JOURNAL OF STATISTICAL METHODS AND DATA SCIENCE

Journal of Statistical Methods and Data Science
Volume 01 Nomor 01 Juni 2023
https://jurnalmipa.unri.ac.id/jsmds

MODELLING GENDER DEVELOPMENT INDEX IN CENTRAL JAVA IN 2021 WITH GEOGRAPHICALLY WEIGHTED REGRESSION

Zulfanita Dien Rizqiana[1]
[1] UIN Raden Mas Said Surakarta
e-mail: zulfanita.dr@staff.uinsaid.ac.id

Article Info:
Received: 31-05-2023
Accepted: 30-06-2023
Available Online: 30-06-2023

Keywords:
gender development index
gwr
kernel fixed gaussian
.

Abstract: *Regression analysis is a statistical analysis used to determine and model the influence of the relationship between one response variable and one or more predictor variables using the Ordinary Least Squares (OLS) method. In some cases involving spatial data, it can lead to violations of heterogeneity and autocorrelation, indicating the presence of spatial effects. One regression analysis that can address spatial effects is Geographically Weighted Regression (GWR). The Gender Development Index is an index used to measure the achievement of basic human development capabilities in various areas within a region, taking gender into consideration. This study aims to model the GDI of Central Java Province in 2021 using the GWR approach. The data used in this study are secondary data from 35 Districts/Cities in Central Java Province in 2021. Based on the GWR modeling analysis above, it can be concluded that the GWR model is the best model for modeling the GDI in Central Java Province in 2021 with an R^2 value by 51.82% and AIC value of 159.0621. In the formed GWR model, the significant variables are X_1 or average length of schooling, X_4 or school participation rate, and X_5 or gender ratio.*

1. INTRODUCTION

Regression analysis is a statistical analysis used to understand and model the relationship between one response variable and one or more predictor variables. Parameter estimation for regression analysis is done using the Ordinary Least Squares (OLS) method (Franklin et al., 2017). There are residual assumptions that must be met when using OLS for parameter estimation, including the assumptions of normally distributed residuals, no multicollinearity, no heteroscedasticity, and no autocorrelation. In some cases involving spatial data, violations of heteroscedasticity and autocorrelation may occur, indicating the presence of spatial effects, so OLS method cannot be used. Therefore, the appropriate model for spatial relationships with spatial effects is spatial regression (Safitri, 2019). One regression analysis that can address spatial effects is Geographically Weighted Regression (GWR), which uses a point-based approach. Spatial heterogeneity is caused by differences in conditions across regions. Thus, parameter estimation in the GWR model requires a weighting matrix (Ratiwi & Ayuningsih, 2023).

The Gender Development Index (GDI) is an index used to measure the achievement of basic human development capabilities in various fields in a region, taking gender into consideration. Gender is not only defined based on biological sex but also related to integrity, personality, etc. The components of GDI include health, knowledge, and economic indicators (Sari, 2018). The Gender Development Index (GDI) for Central Java Province in 2021 ranked 10th nationally with a GDI value of 92.48% (BPS, 2023). This GDI value indicates that the gender gap in Central Java Province is relatively low. However, the distribution of GDI values is still uneven as some districts/cities in Central Java Province have GDI values lower than the provincial GDI value. This is a concern for the Central Java Province government to pay more attention to gender equality.

GDI is influenced by several factors, including average length of schooling, labor force participation rate, open unemployment rate, school participation rate, and sex ratio (Elisa, 2022; Safitri, 2019; Aprilianti & Setiadi, 2022; Amaliah & Riniwati, 2021). Previous studies on GDI include Insiro et al., (2023) which modelling GDI in Province West Java with Nonparametric Penalized Spline Regression that the result was the model had determination coefficient (R^2) value by 78.85% and had six predictor variable significantly , Rachmawati et al., (2022) which prediction of GDI in West Kalimantan using Parabolic Trend Method that had result was the difference obtained from the original data to project data value by 0.00133, and (Kertati, 2021) which analyzed GDI and GEM in Surakarta city that had result was GDI and GEM in Surakarta city showed a position above the average of Central Java GDI. However, previous studies did not use GWR as a modeling approach, so this study aims to model the GDI of Central Java Province in 2021 using the GWR approach. By using GWR approach, it is possible to determine which variables have an impact according to their respective regions.

2. LITERATURE REVIEW

2.1. Multiple Analysis Regression

Multiple linear regression analysis is an extension of simple linear analysis with more than one predictor variable and one response variable. This analysis is used to determine the influence of predictor variables on the response variable. Additionally, multiple linear regression analysis is used for prediction (Mendenhall, 2009). The general form for a multiple linear regression model is as follows.

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_i X_{ij} + \varepsilon_i, \quad i = 1,2,\dots.,n \tag{1}$$

With β_0 as the intercept, β_i as the regression coefficient estimators, and ε_i as the residual assumed to follow a normal distribution with a mean of 0 and a variance of 1, identical, and independent. In this multiple linear regression model, the relationship between the response variable and predictor variables is assumed to be the same at each geographic location (Nurfadilah, 2020). The multiple linear regression model can be written in matrix form as follows.

$$Y = X\beta + \varepsilon \tag{2}$$

$$[y_1\ y_1\ \vdots\ y_n\ ] = [1\ x_{11}\ \dots\ 1\ x_{12}\ \dots\ \vdots\ \vdots\ \ddots\ 1\ x_{n1}\ \dots\ x_{1k}\ x_{2k}\ \vdots\ x_{nk}\ ][\beta_0\ \beta_1\ \vdots\ \beta_k\ ] + [\varepsilon_1\ \varepsilon_1\ \vdots\ \varepsilon_n\ ] \tag{3}$$

According to Mardiatmoko (2020) the assumptions that must be satisfied in multiple linear regression are as follows:

a. The linear regression model is linear in parameters.
b. The residuals of the model are normally distributed.
c. The error variance is constant (homoscedasticity).
d. There is no correlation among the residuals of the observations.
e. There is no correlation among the predictor variables.

2.2. Geographically Weighted Regression

Geographically Weighted Regression (GWR) is an extension of multiple linear regression analysis using the Weighted Least Squares (WLS) method, where the estimated regression coefficients vary according to their geographic locations (Gloria et al., 2022). The estimation of regression coefficients using WLS involves assigning different weights to each observation based on their spatial distances. The closer the distance to the observation, the greater the estimated regression coefficient (Lutfiani et al., 2019). The GWR model can be written as follows.

$$Y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_i(u_i, v_i)X_{ij} + \varepsilon_i, \quad i = 1,2, \ldots, n \tag{4}$$

with Y_i is the value of the response variable at the i-th observation location, $\beta_0(u_i, v_i)$ is the GWR constant, X_{ij} is the value of the k-th predictor variable at the i-th observation location, β_i is the regression coefficient of the k-th variable at the i-th observation location, ε_i is the residual of the k-th observation at the i-th location. The estimation of GWR coefficients for each variable k at observation location i is as follows.

$$\hat{\beta}(u_i, v_i) = \left[X^T W(u_i, v_i) X\right]^{-1} X^T W(u_i, v_i) Y \tag{5}$$

with $\hat{\beta}(u_i, v_i)$ is the estimated coefficient of GWR at location i, X is the predictor variable matrix of size $n\, x(p+1)$, $W(u_i, v_i)$ is the distance weighting matrix at location i of size nxn, Y is the response variable vector of size $nx1$.

2.3 Bandwidth Optimum of GWR

The estimation of GWR coefficients uses weights determined by the spatial coordinates of longitude and latitude. The determination of weights in the GWR model requires a bandwidth or smoothing parameter. One of the methods used in bandwidth selection is Cross Validation or CV (Nursiyono & Apriyani, 2022). The following is the mathematical formula for CV.

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)]^2 \tag{6}$$

with $\hat{y}_{\neq i}$ is estimator of y_i where the observation location (u_i, v_i) is excluded from the estimation process, resulting in the minimum CV value.

2.4 GWR Weighting

The weighting in GWR plays a crucial role in providing parameter estimates at different locations. These weights are expressed in a distance matrix W_{ij} which represents the proximity between location i and location j. The elements of the W_{ij} matrix are weighting functions for each observation location. The function of the weighting matrix is to estimate different parameters at each observation location (Ramadayani et al., 2022). There are four kernel weighting functions:

a) Kernel Fixed Bisquare Function

$$W_j(u_i, v_i) = \{\left(1-\left(\frac{d_{ij}}{b}\right)^2\right)^2, jika\ d_{ij} \leq b\ 0, lainnya \tag{7}$$

with $W_j(u_i, v_i)$ is the spatial weighting matrix for each observation location, d_{ij} is the Euclidean distance and b is the bandwidth that is the same for all observation locations.

b) Kernel Adaptive Bisquare Function

$$W_j(u_i, v_i) = \{\left(1-\left(\frac{d_{ij}}{b}\right)^2\right)^2, jika\ d_{ij} \leq b\ 0, lainnya \tag{8}$$

with $W_j(u_i, v_i)$ is the spatial weighting matrix for each observation location, d_{ij} is the Euclidean distance and b is the bandwidth that is the different for all observation locations.

c) Kernel Fixed Gaussian Function

$$W_j(u_i, v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \tag{9}$$

with $W_j(u_i, v_i)$ is the spatial weighting matrix from kernel fixed Gaussian for each observation location, d_{ij} is the Euclidean distance and b is the bandwidth that is the same for all observation locations.

d) Kernel Adaptive Gaussian Function

$$W_j(u_i, v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \tag{10}$$

with $W_j(u_i, v_i)$ is the spatial weighting matrix from kernel fixed Gaussian for each observation location, d_{ij} is the Euclidean distance and b is the bandwidth that is the different for all observation locations.

2.5 Hypothesis Test of GWR Model

Significance testing of GWR estimates is conducted to assess the influence of predictor variables at each observation location and to determine the model's adequacy. Model adequacy testing is used to identify the best multiple linear regression model or GWR model. The following are the tests conducted for model adequacy.

a) Goodness of Fit Test
Hypothesis for goodness of fit test is the following.
$H_0 : \beta_k(u_i, v_i) = \beta_k, k = 1,2, \ldots, p$
H_1 : minimum have one $\beta_k(u_i, v_i)$ is significant

$$F_{stat} = \frac{\frac{SSE(H_0)}{d_{f1}}}{\frac{SSE(H_1)}{d_{f2}}} \quad (11)$$

The testing criteria for model adequacy are to reject the null hypothesis if $F_{stat} > F_{table}\,(\alpha, d_{f1}, d_{f2})$.

b) Parameter Sig. Test of GWR
Hypothesis for parameter sig. of GWR is the following
$H_0 : \beta_k(u_i, v_i) = 0$
$H_1 : \beta_k(u_i, v_i) \neq 0$

$$t_{stat} = \frac{\hat{\beta}_k(u_i, v_i)}{SE_{\hat{\beta}_k(u_i, v_i)}} \quad (12)$$

The testing criteria for model adequacy are to reject the null hypothesis reject H_0 if $|t_{stat}| > t_{\frac{\alpha}{2}, d_{f2}}$.

2.6 Selection of Best Model

The selection of the best model in this study is based on the values of the coefficient of determination (R^2) and AIC. In multiple linear regression models, the coefficient of determination represents the amount of variation explained by the model. The value of R^2 ranges from 0 to 1. A smaller R^2 indicates a lower ability of the predictor variables to explain the response variable, and vice versa (Lutfiani et al., 2019). The coefficient of determination (R^2) in the GWR model can be expressed as follows

$$R_i^2 = \frac{SST_{GWR} - SSE_{GWR}}{SST_{GWR}} \quad (13)$$

with SST is sum square of total GWR, SSE is sum square of error GWR.

The second criterion for selecting the best model is using the AIC value. The determination of the AIC value can be done using the following formula.

$$AIC = 2n\, ln\, ln\,(\hat{\sigma}) + n\, ln\, ln\,(2\pi) + n + tr(S) \quad (14)$$

3. METHODOLOGY OF RESEACH

3.1. Data and Sources

The data used in this study are secondary data from 35 districts/cities in Central Java Province in 2021. The secondary data for this study were obtained from the official website of the Central Bureau of Statistics. The variables used in this study are the Gender Development Index (Y) as the dependent variable, Average Length of Schooling (X_1), T Open Unemployment Rate for Females (X_2), Labor Force Participation Rate (X_3), School Participation Rate (X_4), and Sex Ratio(X_5) as the independent variables.

3.2. Methods

This research used Geographically Weighted Regression (GWR) approach for modelling GDI in Central Java Province in 2021. The GWR approach will model the regression for each District/City based on weighting factors. This research used kernel bisquare for weighting factors.

3.3. Data Analysis

The steps of data analysis using GWR to model the Gender Development Index in Central Java Province in 2021 are as follows:

1. Perform descriptive analysis to examine the initial distribution of Gender Development Index in Central Java Province in 2021
2. Conduct multiple linier regression analysis (OLS) with following steps
   a. Build the OLS model
   b. Test the significance of parameters
   c. Test of classical assumption of OLS model
3. Perform GWR analysis with the following steps
   a. Calculate the Euclidean distance between location point i and location point j
   b. Determinate the optimal bandwidth using CV method
   c. Calculate weigthed matrix using kernel function. In this study, the Gaussian kernel function is used.
   d. Build the GWR model
   e. Obtain parameter estimates of the GWR model using WLS
   f. Goodness of fit test
   g. Test the significance of parameters in GWR models
4. Determine the best model on the coefficient of determination $\left(R^2\right)$ and AIC values for the GWR model using Gaussian kernel weighting.
5. Draw conclusion

4. RESULT AND DISCUSSION

4.1 Descriptive Analysis

Central Java Province consists of 29 regencies and 6 cities, making a total of 35 regencies/cities administratively. Based on Figure 1, the distribution of the Gender Development Index (GDI) in Central Java Province in 2021 can be observed. The GDI in Central Java can be divided into three intervals based on the GDI score of 92. GDI values between 96 to 92 are classified as high, GDI values between 91 to 83 are classifies as moderate, and GDI values below 82 are classified as low. Regencies/cities with high GDI values include Wonogiri District, Klaten District, Sukoharjo District, Karanganyar District, Blora District, Pati

District, Temanggung District, Wonosobo District, Kebumen District, Cilacap District, Tegal District, and Pekalongan District. Regencies/cities with moderate GDI values include Banyumas District, Brebes District, Pemalang District, Batang District, Kudus District, Grobogan District, Magelang District, Magelang City, Surakarta City, and Tegal City. Regencies/cities with low GDI values include Banyumas District, Banjarnegara District, Semarang District, Sragen District, Rembang District, Salatiga City, Pekalongan City, and Semarang City.

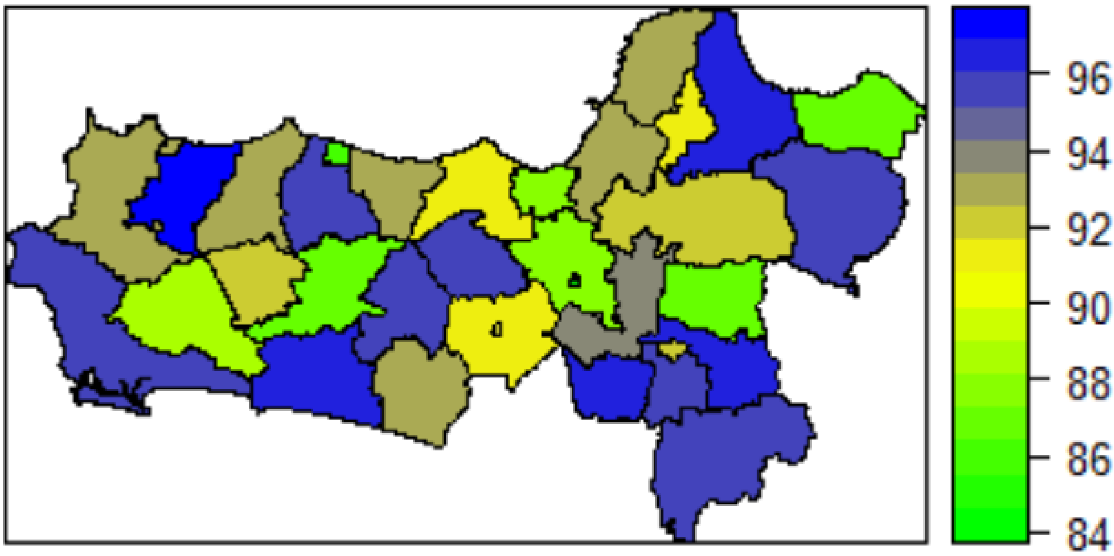

Figure 1. Mapping of Gender Development Index in Central Java Province in 2021

4.2 Multiple Linier Regression

The initial step in analyzing GWR is to first perform multiple linear regression analysis. The results of the multiple linear regression analysis are presented in Table 1.

Table 1. Result of Multiple Linier Regression

	Estimate	Std.Error	T statistics	p-value
Intercept	-48.27656	48.01692	-1.005	0.323017
X_1	2.63099	0.71271	3.692	0.000918*
X_2	-0.45799	0.42852	-1.069	0.293986
X_3	0.22250	0.18147	1.226	0.230015
X_4	0.08249	0.07665	1.076	0.290708
X_5	0.99672	0.07665	1.076	0.032348*

*) sig. in $\alpha = 10\%$

Based on Table 1, the multiple linear regression model can be obtained as follows.

$$\hat{y} = -48.277 + 2.6309X_1 + -04578X_2 + 0.222X_3 + 0.08249X_4 + 0.9967X_5 + \varepsilon \qquad (15)$$

Significance testing of the predictors in the multiple linear regression model can be performed by comparing the p-values of each predictor variable with a significance level (alpha) of 10%. Table 1 shows the p-values of each predictor variable. When compared to the alpha value of

10%, the predictor variables X_1 and X_4 have p-values less than 10%, indicating that these predictor variables have a significant impact on the response variable Y. Therefore, the multiple linear regression model with 2 significant predictor variables is as follows.

$$\hat{y} = -48.277 + 2.6309X_1 + 0.9967X_4 + \varepsilon \quad (16)$$

4.3 Classic Assumption Test

4.3.1 Normality Test

The normality assumption is used to test the normal distribution of the model residuals. The normality assumption in this study is tested using the Shapiro-Wilk test. The results of the Shapiro-Wilk test are shown in Table 2.

Table 2. Result of Normality Test

W	p-value
0.98066	0.7806

Table 2 shows a p-value of 0.7806. When compared to the significance level of 10%, this p-value is greater, indicating that the residuals of the multiple linear regression model are normally distributed.

4.3.2 Multicolinearity Test

Multicollinearity test is used to determine the presence of correlation among predictor variables. The criterion for multicollinearity is when the VIF value is greater than 10. The results of the multicollinearity test are shown in Table 3.

Table 3. Result of Multicolinearity Test

Variable	VIF
X_1	4.712005
X_2	3.148397
X_3	2.222513
X_4	2.370584
X_5	3.440702

Based on Table 3, it can be seen that the VIF values for all predictor variables are less than 10, indicating that there is no multicollinearity.

4.3.3 Spatial Heterogeneity Test

The spatial heterogeneity test is used to determine the heterogeneity/homogeneity of residuals. The spatial heterogeneity test uses the Breusch Pagan test. The results of the Breusch Pagan test are shown in Table 4. The result of the Breusch Pagan test in Table 4 above shows a p-value of 0.06246. When compared to a significance level of 10%, the p-value of the Breusch Pagan test is smaller. Therefore, it can be concluded that there is spatial heterogeneity in the residuals. Consequently, the appropriate analysis in this case is GWR

Table 4. Result of Spatial Heterogeneity Test

Breusch-Pagan Test	Df	p-value
10.491	5	0.06246

4.3.4 Spatial Autocorrelation Test

Spatial autocorrelation test is used to determine the correlation between residuals at each observation location. If spatial autocorrelation occurs, it can be concluded that observation locations will influence each other. The test for spatial autocorrelation can be conducted using Moran's test(Putri & Hajarisman, 2021). The results of the Moran's test are shown in Table 5.

Table 5. Result of Spatial Autocorrelation Test

Index Moran's Test	p-value
0.02199520	-0.02941176

Based on Table 5 above, it can be observed that the p-value for the Moran's Index is -0.0291. When compared to the significance level of 10%, this value is smaller. Therefore, it can be concluded that there is spatial autocorrelation present.

4.4 GWR Analysis

After conducting multiple linear regression analysis, the next step is to analyze the GWR model. The first step in GWR analysis is to determine the Euclidean distance between observation locations. Then, the optimum bandwidth is calculated using CV. Based on the calculation results, the optimum bandwidth for the fixed Gaussian weighting function is 2.838406. Thus, the resulting GWR model estimation is as follows.

Table 6. Estimate of GWR Model

Parameter	Min	Median	Max
Intercept	-86.041383	-57.011467	-9.0769
X_1	2.508417	2.634104	2.7545
X_2	-0.527087	-0.438345	-0.4170
X_3	0.217841	0.225546	0.2308
X_4	0.013025	0.09467	0.1584
X_5	0.668488	1.072398	1.3154

Based on Table 6, it can be seen that the estimated value of variable X_1 has a positive estimate with an influence ranging from 2.508417 to 2.7545. Variable X_2 has a negative estimate with an influence ranging from -0.527087 to -0.4170. Variable X_3 has a positive estimate with an influence ranging from 0.217841 to 0.2308. Variable X_4 has a positive estimate with an influence ranging from 0.013025 to 0.1584. Variable X_5 has a positive estimate with an influence ranging from 0.668488 to 1.3154. Based on the partial significance testing (t-test), it is found that variables X_1, X_4, and X_5 have a significant influence on the variable Y. Therefore, for the GWR model in Cilacap District, the following can be written. Table 7 shows the significant predictor variables for each district/city.

$$\hat{y}_{Cilacap} = -17.9190 + 2.549818X_1 + 0.7462145X_5 + \varepsilon \quad (17)$$

If model above is interpreted, it can be understood for variable X_1, a one unit increase in variable X_1 will result an increase in Y by 2.549818. Simirarly, for variable X_5, a one unit increase in variable X_5 will lead to an increase in Y by 0.7462145.

Table 7. Signifacant Variable of Each District/City

Sig. Variable	District/City
$X_1, X_4, X_5,$	Purworejo District, Wonosobo District, Magelang District, Boyolali District, Klaten District, Wonogiri District, Sragen District, Grobogan District, Blora District, Rembang District, Pati District, Semarang District, Temanggung District
X_1 and X_5	Cilacap District, Banyumas District, Purbalingga District, Banjarnegara District, Kebumen District, Sukoharjo District, Karanganyar District, Jepara District, Demak District, Kendal District, Batang District, Pekalongan District, Pemalang District, Tegal District, Brebes District, Magelang City, Surakarta City, Salatiga City, Semarang City, Pekalongan City, Tegal City.

The model's goodness of fit is tested using the F-test. The detailed results of the F-test can be seen in Table 8.

Table 8. Result of Goodness of Fit

	Df	SS	MS	F Stat	F Table
GWR Improv	26.74872	8.251278	0.30847375		
GWR Res	8.25128	156.5899	18.9776495	61.5211163	2.396054

Based on Table 8, it can be seen that the Fstat (61.521) is greater than the critical F-value. Therefore, it can be concluded that the GWR model is appropriate.

4.5 Selection of Best Model

The criteria for selecting the best model between multiple linear regression and GWR are compared based on the values of the coefficient of determination(R^2) and and the AIC in each model. The values of the coefficient of determination (R^2) and the AIC are shown in Table 9.

Table 9. Result of R^2 and AIC

Model	R^2	AIC
OLS	0.5325	168.8745
GWR	0.5181791	159.0621

Based on Table 9, it can be seen that the multiple linear regression model has a higher coefficient of determination (R^2) compared to the GWR model. However, the multiple linear regression model has a higher AIC value compared to the GWR model. Therefore, it can be concluded that the GWR model is better for modeling the Gender Development Index in Central Java Province in 2021 because it has a lower AIC value.

5. CONCLUSION

Based on the GWR modeling analysis above, it can be concluded that the GWR model is the best model for modeling the GDI in Central Java Province in 2021. This model had R^2 value by 51.82% and AIC value of 159.0621. In the formed GWR model, the significant variables are X_1 or average length of schooling, X_4 or school participation rate, and X_5 or gender ratio.

REFERENCES

Amaliah, K. N., & Riniwati, H. (2021). Pemodelan Dinamika Sistem Indeks Pembangunan Manusia Dan Indeks Pembangunan Gender Di Wilayah Pesisir Kabupaten Sumenep. *Prosiding SNasPPM*, *6*(1), 145–150. https://snasppm.unirow.ac.id/prosiding/index.php/SNasPPM/article/view/645

Badan Pusat Statistika. 2023. Jakarta.

Aprilianti, S., & Setiadi, Y. (2022). Faktor-faktor Yang Memengaruhi Indeks Pembangunan Gender di Indonesia Tahun 2020. *Seminar Nasional Official Statistics*, *2022*(1), 245–254. https://doi.org/10.34123/semnasoffstat.v2022i1.1351

Elisa, I. (2022). Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Gender (IPG) Provinsi Sumatera Barat Menggunakan Analisis Regresi Data Panel. *Journal of Mathematics UNP*, *7*(2), 8. https://doi.org/10.24036/unpjomath.v7i2.12666

Franklin, C. A., Klingenberg, B., & Agresti, A. (2017). *Statistics: The Art and Science of Learning from Data, Global Edition*.

Gloria, M., Bele, L., Mustikawati, E., Hermanto, P., & Fitriani, F. (2022). *Pemodelan Geographically Weighted Regression pada Kasus Stunting di Provinsi Nusa Tenggara Timur Tahun 2020*. *6*(2), 179–191.

Insiro, A. R., Handajani, S. S., Subanti, S., Studi, P., & Fmipa, S. (2023). *Pemodelan Indeks Pembangunan Gender Provinsi Jawa Barat Menggunakan Regresi Nonparametrik Penalized Spline*. *8*(2721).

Kertati, I. (2021). Analisis Indeks Pembangunan Gender (IPG) Dan Indeks Pemberdayaan Gender (Idg) City Surakarta. *Public Service and Governance Journal*, *2*(01), 1. https://doi.org/10.56444/psgj.v2i01.1960

Lutfiani, N., Sugiman, & Mariani, S. (2019). Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. *UNNES Journal of Mathematics*, *5*(1), 82–91. http://journal.unnes.ac.id/sju/index.php/ujmUJM8

Mardiatmoko, G. (2020). Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Berganda. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *14*(3), 333–342. https://doi.org/10.30598/barekengvol14iss3pp333-342

Mendenhall, W. R. J. B. B. M. B. (2009). *Introduction to Probability and Statistics* (3rd ed.). Brooks/Cole.

Nurfadilah, K. (2020). Geographically Weighted Regression (GWR) pada Data Jumlah Penderita Penyakit AIDS. *Jurnal Matematika Dan Satatistika Serta Aplikasinya*, *8*(1).

Nursiyono, J. A., & Apriyani, M. (2022). Determinan Kematian Ibu di Jawa Timur Tahun 2020: Analisis Geographically Weighted Regression (GWR). *Poltekita : Jurnal Ilmu Kesehatan*, *16*(1), 89–97. https://doi.org/10.33860/jik.v16i1.844

Putri, S. A. H., & Hajarisman, N. (2021). Pemodelan data kemiskinan di provinsi Jawa Barat menggunakan metode geographically weighted regression (GWR) dengan fungsi pembobot kernel bi-square. *Prosiding Statistika*, *7*(1), 208–215.

Rachmawati, F., Sigalingging, N. Y., Kiftiah, M., Tanjungpura, U., Prof, J., Profesor, J., Nawawi, D. H. H., Pontianak, K., & Penulis, ; *. (2022). *Proyeksi Indeks Pembangunan*

Gender (IPG) di Kalimantan Barat dengan Metode Trend Parabolik Projection of the Gender Development Index (GPI) in West Kalimantan using the Parabolic Trend Method. *2*(2), 83–91. http://dx.doi.org/10.xxxxx/formasi.2021.1.1.1-12

Ramadayani, M. R., Fariani Hermin Indiyah, & Ibnu Hadi. (2022). Pemodelan Geographically Weighted Regression Menggunakan Pembobot Kernel Fixed dan Adaptive pada Kasus Tingkat Pengangguran Terbuka di Indonesia. *JMT : Jurnal Matematika Dan Terapan*, *4*(1), 51–62. https://doi.org/10.21009/jmt.4.1.5

Safitri, L.D.A. (2019). Pemodelan Indeks Pembangunan Manusia Dan Indeks Pembangunan Gender Di Indonesia Dengan Pendekatan Regresi Probit Biner Bivariat. *Jurnal Matematika*, *Statistika Dan Komputasi*, *16*(2), 150. https://doi.org/10.20956/jmsk.v16i2.7436

Safitri L. P. S., & Ayuningsih, N. P. M. (2023). Pemodelan Mixed Geographically Weighted Regression (MGWR) pada Kasus Penderita Diare di Provinsi Bali. *Saintifik*, *9*(1), 18–27. https://doi.org/10.31605/saintifik.v9i1.384

Sari, S. U. R. (2018). Aplikasi Metode Regresi Nonparametrik Spline Multivariabel Untuk Pemodelan Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Gender di Provinsi Jawa Barat. *Jurnal Statistika Universitas Muhammadiyah, 6*(2), 119–129. http://103.97.100.145/index.php/statistik/article/view/4316