauan Yapen, Waropen, Yalimo, Biak Numfor, Supiori, Puncak, Paniai, Mamberamo Raya, Dogiyai, Puncak Jaya, Nduga, Intan Jaya, Tolikara, Lanny Jaya, Deiyai
1. Student-Teacher Ratio at Junior High School Level (X1) 2. Student-Classroom Ratio at Junior High School Level (X2)	Merauke, Asmat, Jayawijaya, Yahukimo, Boven Digoel, Pegunungan Bintang, Mappi, Keerom

Based on Table 9, the factors influencing school dropout cases in disadvantaged regions of Papua Province vary across regions. The GWR model results, as shown in Table 9, indicate that three variables significantly affect school dropout rates in Papua's disadvantaged regions: the student-teacher ratio at the junior high school level (X1), the student-classroom ratio at the junior high school level (X2), and the elementary school dropout rate (APTs) (X5). A total of 18 regencies are influenced by all three factors, while eight regencies are influenced by two factors. The following presents the mapping derived from Table 9.

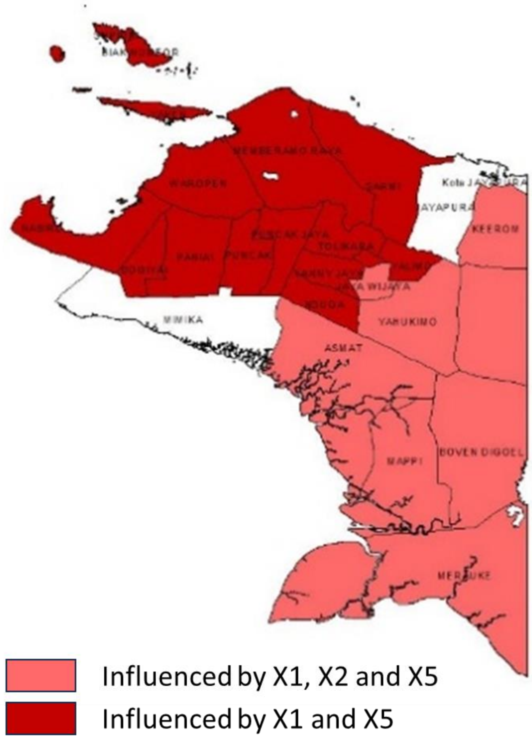

Figure 3. Mapping of Disadvantaged Regions Based on the Influence of Predictor Variables on School Dropout Cases in Papua

5. CONCLUSIONS

This study successfully identifies key factors influencing dropout rates in Papua's disadvantaged regions. Linear regression results show that the student-teacher ratio (X1), student-classroom ratio (X2) at the junior high school level, and the elementary school dropout rate (X5) significantly impact dropout cases. The Geographically Weighted Regression model was selected using a fixed Gaussian kernel for its lowest cross-validation (CV) score. GWR outperforms multiple linear regression with a higher R^2 value (67.70%) and lower SSE. The analysis reveals spatial variation in dropout determinants, with GWR identifying two distinct clusters characterized by different significant variables. These findings highlight the need for spatially adaptive education policies tailored to local conditions.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Study Program in Informatics Engineering and the Research Center for Quantum Computing and Materials

Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, for the invaluable support and facilities provided throughout this research. Special thanks are also extended to the Department of Statistics, Faculty of Science and Technology, Universitas PGRI Adi Buana, for the academic contributions and collaboration.

REFERENCES

[1] N. P. Stromquist, "The Learning Generation: Investing in Education for a Changing World; A Report by the International Commission on Financing Global Education Opportunity," *https://doi.org/10.1086/690064*, vol. 61, no. 1, pp. 214–217, Feb. 2017, doi: 10.1086/690064.

[2] L. Sumardi, "Why Students Dropout? Case Study of Dropout Attributions in West Nusa Tenggara Province, Indonesia," *Cross-Currents: An International Peer-Reviewed Journal on Humanities & Social Sciences*, vol. 6, no. 6, pp. 85–91, Jun. 2020, doi: 10.36344/CCIJHSS.2019.V06I06.006.

[3] S. M. Shahidul and A. H. M. Z. Karim, "FACTORS CONTRIBUTING TO SCHOOL DROPOUT AMONG THE GIRLS: A REVIEW OF LITERATURE," *European Journal of Research and Reflection in Educational Sciences*, vol. 3, no. 2, 2015, Accessed: May 28, 2025. [Online]. Available: www.idpublications.org

[4] J. R. Behrman and A. B. Deolalikar, "School Repetition, Dropouts, and the Rates of Return to Schooling: The Case of Indonesia," *Oxf Bull Econ Stat*, vol. 53, no. 4, pp. 467–480, 1991, Accessed: May 28, 2025. [Online]. Available: https://ideas.repec.org/a/bla/obuest/v53y1991i4p467-80.html

[5] F. Pastore and K. F. Zimmermann, "Understanding school-to-work transitions," *Int J Manpow*, vol. 40, no. 3, pp. 374–378, May 2019, doi: 10.1108/IJM-06-2019-343.

[6] Setyadharma and Andryan, "Upper secondary school dropout : lessons from central Java province, Indonesia : a dissertation presented in partial fulfilment of the requirement for the degree of Doctor of Philosophy in Economics at Massey University, Manawatū, New Zealand," 2017, *Massey University*. Accessed: May 28, 2025. [Online]. Available: http://hdl.handle.net/10179/11854

[7] A. Sharma and A. Samantaray, "Demographic Analytical Study of Girl Child Dropout from Schools in India," *International Journal of Engineering Technology Science and Research*, vol. 4, no. 10, pp. 921–26, 2017, [Online]. Available: www.ijetsr.com

[8] O. Hetlevik, T. Bøe, and M. Hysing, "GP-diagnosed internalizing and externalizing problems and dropout from secondary school: a cross-sectional study," *Eur J Public Health*, vol. 28, no. 3, pp. 474–479, Jun. 2018, doi: 10.1093/EURPUB/CKY026.

[9] M. Cui, C. A. Darling, M. Lucier-Greer, F. D. Fincham, and R. W. May, "Parental Indulgence: Profiles and Relations to College Students' Emotional and Behavioral Problems," *J Child Fam Stud*, vol. 27, no. 8, pp. 2456–2466, Jul. 2018, doi: 10.1007/S10826-018-1076-6/METRICS.

[10] J. E. Lansford, K. A. Dodge, G. S. Pettit, and J. E. Bates, "A Public Health Perspective on School Dropout and Adult Outcomes: A Prospective Study of Risk and Protective Factors From Age 5 to 27 Years," *Journal of Adolescent Health*, vol. 58, no. 6, pp. 652–658, Jun. 2016, doi: 10.1016/J.JADOHEALTH.2016.01.014.

[11] UNESCO Institute for Statistic (UIS), "Leaving no one behind: how far on the way to universal primary and secondary education? UIS UNESCO," 2016.

[12] Kemdikbud, "Data Pokok Pendidikan Dasar dan Menengah Tahun 2015," Jakarta, 2015.

[13] L. Mubarokah, i N. Budiantara, and M. Ratna, "Pemodelan Angka Putus Sekolah Usia SMP Menggunakan Metode Regresi Nonparametrik Spline di Papua," *Jurnal Sains dan Seni ITS*, vol. 5, no. 1, pp. D103–D108, Mar. 2016, doi: 10.12962/J23373520.V5I1.14697.

[14] R. A. Rahma and I. M. Arcana, "TINGKAT RISIKO PUTUS SEKOLAH PADA REMAJA DI PROVINSI PAPUA TAHUN 2018," *Seminar Nasional Official Statistics*, vol. 2020, no. 1, pp. 672–681, Jan. 2020, doi: 10.34123/SEMNASOFFSTAT.V2020I1.468.

[15] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression*. Chichester UK: John Wiley & Sons, 2002.

[16] R. Hidayat N, B. W. Otok, Z. Mahsyari, S. H. Sa'diyah, and D. A. Fadhila, "Geographically Weighted Regression for Prediction of Underdeveloped Regions in East Java Province Based on Poverty Indicators," pp. 898–907, Jun. 2019, doi: 10.5220/0007553708980907.

[17] F. Cholid, D. Trishnanti, and H. Al Azies, "Pemetaan Faktor-Faktor yang Mempengaruhi Stunting pada Balita dengan Geographically Weighted Regression

(GWR)," *SEMNAKes*, pp. 156–165, 2019, doi: 10.17605/OSF.IO/9MZU7.
[18] D. R. S. Saputro, R. D. Hastutik, and P. Widyaningsih, "The modeling of human development index (HDI) in Papua - Indonesia using geographically weighted ridge regression (GWRR)," *AIP Conf Proc*, vol. 2326, no. 1, Feb. 2021, doi: 10.1063/5.0040329/1000572.
[19] H. Al Azies, "Analisis Pengaruh Pengendalian Pencemaran Dan Kerusakan Lingkungan Terhadap Pertumbuhan Ekonomi Di Indonesia Menggunakan Pendekatan Geographically Weighted Regression Principal Components Analysis (GWRPCA)," *Prosiding Seminar Nasional Energi*, vol. 8, no. 1, pp. 18–27, 2019.
[20] J. T. Kilmer and R. L. Rodríguez, "Ordinary least squares regression is indicated for studies of allometry," *J Evol Biol*, vol. 30, no. 1, pp. 4–12, Jan. 2017, doi: 10.1111/JEB.12986.
[21] A. Mohammadinia, A. Alimohammadi, and B. Saeidian, "Efficiency of Geographically Weighted Regression in Modeling Human Leptospirosis Based on Environmental Factors in Gilan Province, Iran," *Geosciences 2017, Vol. 7, Page 136*, vol. 7, no. 4, p. 136, Dec. 2017, doi: 10.3390/GEOSCIENCES7040136.
[22] B. Lu, C. Brunsdon, M. Charlton, and P. Harris, "Geographically weighted regression with parameter-specific distance metrics," *International Journal of Geographical Information Science*, vol. 31, no. 5, pp. 982–998, May 2017, doi: 10.1080/13658816.2016.1263731.
[23] K. Isbiyantoro, Y. Wilandari, and M. Jurusan Statistika FSM UNDIP, "PERBANDINGAN MODEL PERTUMBUHAN EKONOMI DI JAWA TENGAH DENGAN METODE REGRESI LINIER BERGANDA DAN METODE GEOGRAPHICALLY WEIGHTED REGRESSION," *Jurnal Gaussian*, vol. 3, no. 3, pp. 461–469, 2014, doi: 10.14710/J.GAUSS.3.3.461-469.
[24] L. Chao, K. Zhang, Z. Li, Y. Zhu, J. Wang, and Z. Yu, "Geographically weighted regression based methods for merging satellite and gauge precipitation," *J Hydrol (Amst)*, vol. 558, pp. 275–289, Mar. 2018, doi: 10.1016/J.JHYDROL.2018.01.042.
[25] M. Charlton and S. Fotheringham, "Geographically Weighted Regression," *White Paper. National Centre for Geocomputation, National University of Ireland Maynooth*, pp. 1–14, 2009.
[26] S. P. Ganiswari, H. Al Azies, A. Nugraha, A. Luthfiarta, and G. A. Firmansyah, "Data-Driven Modeling of Human Development Index in Eastern Indonesia's Region Using Gaussian Techniques Empowered by Machine Learning," *Journal of Applied Geospatial Information*, vol. 7, no. 2, pp. 1004–1010, Nov. 2023, doi: 10.30871/JAGI.V7I2.6757.
[27] R. Hidayat N, B. W. Otok, Z. Mahsyari, S. H. Sa'diyah, and D. A. Fadhila, "Geographically Weighted Regression for Prediction of Underdeveloped Regions in East Java Province Based on Poverty Indicators," *Proceedings of the 2nd International Conference Postgraduate School*, pp. 898–907, Jan. 2018, doi: 10.5220/0007553708980907.
[28] G. A. Firmansyah, J. Zeniarja, H. Al Azies, S. winarno, and S. P. Ganiswari, "Machine Learning-Enhanced Geographically Weighted Regression for Spatial Evaluation of Human Development Index across Western Indonesia," *Journal of Applied Geospatial Information*, vol. 7, no. 2, pp. 996–1003, Nov. 2023, doi: 10.30871/JAGI.V7I2.6755.
[29] N. F. T. A. Nugroho and I. Slamet, "Geographically Weighted Regression Model with Kernel Bisquare and Tricube Weighted Function on Poverty Percentage Data in Central Java Province," *J Phys Conf Ser*, vol. 1025, no. 1, p. 012099, May 2018, doi: 10.1088/1742-6596/1025/1/012099.
[30] J. A. Yacim and D. G. B. Boshoff, "A comparison of bandwidth and kernel function selection in geographically weighted regression for house valuation," *International Journal of Technology*, vol. 10, no. 1, pp. 58–68, 2019, doi: 10.14716/IJTECH.V10I1.975.
[31] A. R. da Silva and F. F. Mendes, "On comparing some algorithms for finding the optimal bandwidth in Geographically Weighted Regression," *Appl Soft Comput*, vol. 73, pp. 943–957, Dec. 2018, doi: 10.1016/J.ASOC.2018.09.033.
[32] A. G. Klein *et al.*, "The detection of heteroscedasticity in regression models for psychological data," *Psychol Test Assess Model*, vol. 58, no. 4, pp. 567–592, 2016.