

Spatial Analysis of Human Development Index in Indonesia

Jesica Putri Widyastuti^{1*}, Dewi Retno Sari Saputro², Purnami Widyaningsih³

^{1,2,3} Department of Mathematics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University
Ir. Sutami No. 36, Surakarta, 57126, Indonesia

Corresponding Author.

*Email: jesjesica@student.uns.ac.id

Abstract: This study develops a province level HDI classification model for Indonesia by applying the GWOLR approach with an Adaptive Gaussian Kernel weighting scheme. HDI values are grouped into three ordered categories: low, medium, and high. The research data covers 38 provinces in Indonesia in 2024, sourced from the Central Statistics Agency. The predictor variables used include the percentage of the poor population, the open unemployment rate, the school enrollment rate for ages 16-18, the reading interest rate, the number of villages/subdistricts with hospital facilities, and the percentage of households with access to adequate sanitation. The selection of the optimal bandwidth was performed using a k-nearest neighbor based CV method, resulting in an optimal bandwidth of $k = 13$ with a CV value of 11.5577. The results of the simultaneous test indicate that, collectively, the predictor variables have a significant effect on the HDI category, with a test statistic value G^2 of 71.1677. Based on the results of the partial test, 13 out of 38 provinces have at least one variable with a significant effect. The variable representing the percentage of households with access to adequate sanitation was the most influential variable in 13 provinces, followed by the variable representing school enrollment rates for 16-18 year olds in 12 provinces. Model evaluation using APER shows a classification error rate of 7.89% and an accuracy of 92.11%.

Keywords: HDI, GWOLR, Adaptive Gaussian Kernel, Spatial Analysis, Cross Validation

© 2026 International Conference on Multidisciplinary Engagement. All rights reserved.

1. INTRODUCTION

The Human Development Index (HDI) serves as a composite measure capturing three fundamental aspects of human well-being: longevity and health, educational attainment, and adequate living standards [1]. The HDI is widely recognized as a key indicator for assessing the quality of life of a population and the development performance of a country or region because it combines the dimensions of health, education, and income into a single index that can be compared across regions and over time [2], [3]. Based on figures released by the Badan Pusat Statistik [4], Indonesia's HDI increased from 72.81 in 2020 to 75.08 in 2024. Although the national trend shows an increase, disparities among provinces remain clearly evident. In 2024, DKI Jakarta recorded the highest HDI value of 83.08, while Papua Pegunungan reached only 54.24. This range illustrates significant development disparities between Western and Eastern Indonesia. Tracing the socioeconomic determinants of such variation calls for a regression based framework capable of pinpointing the factors that drive provincial human development outcomes [5].

In various studies, the HDI is often classified into several categories of development levels, such as low, medium, and high [6], [7]. When the response variable is a categorical variable with a natural order, the commonly used approach is Ordinal Logistic Regression (OLR). This model allows for modeling the relationship between predictor variables and the probability that a region falls into a specific development category. Differences in geographical, socioeconomic, and regional characteristics can mean that the relationship between factors determining human development is not always the same across all regions. Consequently, a modeling approach that accounts for location specific aspects is essential to gain a more comprehensive understanding of the factors influencing the HDI across various regions

The response variable in this study consists of ordinal-scale categorical data, making the OLR approach a methodologically appropriate choice. However, the OLR model produces global parameters that ignore locational diversity [8]. This shortcoming is overcome by the Geographically Weighted Ordinal Logistic Regression (GWOLR) model. The GWOLR model produces unique parameter estimates at each observation location, allowing the relationship between predictor variables and HDI categories to vary across provinces [9], [10].

Recent studies have applied spatial regression approaches to socio-economic indicators, particularly using geographically weighted regression and its extensions. Dong et al. [9] applied GWOLR to life satisfaction data, while Fariz et al. [12] modeled the Human Development Index in Indonesia using a GWOLR approach. However, most previous studies rely on fixed kernel functions and do not explore adaptive bandwidth selection in depth. Therefore, this study extends previous work by implementing an Adaptive Gaussian Kernel with k-nearest neighbor bandwidth selection for modeling ordinal HDI data.

2. METHOD

This section presents the statistical foundations underlying the proposed GWOLR approach, covering the model specification, kernel weighting function, bandwidth selection, hypothesis testing, multicollinearity detection, and model evaluation.

2.1 Visualization of Bibliometric Networks (VOSviewer)

A bibliometric review of 332 Scopus-indexed publications (2015–2025) using the keywords “geographically weighted regression” and “logistic regression” via VOSviewer [14], [15] revealed three dominant research clusters, as shown in Figure 1.

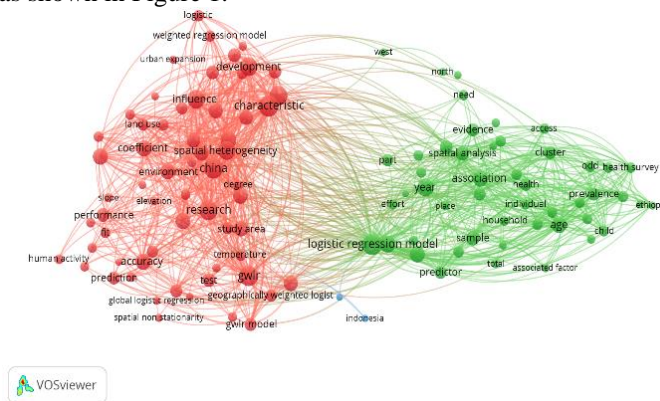


Figure 1. Bibliometric network of GWR and logistic regression models

Figure 1 shows three clusters of the bibliometric network. The red cluster primarily contains terms such as spatial heterogeneity, coefficient of variation, geographically weighted logistic regression (GWLR), and spatial nonstationarity; the green cluster primarily covers topics related to logistic regression models, spatial analysis, and health studies; and the blue cluster includes regional topics such as Indonesia.

2.2 Geographically Weighted Ordinal Logistic Regression (GWOLR) Model

The GWOLR model is a development of the Geographically Weighted Regression (GWR) model and the ordinal logistic regression model. This model is used to model the relationship between an ordinal scaled categorical response variable and predictor variables whose regression coefficients depend on the location where the data is observed [17]. For a response variable with G categories, the GWOLR model for location i can be written as

$$\text{logit} [P(Y_i \leq g|X_i)] = \alpha_g(u_i, v_i) + X_i^T \beta(u_i, v_i), \quad g = 1, 2, \dots, G - 1 \tag{1}$$

Where $P(Y_i \leq g|X_i)$ is the cumulative probability of the k-th category g for X_i , $X_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip}]^T$ is the vector of predictor variables at the k-th observation location i, $\alpha_g(u_i, v_i)$ is the intercept parameter that satisfies $\alpha_1(u_i, v_i) \leq \alpha_2(u_i, v_i) \leq \dots \leq \alpha_{G-1}(u_i, v_i)$, $\beta(u_i, v_i) = [\beta_1(u_i, v_i) \ \beta_2(u_i, v_i) \ \dots \ \beta_p(u_i, v_i)]^T$ is the vector of regression coefficients for the k-th location i, and (u_i, v_i) are the geographic coordinates (longitude and latitude) of the k-th location i. The cumulative probability can be written as

$$P(Y_i \leq g|X_i) = \frac{e^{\alpha_g(u_i, v_i) + X_i^T \beta(u_i, v_i)}}{1 + e^{\alpha_g(u_i, v_i) + X_i^T \beta(u_i, v_i)}}, \quad g = 1, 2, \dots, G - 1. \quad (2)$$

GWOLR parameters are estimated by maximising the likelihood function through iterative numerical procedures approach, which yields a system of nonlinear equations; thus, a numerical method is required to obtain the parameter estimation solution, written as follows [18]

$$\ell(\theta(u_i, v_i)) = \sum_{i=1}^n \sum_{g=1}^G y_{ig} \ln(\pi_g(X_i)) w_j(u_i, v_i). \quad (3)$$

Parameters are estimated via the Newton-Raphson iteration [19], [20].

2.3 Adaptive Gaussian Kernel Weighting Function

The Adaptive Gaussian Kernel function is employed within the GWOLR framework to assign location-specific weights, expressed as

$$w_j(u_i, v_i) = \exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{b_{i(p)}}\right)^2\right) \quad (4)$$

where $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ is the Euclidean distance between locations i and j , and $b_{i(p)}$ is the adaptive bandwidth that defines p as the distance between locations i and j [21].

2.4 Selection of Optimal Bandwidth

A common approach for identifying the optimal bandwidth value is Cross Validation (CV) [10]. The CV criterion is expressed as

$$CV(h) = \sum_{i=1}^n \sum_{g=1}^G (y_{\neq i, g} - \hat{\pi}_{\neq i, g}(h))^2 \quad (5)$$

Where $y_{\neq i, g}$ is an indicator variable such that $y_{\neq i, g} = 1$ if the i -th observation falls into the category g , and 0 otherwise. The value $\hat{\pi}_{\neq i, g}(h)$ is the estimated probability obtained by omitting the i -th location from the estimation process. The value h that yields the minimum CV is selected as the optimal bandwidth.

2.5 Parameter Testing

Hypothesis testing in the GWOLR model involves simultaneous parameter testing as well as partial parameter testing [5] [22].

a. Simultaneous testing uses the statistic G^2

Hypothesis:

$H_0: \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0$ (all predictor variables together have no effect)

$H_1: \beta_j(u_i, v_i) \neq 0; j = 1, 2, \dots, p$ (there are predictor variables that influence the response variable)

Test statistic:

$$G^2 = -2 \left(\sum_{i=1}^n \sum_{g=1}^G \ln \left[\frac{\sum_{j=1}^n y_{jg} w_j(u_i, v_i)}{\sum_{j=1}^n w_j(u_i, v_i)} \right] - \sum_{i=1}^n \sum_{g=1}^G \ln[\hat{\pi}_g^*(x_i)] \right)$$

Test criteria: Reject the null hypothesis (H_0) if $G^2 > \chi_{(\alpha, p)}^2$ or p -value $< \alpha$.

b. Partial test using the Wald test statistic

Hypothesis:

$H_0: \beta_j(u_i, v_i) = 0; j = 1, 2, \dots, p$ (the predictor variable has no effect on the response variable)

$H_1: \beta_j(u_i, v_i) \neq 0; j = 1, 2, \dots, p$ (the predictor variable affects the response variable)

Test statistic: $Z_j = \frac{\hat{\beta}_j(u_i, v_i)}{SE(\hat{\beta}_j(u_i, v_i))}$ or $W_j^2 = \frac{(\hat{\beta}_j(u_i, v_i))^2}{(SE(\hat{\beta}_j(u_i, v_i)))^2}$ with $SE(\hat{\beta}_j(u_i, v_i)) = \sqrt{\text{var}(\hat{\beta}_j(u_i, v_i))}$

Test criteria: reject the null hypothesis (H_0) if the alternative hypothesis ($|Z_j| > Z_{\alpha/2}$) holds, or reject the alternative hypothesis (H_0) if the null hypothesis ($W_j^2 > \chi_{(a, 1)}^2$) holds.

2.6 Multicollinearity Test

Multicollinearity arises when two or more predictor variables exhibit strong linear dependence [23]. It is detected using the Variance Inflation Factor (VIF) test statistic. The VIF values for the first through j predictor variables is defined as

$$VIF_j = \frac{1}{1-R_j^2} \tag{6}$$

where R_j^2 is the coefficient of determination of the regression model when the k -th predictor variable (j) is regressed against the other predictor variables. The test criterion used is that the assumption of non-multicollinearity is satisfied if the value of $VIF_j < 10$; however, if the value of $VIF_j \geq 10$, then the variable indicates high multicollinearity, meaning the assumption of non-multicollinearity is not satisfied [24].

2.7 Model Evaluation

The model’s classification performance is quantified using the Apparent Error Rate (APER), which is the proportion of misclassified observations [25]. The APER evaluation is expressed as

$$APER = \left(\frac{\text{Number of incorrect predictions}}{\text{Number of correct predictions}} \right) \times 100\% \tag{7}$$

Therefore, to determine the classification accuracy of the model, the following equation can be used is total accuracy rate=1-APER.

2.8 Research Data and Analysis Steps

This study is an applied research study that applies the GWOLR model to provincial HDI data in Indonesia. The research method is divided into two parts: research data and research steps. All data are secondary, retrieved from <http://www.bps.go.id>, and encompass all 38 Indonesian provinces for the reference year 2024. The complete variable list is summarised in Table 1.

Table 1. Response and predictor variables of the research data

Variable	Research Data
Y	Human Development Index (HDI) Categories: low category (IPM < 70), medium category (70 ≤ IPM ≤ 75), high category (IPM > 75)
X ₁	Percentage of the poor population
X ₂	Open unemployment rate
X ₃	School Enrollment Rate (SER) for ages 16–18
X ₄	Reading interest rate
X ₅	Villages/subdistricts with health facilities such as hospitals
X ₆	Percentage of households with access to adequate sanitation
u _i	Longitude
v _i	Latitude

The analysis involves the following six steps.

1. Perform descriptive statistics to describe the characteristics of the data.
2. Detect multicollinearity using the VIF value.
3. Performing GWOLR modeling: determining provincial coordinates, calculating Euclidean distances, select optimal bandwidth via cross validation, calculating the Adaptive Gaussian Kernel weights, and estimating parameters using the Newton-Raphson iteration.
4. Conducting simultaneous and partial parameter testing.
5. Evaluating classification accuracy using APER.
6. Model interpretation.

3. RESULTS AND DISCUSSION

3.1 Description of Research Data

The research data covers 38 provinces in Indonesia in 2024 with varying HDI distributions across provinces. Based on this data, a diverse distribution of HDI categories was obtained. There are 7 provinces with a low HDI category, 25 provinces with a medium HDI category, and 6 provinces with a high HDI category. The description of the research variables is presented in Table 2.

Table 2. Description of research variables

Variable	Min	Max	Mean	Median	Std. Dev
Y	53.42	83.08	72.39	73.18	5.15
X ₁	3.90	31.32	10.91	9.85	6.55
X ₂	1.32	6.75	4.38	4.19	1.41
X ₃	47.65	90.36	75.07	75.53	7.53
X ₄	38.83	79.99	67.39	69.18	8.34
X ₅	9	346	69.39	42	86.15
X ₆	12.61	96.83	81.15	83.36	15.07

Table 2 presents the descriptive statistics of all variables. The HDI values range from 53.42 to 83.08, with a mean of 72.39, indicating moderate variation in development levels across provinces. Among the predictor variables, sanitation access (X₆) shows a relatively high average value of 81.15%, while school enrollment rates (X₃) also demonstrate a high mean of 75.07%. In contrast, the number of health facilities (X₅) exhibits substantial variability, with values ranging from 9 to 346 and a large standard deviation of 86.15, reflecting significant disparities in infrastructure availability across regions. Overall, these variations suggest that differences in basic infrastructure and education access may contribute to the observed disparities in HDI across provinces.

3.2 Multicollinearity Test

To assess whether problematic linear dependence exists among the predictors, VIF values were computed according to Equation (6). The resulting VIF estimates for the six predictor variables are presented in Table 3.

Table 3. VIF values of predictor variables

Variable	VIF Value
X ₁	3.72
X ₂	1.39
X ₃	2.13
X ₄	4.62
X ₅	1.59
X ₆	4.43

Table 3 shows that all predictor variables have VIF values below 10. The highest VIF value is 4.62 for the variable X₄. This indicates that there is no multicollinearity in the research data.

3.3 Selection of Optimal Bandwidth

Bandwidth optimisation used the CV criterion of Equation (5), evaluating k from 4 to 15 nearest neighbors. CV values and associated misclassification rates are tabulated in Table 4.

Table 4. Results of bandwidth selection using the CV method

k	CV Value	Misclassification Rate
4	24.8735	0.3421
5	19.4363	0.2632
6	19.8493	0.2895
7	19.2521	0.2895
8	15.2376	0.2105
9	15.0818	0.2368
10	13.3133	0.2105
11	12.1349	0.2105
12	12.4801	0.2368
13	11.5577	0.2105
14	11.7764	0.1842
15	13.2996	0.2105

Table 4 shows that the smallest CV value was obtained at k = 13 with a CV value of 11.5577. Therefore, the optimal bandwidth used in the modeling is k = 13 nearest neighbors. In the parameter estimation process for each province, the model uses weighting information calculated based on the distance to the 13 nearest provinces. Provinces with closer geographical distances will receive larger weights compared to those farther away, ensuring that local parameter estimates more accurately reflect the spatial characteristics of the region. Inter-provincial weights are computed via the Adaptive Gaussian Kernel (Equation 4).

3.4 Estimation of GWOLR Model Parameters

GWOLR parameter estimates were conducted locally for each province using the maximum likelihood approach with the weighted likelihood function presented in Equation (3). The estimation process utilized the Newton-Raphson iteration for each province. The cumulative logit model used in this study follows the form of the model in Equation (1). Table 5 presents estimation results and partial test outcomes for West Sumatra Province, selected as a representative case given its comparatively high number of significant predictors.

Table 5. GWOLR model parameter estimates for West Sumatra Province

Parameter	Estimate	Standard Error	Calculated Z	p-value
α_1	82.2297	9.0160	9.1205*	0.0000
α_2	101.3860	3.6697	27.6278*	0.0000
β_1	-0.8145	0.8268	-0.9852	0.3245
β_2	2.0480	1.7086	1.1987	0.2307
β_3	0.7316	0.1003	7.2955*	< 0.0001
β_4	-0.1463	0.0283	-5.1762*	< 0.0001
β_5	-0.0020	0.0432	-0.0470	0.9625
β_6	0.5371	0.0853	6.2940*	< 0.0001

*) Significant at $\alpha = 5\%$ ($|Z| > 1,96$)

Table 5 shows that of the six predictor variables, three have a significant effect in West Sumatra Province, namely the school participation rate for ages 16-18 (X_3), the reading interest rate (X_4), and the percentage of households with access to adequate sanitation (X_6). Meanwhile, the percentage of the poor population (X_1), the open unemployment rate (X_2), and villages/subdistricts with health facilities in the form of hospitals (X_5) do not have a significant effect in that location. These findings highlight that educational access and sanitation infrastructure are pivotal determinants of human development in West Sumatra.

3.5 Simultaneous Test of the GWOLR Model

A global test was performed to assess whether the full predictor set jointly determines HDI category. Based on the calculation results, the test statistic value was obtained as $G^2 = 71.1677$ with degrees of freedom $df = 6$. This value is greater than the critical value $\chi^2(0,05; 6) = 12,592$, triggering rejection of H_0 . Collectively, therefore, at least one predictor variable exerts a statistically significant influence on HDI category in Indonesia.

3.6 Partial Test of the GWOLR Model

Province-level partial tests identified locally significant predictors at $\alpha = 5\%$ with a critical value of $Z_{0,05} = 1,96$. The partial test for the first location, namely West Sumatra Province, Z-statistics were benchmarked against the critical value, with standard errors and Z values from RStudio summarised in Table 6.

Table 6. Partial testing of GWOLR model parameters for West Sumatra Province.

Parameter	Estimate	Standard Error	Calculated Z
β_3	0.7316	0.1003	7.2955
β_4	-0.1463	0.0283	-5.1762
β_6	0.5371	0.0853	6.2940

All 38 provinces are classified by their significant variable profiles in Table 7 and spatially visualised in Figure 2.

Table 7. Provincial groups based on significant variable patterns

Group	Province	Number
No significant variables	Aceh, North Sumatra, East Java, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Southwest Papua, Papua, South Papua, Central Papua, Papua Pegunungan	25
Significant in X_3 and X_6	Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Jakarta, West Java, Central Java, Special Region of Yogyakarta, Banten	11
Significant $X_3, X_4,$ and X_6	West Sumatra	1
Significant X_1 and X_6	Riau Islands	1

Table 7 shows that 13 of 38 provinces have at least one significant predictor. Sanitation access (X_6) is significant in the most provinces 13, followed by school enrollment for ages 16–18 (X_3) in 12 provinces, reading preference (X_4) in West Sumatra alone, and poverty rate (X_1) only in Riau Islands. Unemployment (X_2) and hospital-equipped settlements (X_5) show no local significance in any province.



Figure 2. Mapping of provinces based on significant variables in the GWOLR model

3.7 Evaluation of the GWOLR Model

Predictive performance was assessed by cross-tabulating predicted and observed HDI categories in a confusion matrix, detailed in Table 8.

Table 8. Classification matrix of HDI category prediction accuracy

	Prediction Cat. 1	Prediction Cat. 2	Prediction Cat. 3
Actual Cat. 1 ($n = 7$)	5	2	0
Actual Category 2 ($n = 25$)	0	25	0
Actual Category 3 ($n = 6$)	0	1	5

Applying Equation (7) to Table 8 gives $APER = 3/38 = 7.89\%$. Thus, the classification accuracy of the GWOLR model is 92.11%. These results indicate that the model correctly classified 35 provinces, consisting of 5 provinces in the low HDI category, 25 provinces in the medium category, and 5 provinces in the high category. Classification errors occurred in two provinces with a low HDI category that were predicted as medium, as well as one province with a high HDI category that was predicted as medium. This accuracy level indicates that the GWOLR model has good capability in classifying the HDI categories of provinces in Indonesia.

3.8 Interpretation of the GWOLR Model

Drawing on the significant West Sumatra parameters β_3 , β_4 , and β_6 . Therefore, based on Table 6 an HDI logit model can be formulated for West Sumatra Province, expressed as

$$\begin{aligned} \text{logit}[P(Y \leq 1 | x_i)] &= 82.2297 + 0.7316X_3 - 0.1463X_4 + 0.5371X_6 \\ \text{logit}[P(Y \leq 2 | x_i)] &= 101.3860 + 0.7316X_3 - 0.1463X_4 + 0.5371X_6 \end{aligned}$$

From this model, the cumulative odds for each HDI category can be calculated using the logistic function as shown in Equation (2). The variable for school enrollment rates among 16–18 year olds (X_3) has a positive coefficient of 0.7316 with an odds ratio of $e^{0.7316} = 2.078$. This value indicates that every 1 unit increase in the school enrollment rate for ages 16–18 increases the odds of a province being in a higher HDI category by 2.078 times. Meanwhile, the variable for reading enthusiasm (X_4) has a negative coefficient of -0.1463 with an odds ratio of $e^{-0.1463} = 0.864$. The coefficients obtained reflect the influence of each variable while accounting for other variables in the model, particularly SER, which is also closely related to educational aspects. There may be indirect influences between variables that cause the reading preference coefficient to be negative in certain regions. Therefore, these results do not imply that reading preference is unimportant, but rather indicate that relationships between variables in a spatial context can be complex. Furthermore, the variable percentage of households with access to adequate sanitation (X_6) has a positive coefficient of 0.5371 with an odds ratio of

$e^{0.5371} = 1.711$. This implies that each one-unit improvement in household sanitation coverage raises the provincial odds of attaining a higher HDI category by a factor of 1.711.

The results indicate that sanitation access (X_6) is the most dominant variable influencing HDI classification across multiple provinces. This finding highlights the critical role of basic infrastructure in improving human development outcomes. Access to adequate sanitation directly contributes to better health conditions, which is one of the main components of the HDI. The school enrollment rate (X_3) is also found to be a significant predictor in many provinces. This result emphasizes the importance of education in determining human development levels. Compared to previous studies, this finding suggests that the effect of education varies spatially, meaning that improvements in educational access may have different impacts depending on regional characteristics.

Interestingly, the reading interest variable (X_4) shows a negative coefficient in West Sumatra. This result may be influenced by indirect relationships with other educational variables or overlapping effects within the model. It indicates that the relationship between literacy indicators and HDI is complex and may depend on broader structural and socio-economic factors. Overall, the findings confirm that the determinants of human development are not spatially uniform, and region-specific analysis is necessary to better understand development disparities.

4. CONCLUSION

This study applies the GWOLR model with an Adaptive Gaussian Kernel to analyze the spatial variation of HDI classification across 38 provinces in Indonesia in 2024. The results of the simultaneous test $G^2 = 71.1677$ indicate that the predictor variables collectively have a statistically significant effect on HDI categories. Based on the partial test, 13 out of 38 provinces show at least one significant predictor. Among the variables, sanitation access (X_6) is significant in 13 provinces, followed by school enrollment rates (X_3) in 12 provinces, indicating that these factors are associated with regional differences in human development. Model evaluation using APER yields a classification error rate of 7.89%, corresponding to an accuracy of 92.11% with 35 correctly classified provinces out of 38. The analysis is based on cross-sectional data from a single year and captures spatial heterogeneity through location-specific weighting, but it does not explicitly model spatial dependence between regions. Future studies may consider extending this approach using spatio-temporal models or spatial autoregressive frameworks.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all parties who contributed to this study, particularly in the development of the mathematical formulation, the selection of appropriate statistical models, as well as the estimation and validation procedures. Intensive scientific discussions regarding model assumptions, inferential consistency, and predictive performance evaluation have significantly strengthened the methodological validity of this research. The authors also acknowledge the support in data provision and computational resources, which enabled the efficient implementation of algorithms and numerical simulations. Such contributions have ensured that the results obtained are both reliable and reproducible.

REFERENCES

- [1] United Nations Development Programme, "Human Development Report 1990: Concept and Measurement of Human Development," UNDP, New York, NY, USA: Oxford University Press, 1990. Accessed: Aug. 20, 2025. [online] available: <https://hdr.undp.org/content/human-development-report-1990>.
- [2] S. Susilawati and D. Didiharyono, "Application of principal component regression in analyzing factors affecting human development index," *J. Varian*, vol. 6, no. 2, 2023, doi: 10.30812/varian.v6i2.2366.
- [3] B. Dašić et al., "Human development index in a context of human development: Review on the Western Balkans countries," *Brain Behav.*, vol. 10, 2020, doi: 10.1002/brb3.1755.
- [4] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2024*. Jakarta, Indonesia: Badan Pusat Statistik, 2024.
- [5] D. R. S. Saputro, K. R. Demu, and P. Widyaningsih, "Nonparametric truncated spline regression model on the data of human development index (HDI) in Indonesia," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, p. 012219, 2018.

-
- [6] E. Embuai, R. Salhuteru, M. Sopaheluwakan, and R. Djami, “Ordinal probit regression modeling on the human development index Maluku and North Maluku provinces,” *Parameter: J. Mat. Stat. Terap.*, 2025, doi: 10.30598/parameter.v4i1pp73-80.
- [7] N. Senaviratna and T. Cooray, “Diagnosing multicollinearity of logistic regression model,” *Asian Journal of Probability and Statistics*, 2019. doi: 10.9734/ajpas/2019/v5i230132.
- [8] A. Agresti, *Categorical Data Analysis*, 2nd ed. New York, NY, USA: John Wiley & Sons, 2002.
- [9] G. Dong, T. Nakaya, and C. Brunson, “Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data,” *Comput. Environ. Urban Syst.*, vol. 70, pp. 35–42, 2018.
- [10] M. Rifada and Purhadi, “Pemodelan tingkat kerawanan demam berdarah dengue di Kabupaten Lamongan dengan pendekatan geographically weighted ordinal logistic regression,” in *Pros. Semin. Nas. Statistika Univ. Diponegoro*, 2011, pp. 114–126.
- [11] A. Rahman, N. Gusriani, and D. A. Kusuma, “Pemetaan zonasi risiko COVID-19 di Provinsi Jawa Barat menggunakan model geographically weighted ordinal logistic regression (GWOLR),” *Teorema: Teori dan Riset Mat.*, vol. 7, no. 1, pp. 193–204, 2022.
- [12] F. Fariz, I. Zain, and E. O. Permatasari, “Pemodelan indeks pembangunan manusia (IPM) di Indonesia menggunakan geographically weighted ordinal logistic regression (GWOLR),” *J. Sains dan Seni ITS*, vol. 12, no. 2, pp. D159–D166, 2023.
- [13] A. S. Fotheringham, C. Brunson, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, U.K.: John Wiley & Sons, 2002.
- [14] D. Fitria, A. Husaeni, A. Bayu, and D. Nandiyanto, “Bibliometric using VOSviewer with Publish or Perish (using Google Scholar data): From step-by-step processing for users to the practical examples in the analysis of digital learning articles in pre and post COVID-19 pandemic,” *ASEAN Journal of Science and Engineering.*, vol. 2, pp. 19-46, 2021. doi: 10.17509/ajse.v2i1.37368.
- [15] A. Kirby, “Exploratory bibliometrics: Using VOSviewer as a preliminary research tool,” *Publications*, vol. 11, no. 1, 2023.
- [16] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas., “Cross validation pitfalls when selecting and assessing regression and classification models,” *J. Cheminformatics*, vol. 6, Art. no. 10, 2014.
- [17] H. Khaulasari and F. Suhaeri, “Mapping of gender inequality in Indonesia based on influencing factors using geographically weighted ordinal logistic regression,” *BAREKENG: J. Math. Its Appl.*, vol. 18, no. 1, pp. 233–244, 2024.
- [18] P. Widyaningsih, D. R. S. Saputro, and A. N. Putri, “Fisher scoring method for parameter estimation of geographically weighted ordinal logistic regression (GWOLR) model,” *J. Phys. Conf. Ser.*, vol. 855, no. 1, p. 012060, 2017.
- [19] S. Zuhdi, D. R. S. Saputro, and P. Widyaningsih, “Parameters estimation of geographically weighted ordinal logistic regression (GWOLR) model,” *J. Phys. Conf. Ser.*, vol. 855, no. 1, p. 012064, 2017.
- [20] M. Ohyver, P. Purhadi, and A. Choiruddin, “Parameter estimation of geographically and temporally weighted elastic net ordinal logistic regression,” *Mathematics*, 2025, doi: 10.3390/math13081345.
- [21] C. Chasco, I. Garcia, and J. Vicens, “Modeling spatial variations in household disposable income with geographically weighted regression,” *Munich Personal RePEc Archive (MPRA)*, Paper No. 1682, 2007.
- [22] M. Ohyver, P. Purhadi, and A. Choiruddin, “Hypothesis testing of spatial and temporal weighted ordinal logistic regression,” in *Proc. Int. Conf. Math. Stat.*, 2024, doi: 10.1145/3686592.3686608.
- [23] D. N. Gujarati and D. C. Porter, *Dasar-Dasar Ekonometrika*, 5th ed. Jakarta, Indonesia: Salemba Empat, 2010.
- [24] Z. Xue, Y. Xue, and G. Hu, “Geographically weighted regression analysis for spatial economics data: A Bayesian recourse,” *Int. Reg. Sci. Rev.*, vol. 44, pp. 582–604, 2020, doi: 10.1177/0160017620959823.
- [25] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey, NJ, USA: Pearson Education, 2007.