Impacts of Human Development Index and Percentage of Total Population on Poverty using OLS and GWR models in Central Java, Indonesia

Duhania Oktasya Mahara[1], Achmad Fauzan[1],*

[1] Statistics Department, Faculty of Mathematics and Natural Sciences Universitas Islam Indonesia, Jl. Kaliurang KM 14.5, Sleman-Yogyakarta, Indonesia

* *Corresponding author: achmadfauzan@uii.ac.id.*

Abstract: Central Java province is one of the provinces with the highest number of poor people on the island of Java, with the number of poor people in 2020 increasing by 0.44 million people from the previous year. Poverty is caused by several factors, one of which is the Human Development Index (HDI) and the Total Population level. Each region has different characteristics from other regions. These differences in characteristics cause more specific spatial effects, namely spatial heterogeneity. Geographically Weighted Regression (GWR) is a statistical method that can analyze spatial heterogeneity by assigning different weights and models to each observation location. This study aims to determine whether the HDI variable and percentage of total population significantly impact the number of poor people in Central Java Province in 2020 without eliminating the spatial effect. There are three groupings of variables that affect the Number of Poor People for GWR with the Adaptive Kernel Bisquare weighting function and four groups for the Adaptive Kernel Tricube weighting function. The Key Performance Indicators (KPI) used are Mean R^2, Akaike Information Criterion (AIC), Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE). Based on these KPIs, the GWR model with the Adaptive Kernel Bisquare weighting function provides better results when compared to the OLS model.

Keywords: GWR, OLS, Poverty, Spatial effects.

Introduction

Poverty is one of the fundamental economic problems for a country that is the center of government attention, which is a very complex problem that must be handled appropriately. In 2020 the number of poor people will increase by 2.76 million [1]. There were several things that become poverty criteria, one of them was number of poor populations. Central Java Province is one of the provinces with the highest number of poor people on the island of Java. It is known that the number of poor people in Central Java Province in 2020 has increased by 0.44 million people from the previous year [2]. Poverty is caused by several factors, one of which is the Human Development Index (HDI) level and the population in an area.

Each region has characteristics that are different from other regions. These differences in characteristics cause spatial effects. The use of spatial effects in research will cause problems in the form of spatial dependence (spatial autocorrelation) and the emergence of spatial heterogeneity [3], [4].

Spatial heterogeneity occurs because of random location effects, or it can be said that the same response variable gives a different response from one location to another [5]. This trait arises due to differences in characters or traits which are the special characteristics of each location. The existence of spatial heterogeneity will have an impact on the parameter estimator in the model that is not Best Linear Unbiased Estimator (BLUE). Moreover, the existence of heterogeneity can result in the conclusions of the model being misleading; this is due to the tendency of the variance to increase which results in an enlarged standard error [6].

One of the statistical methods that can be used to analyze spatial heterogeneity is Geographically Weighted Regression (GWR). GWR is different from the linear regression (OLS) method. If in linear regression using parameter estimation using the least-squares method [7], GWR uses the weighted least

squares method in parameter estimation [8]. Several previous studies related to previous methods or cases, such as Andrietya et al. [9], analyze poverty determinants in Central Java's province using linear regression. Haryanto & Andriani [10] analyzed GWR with a fixed kernel exponential weighting function to predict the number of poor people in Central Java in 2018 with predictors of the district/city minimum wage, Human Development Index (HDI), and the Open Unemployment Rate. Fotheringham et al. [11] evaluate the impacts of air pollution in China using a latterly proposed model—Multi-scale Geographically Weighted Regression (MGWR). Koh et al. [12] learn the relationships between the NO3–N concentration and various parameters (topography, hydrology, and land use) across the island in Jeju Island, South Korea, using OLS regression and GWR. Zhao et al. [13] studied the use of principal component analysis (PCA) and the GWR model to predict the spatial distribution of frozen ground temperature. Ma et al. [14] discuss about Bayesian resource of GWR. Muche et al. [15] using GWR analysis to cluster under-nutrition and its predictors among under-five children in Ethiopia: Evidence from demographic and health survey.

Based on this, this study aims to measure the impact of HDI and Percentage of Population on Poverty in the province of Central Java while still paying attention to the spatial effects it has, and this is because each region has unique differences in nature and character.

Materials and Methods
Data

The data used in this study is secondary data sourced from the Central Java Central Bureau of Statistics (BPS) website, which can be accessed at https://jateng.bps.go.id/. The dependent variable used is the Number of Poor Populations (NPP), and the independent variables consist of the human development index (HDI) and the percentage of Total Population (PP) in Central Java Province in 2020. More explanation is represented in Table 1 below.

Table 1. Description of the data

Variable	Explanation	Denomination	Meaning
Number of Poor Populations (NPP)	Y	Thousand People	Population with an average of expenditure per capita below the minimum limit of Rupiah spent on food and non-food
Human Development Index (HDI)	X_1	Percent	Percentage of population's ability to access development outcomes in obtaining good income, health, and education.
Percentage of Total Population (PP)	X_2	Percent	Number of populations in an area

Linear Regression (OLS) and Geographically Weighted Regression (GWR)

Linear regression is an analytical method that can be used to see a linear relationship between two variables where one variable is considered the affected variable while the other variable is regarded as the influencing variable; the regression parameter or coefficient determines this. The variables that influence are called independent variables or predictor variables, and the affected variables are called dependent variables or response variables [16]. One of the most used linear regression analysis approaches is Ordinary Least Square (OLS). OLS is a method used to estimate the regression coefficient by minimizing the number of squares of errors [17].

While the GWR method is a development of the regression method, this method produces a model that has its parameter equation for each observation location [18]. The occurrence of spatial heterogeneity causes the different model equations from each location due to variance from one observation to another observation [8]. Spatial heterogeneity is a condition if one of the independent variables gives a different response to several different locations in an observation [5]. Due to the different reactions to several places, different regression parameter results appear at each observation location [19] This test can be carried out using the Breusch-Pagan method with the hypothesis H_0 that there is no spatial heterogeneity. The number of parameter equations obtained is as many as the observation locations used in the study [20], in which the model that has been obtained at one observation location cannot be used to estimate parameters at other observation locations [21].

The method used to estimate the parameters for the GWR model is the Weighted Least Square (WLS) method or the Weighted Least Square Method, using weights [22]. WLS is known to be able to neutralize the consequences of violating the heterogeneity assumption. The GWR method is one of the local analytical methods, while linear regression is a global analysis method. In addition, GWR also pays attention to spatial or location effects. The GWR model is presented in Equation 1.

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)X_{ik} + \varepsilon_i \tag{1}$$

Y_i is the value of the dependent variable at the ith observation location, X_{ik} is the kth independent variable value at the ith observation location, (u_i, v_i) is the coordinates of the ith observation location, $\beta_0(u_i, v_i)$ is the GWR Intercept, $\beta_k(u_i, v_i)$ is the kth regression coefficient at the ith observation location, and ε_i is the error value at the ith observation location.

Neighborhood relationships that show proximity between locations are expressed in a spatial weighting matrix [23]. The weights in the GWR model play a significant role because the weighting values represent the location of one observation with another. The spatial weighting of the GWR model is in the form of a diagonal matrix in which the elements are a weighting function of each observation location, based on a point location approach. It is known that there are several methods used to determine the amount of weight for each different place in the GWR model, one of which is the kernel function.

The kernel function or K(u) is a continuous, symmetrical, finite and function $\int_{\infty}^{\infty} K(u)\,du = 1$. The kernel function is used to estimate the parameters in the GWR model if the distance function is continuous and monotonically decreasing. The weighting of the GWR kernel model can be formed using the distance function presented in Equation 2.

$$w_{ij} = \begin{cases} 1 & \text{, when } d_{ij} < h \\ 0 & \text{, when } d_{ij} \geq h \end{cases} \tag{2}$$

d_{ij} is the Euclidean distance between the ith observation point and the jth observation point [22], and h is the bandwidth. The value of d_{ij} is obtained by Equation 3.

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{3}$$

Bandwidth (h) measures the distance of the weighting function, which states to what extent the influence of the location on other locations [24]. The weighting matrix W (u_i, v_i) can be determined using a kernel function. The kernel function is known to give weighting according to the optimum bandwidth. Optimum bandwidth selection in the GWR method is essential because it will affect the model's data accuracy. There are two types of kernel functions, namely Adaptive and Fixed. The Adaptive kernel function obtains a different bandwidth value at each observation location by adjusting the observation location. In contrast, the Fixed kernel function will receive only one bandwidth value, which means the bandwidth value for all observation locations.

Some of the adaptive kernel functions are (1) Adaptive Kernel Bisquare, (2) Adaptive Kernel Gaussian, and (3) Adaptive Kernel Tricube. In comparison, the types of Fixed kernel functions include (1) Fixed Kernel Bisquare, (2) Fixed Kernel Gaussian, and (3) Fixed Kernel Tricube [25]. Estimation of the parameter (u_i, v_i) in the GWR model uses the Weighted Least Squares (WLS) method, namely by giving different spatial weights for each observation location [5]. The amount of spatial weighting is obtained based on the distance between the observation locations. The closer the distance between the observation locations, the greater the spatial weighting.

After obtaining the estimated GWR model parameters, the GWR hypothesis testing is carried out, namely testing the model's suitability, and testing the model parameters. The goodness of fit was conducted to determine the significance of the geographical factors, which are the core of the GWR model with H_0 defined there is no significant difference between the classical regression model and the GWR model. The parameter testing of the GWR model is carried out to partially test the parameters. This test is carried out to determine which parameters have a significant effect on the dependent variable at each observation location. With H_0, there is no effect of the independent variable on the dependent variable. This test is carried out by looking at the t_{cal} value of each parameter. Significance level (α) is the threshold that used to determine the significance, if the p$_{value}$ is less than or equal to significance level the data is considered statistically significant. The significant level used is 95% or $\alpha = 0.05$.

Key Performance Indicator

The Key Performance Indicator (KPI) used is based on the value of R^2, Akaike Information Criterion/AIC [26], Mean Absolute Error/MAE [27], Mean Square Error/MSE [28], and Mean Absolute Percentage Error/ MAPE [29]. This KPI used as comparison criteria to choose the best models, which are presented sequentially in Equation 4-7.

$$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100\% \tag{7}$$

k is the number of parameters estimated in the regression model, n: number of observations, e: 2,718, and u: residual. Y_i is the actual value of the dependent variable at location i, while Y_i is the predicted value of the dependent variable at location i.

The model is better if it has a more significant coefficient of determination or R^2 because it means that the existing factor variables can explain the model more. AIC is one method that can be used to select the best regression model [30], and it is said to be the best regression model if it has the smallest AIC value [31]. Likewise, the smaller MAPE, MAE, and MSE indicators show that the model is getting better.

Methods

This study was started by implementing linear regression and then compared with the GWR method. The first step is data collection on the variable Number of Poor Population (NPP), Human Development Index (HDI), and Percentage of Total Population (PP) in Central Java Province. Then arrange a descriptive analysis to recognize the general description of the dependent and independent variables by making a table or graph. Spatial visualization was used using the Quantum Geographic Information System (QGIS). Spatial visualization is used to invent it more accessible to draw the initial image spatially. QGIS is used because it is a free and powerful GIS software whose functions are almost the same as paid GIS software [32]. Meanwhile, R software is used for inferential analysis because of its convenience, power, and other resources [33].

The second step is linear regression analysis to recognize whether the independent variables/factors significantly influence the dependent variable and continue with the assumption test. Then arrange a test of spatial heterogeneity because the GWR method is used to analyze data that has spatial heterogeneity. If there is spatial heterogeneity, the next step is to determine the weighting with the kernel function to find the optimum bandwidth using the cross-validation method.

The third step is to resemble parameter estimates for the GWR model and continue to carry out the goodness of fit test or model suitability test. This is done to see whether the GWR model has a significant difference from the linear regression model. Suppose the GWR model with the predetermined weighting function does not significantly differ from the classical regression model. In that case, the GWR model's bandwidth search and parameter estimation with other weighting functions are carried out again until the best weighting function is found.

The fourth step is to find the estimated parameters of the GWR model, find the t-count of the parameter estimates, and determine the best model for each observation location by looking at the t-count. The last step is to calculate the AIC, R^2, MAPE, MSE, and SSE values from the linear regression and the GWR model to compare the best model for analyzing case studies and concluding.

Results and Discussions
Descriptive Statistics

Based on data, in 2020, the average number of Number of Poor Population (NPP) in Central Java is known to be worth 113.74 thousand people, while the HDI is on average 72.51%, and the PP is on average worth 2.86%. Visualizations of NPP, HDI, and PP are presented in Figures 1(a), 1(b), and 1(c).
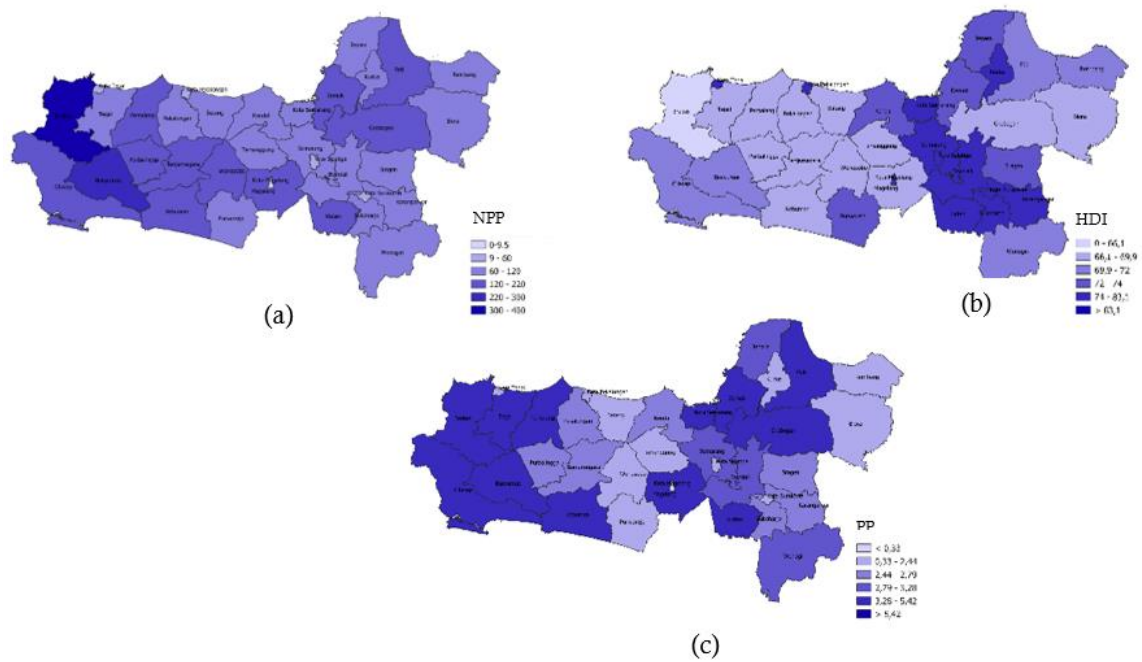
Figure 1. NPP Thematic Map and the factors that influence NPP.

Figure 1 is a thematic map of NPP and the factors that influence it. The darker the color in the area, the higher the value for the variable. As an illustration, the area with the highest NPP value in Central Java Province is Brebes regency, with a NPP worth 308.8 thousand people, while the lowest NPP is Magelang city which is 9.3 thousand people. Furthermore, the HDI value shows that the area with the lowest HDI value is Brebes Regency with an HDI value of 66.11%, and the highest is Salatiga City, which is 83.14%. While the PP value based on Figure 1(c) tells that the City of Salatiga is the area with the lowest PP, it is known to be worth 0.33%, and the area with the highest PP value is Brebes Regency which is worth 5.42%.

Analysis of Linear Regression and GWR

Linear regression analysis aims to determine the effect of the independent variables/ factors on the Number of Poor Population variables without considering the spatial impact. With $\alpha = 0.05$, the results of the linear regression are presented in Table 2.

Table 2. Results of linear regression analysis.

Model	Value	Significance level	Decision
F_{cal}	62.57	1.79	H_0 rejected
p_{val}	$8.74e^{-12}$	0.05	H_0 rejected
R^2	79.64%	-	-
Adj.R^2	78.36%		

Based on Table 1, the value of the F test is obtained by $F_{cal} = 62.57 > 1.78 = F_{table}$ so that it can be explained that the model is significant. Then, the value of R^2 received a weight of 79.64%, which means that the independent variable can define the variance of the dependent variable of 79.64%. Other variables outside the model describe the remaining 20.36%. Then continue with the partial test or t-test to partially test the parameter estimation, or in other words, to find out whether the independent variable/predictor (X) has a significant effect on the response variable (Y). The value of t_{cal} from the two independent variables is more than t_{table}. In other words, it can be said that the two independent variables used are significant or affect the dependent variable. The result of t-test shows in table 3 below.

Table 3. Results of linear regression analysis (t-test).

Variabel	t_{cal}	P-value	Decision	Conclusion
HDI (X_1)	-3.67	$8.62e^{-4}$	H_0 Rejected	Significant
PP (X_2)	7.84	$6.03e^{-09}$	H_0 Rejected	Significant

Meanwhile, the results of the assumption test for the linear regression model, including the normality test, autocorrelation test, and homoscedasticity test, are presented in Table 4 [34].

Table 4. Linear regression assumption test results.

Assumption test	p_{value}	Decision	Conclusion
Normality	0.98	Failed to reject H_0	Residuals are normally distributed
Autocorrelation	0.65	Failed to reject H_0	There is no autocorrelation
Homoscedasticity	0.01	H_0 rejected	There is heteroscedasticity

Based on Table 4, the data has a normal distribution of residuals, and no autocorrelation occurs. However, the assumption of homoscedasticity is not fulfilled, or heteroscedasticity befalls. While the results of the multicollinearity test showed that the two variables did not occur multicollinearity. Based on the Breusch Pagan test, it was obtained that the $p_{value} = 001889 < \alpha = 0.05$. From this, it can be concluded that there is heterogeneity so that the analysis can be continued using the GWR method.

The first step in the GWR method is the selection of a weighting function. The spatial weighting for the GWR model in this study was determined by performing a goodness of fit test on all types of bandwidth and kernel functions. As explained earlier, the bandwidth used is adaptive and fixed. While the kernel functions used are Bisquare, Gaussian, and Tricube. The weighting function used is a weighting function whose results show that there is a difference between linear regression and GWR as evidenced if $p_{value} < \alpha$. The goodness of fit test results from all weighting functions are presented in Table 5.

Table 5. The results of the linear regression assumption test.

Weight Function	p_{value}	Decision
Adaptive Kernel Bisquare	0.04	H_0 rejected
Fixed Kernel Bisquare	0.40	Failed to reject H_0
Adaptive Kernel Gaussian	0.21	Failed to reject H_0
Fixed Kernel Gaussian	0.38	Failed to reject H_0
Adaptive Kernel Tricube	0.00	H_0 rejected
Fixed Kernel Tricube	0.43	Failed to reject H_0

Based on Table 5, two weighting functions that result in GWR modeling have significant differences between linear regression and GWR, namely Adaptive Kernel Bisquare and Adaptive Kernel Tricube. Next, GWR modeling is carried out using the two weighting functions. The bandwidth value of each weight is presented in Table 6.

Table 6. The bandwidth of each observation location

Number	District (D)/ Municipality (M)	Bandwidth with Adaptive Kernel Bisquare weighting function	Bandwidth with Adaptive Kernel Tricube weighting function
1	Cilacap (D)	0.69	0.69
2	Banyumas (D)	0.54	0.54
3	Purbalingga (D)	0.54	0.54
4	Banjarnegara (D)	0.82	0.82
5	Kebumen (D)	0.48	0.48
6	Purworejo (D)	0.61	0.61
7	Wonosobo (D)	0.54	0.54
8	Magelang (D)	0.65	0.65
9	Boyolali (D)	0.65	0.64
10	Klaten (D)	0.62	0.62
11	Sukoharjo (D)	0.54	0.54
12	Wonogiri (D)	0.73	0.73
13	Karanganyar (D)	0.89	0.89
14	Sragen (D)	0.58	0.58
15	Grobogan (D)	0.57	0.57
16	Blora (D)	0.48	0.48
17	Rembang (D)	0.77	0.77
18	Pati (D)	0.78	0.78

Number	District (D)/ Municipality (M)	Bandwidth with Adaptive Kernel Bisquare weighting function	Bandwidth with Adaptive Kernel Tricube weighting function
19	Kudus (D)	0.81	0.81
20	Jepara (D)	0.93	0.93
21	Demak (D)	0.60	0.59
22	Semarang (D)	0.86	0.86
23	Temanggung (D)	0.54	0.53
24	Kendal (D)	0.75	0.75
25	Batang (D)	0.93	0.93
26	Pekalongan (D)	0.66	0.66
27	Pemalang (D)	0.61	0.61
28	Tegal (D)	0.57	0.57
29	Brebes (D)	0.47	0.47
30	Magelang (M)	0.66	0.61
31	Surakarta (M)	0.57	0.57
33	Salatiga (M)	0.66	0.66
33	Semarang (M)	0.58	0.58
34	Pekalongan (M)	1.09	1.09
35	Tegal (M)	0.96	0.96

Because the parameter values in each location are different, the parameter estimates in the GWR model are presented in Table 7.

Table 7. Summary of Estimation Parameter each observation location

Weighting function	Variable	Parameter Coefficient					Global
		Min	1st Quartile	Median	3rd Quartile	Max	
Adaptive Kernel Bisquare	Intercept	-264.91	139.69	287.84	534.94	649.60	357.78
	HDI	-9.38	-6.68	-4.01	-1.66	3.78	-4.79
	PP	7.22	27.45	40.36	46.08	61.09	36.26
Adaptive Kernel Tricube	Intercept	-300.19	149.08	317.33	541.87	649.30	357.78
	HDI	-9.39	-6.86	-4.06	-1.78	4.32	-4.79
	PP	4.86	27.19	38.55	46.92	61.96	36.26

Furthermore, to find the weighting function with the Adaptive Kernel Bisquare, the bandwidth and the Euclidean distance from each location are substituted into the Adaptive Kernel Bisquare equation. Based on Table 5, it is shown that the parameter estimation HDI variable has a minimum value of -9.38 and a maximum value of 3.78. The HDI variable can affect NPP in Central Java Province with an estimated value range of -9.38 to 3.78. Likewise, the 1st quartile and 3rd quartile values, the HDI predictor variable has a 1st quartile value of the estimated GWR model parameter, which is -6.68, and the 3rd quartile value is -1.67, with the median or estimated value of the HDI variable parameter is -4.01. The Global meaning in Table 5 above shows the estimated global parameters or parameter estimates from the linear regression model, worth -4.79 for the HDI variable. Besides other predictor variables.

From the goodness of fit test results, the value $F_{cal} = 4.22 > F_{table} = 2.35$, the decision to reject H_0 is obtained, which means that there is a significant difference between the linear regression and the GWR model. In other words, GWR model with the Adaptive Kernel Bisquare function can be used. After obtaining parameter estimates to perform GWR modeling with the Adaptive Kernel Bisquare weighting function, the parameter estimation results are obtained for each district/ municipality in Central Java Province. A partial test on each variable by studying at the t-count value obtained from each variable in each location is carried out to obtain the best model for each district/ municipality, by looking at the t-count value obtained from each variable in each location. Table 8 below shows the estimation parameter obtained of every location and the t_{cal} value.

Table 8. Estimation Parameter and Statistics Test Value from every location

No	District (D)/ Municipality (M)	Adaptive Kernel Bisquare					Adaptive Kernel Tricube				
		Estimation Parameter			t_{cal}		Estimation Parameter			t_{cal}	
		Intercept	X_1	X_2	X_1	X_2	Intercept	X_1	X_2	X_1	X_2
1	Cilacap (D)	-221.42	2.87	41.7	1.11	4.6	-231.24	3	41.82	1.13	4.52
2	Banyumas (D)	-264.91	3.78	41.92	1.51	5.66	-300.2	4.32	40.88	1.54	5.29
3	Purbalingga (D)	353.86	-5.53	59.27	-1.19	8.3	353.39	-5.55	59.83	-1.06	7.69
4	Banjarnegara (D)	502.53	-6.54	27.41	-4.13	4.3	497.48	-6.48	27.36	-3.97	4.1
5	Kebumen (D)	242.57	-4.01	61.09	-1.21	6.84	239.49	-3.96	60.97	-1.13	6.58
6	Purworejo (D)	284.93	-3.84	40.36	-1.77	4.21	317s.34	-4.18	37.03	-1.78	3.39
7	Wonosobo (D)	341	-4.21	29.31	-1.89	2.72	328.13	-4.06	30.09	-1.78	2.63
8	Magelang (D)	628.15	-8.96	45.31	-6.39	5.56	636.79	-9.07	45.22	-6.31	5.3
9	Boyolali (D)	590.04	-8.57	49.32	-5.97	6.56	586.27	-8.53	49.7	-5.77	6.42
10	Klaten (D)	-172.44	2.09	47.01	1.02	6.77	-188.98	2.28	47.6	1.08	6.58
11	Sukoharjo (D)	635.63	-8.17	28.66	-6.34	4.92	637.61	-8.15	27.6	-5.85	4.56
12	Wonogiri (D)	10.32	-0.15	41.29	-0.08	5.85	18.43	-0.26	41.49	-0.14	5.55
13	Karanganyar (D)	623.78	-8.54	38.06	-5.9	4.2	624.13	-8.58	38.55	-5.75	4.13
14	Sragen (D)	95.12	-1.28	42.53	-0.39	6.36	80.78	-1.07	42.27	-0.31	6
15	Grobogan (D)	-24.87	0.42	42.04	0.12	6.39	-65.69	1.02	41.52	0.29	6.03
16	Blora (D)	251.64	-4.09	60.27	-1.16	7.72	265.6	-4.25	59.6	-1.16	7.45
17	Rembang (D)	318.48	-3.98	24.85	-1.83	2.82	323.47	-4.06	25.02	-1.79	2.75
18	Pati (D)	637.69	-8.97	42.68	-6.65	5.62	642.57	-9.02	42.62	-6.46	5.38
19	Kudus (D)	393.69	-4.9	20.4	-3.13	2.89	391.74	-4.88	20.44	-2.89	2.64
20	Jepara (D)	263.64	-3.28	26.67	-1.55	3.13	268.59	-3.36	26.98	-1.51	3.08
21	Demak (D)	123.9	-1.54	37.59	-0.8	4.65	132.79	-1.66	37.73	-0.82	4.4
22	Semarang (D)	567.35	-6.82	7.22	-2.35	0.45	604.91	-7.24	4.86	-2.22	0.28
23	Temanggung (D)	32.15	-0.69	49.77	-0.2	5.29	76.93	-1.34	50.48	-0.35	4.99
24	Kendal (D)	287.84	-3.71	30.7	-1.85	2.98	283.29	-3.65	31.05	-1.78	2.92
25	Batang (D)	335.09	-4.24	23.59	-1.71	2.52	367.1	-4.66	22.98	-1.78	2.38
26	Pekalongan (D)	155.48	-1.8	27.48	-0.82	3.38	165.39	-1.91	27.04	-0.86	3.28
27	Pemalang (D)	639.84	-9.31	48.79	-6.83	6.35	646.3	-9.39	48.79	-6.73	6.12
28	Tegal (D)	-264.03	3.23	45.91	1.54	6.73	-273.27	3.3	47.2	1.49	6.43
29	Brebes (D)	631.38	-9.38	54.53	-6.83	6.67	628.79	-9.37	54.92	-6.56	6.52
30	Magelang (M)	263.15	-3.28	30.97	-1.82	3.74	270.93	-3.36	30.32	-1.81	3.54
31	Surakarta (M)	198.52	-2.24	26.06	-1.14	3.78	216.92	-2.42	24.19	-1.19	3.22
33	Salatiga (M)	474.06	-5.69	11.15	-2.47	0.96	492.91	-5.9	9.54	-1.97	0.64
33	Semarang (M)	649.6	-9.33	46.26	-6.78	5.44	644.78	-9.29	46.64	-6.52	5.16
34	Pekalongan (M)	448.03	-5.78	26.23	-3.81	4.05	441.2	-5.65	24.91	-3.57	3.72
35	Tegal (M)	281.15	-3.65	30	-1.77	2.84	297.5	-3.85	29.12	-1.77	2.65

The variable is said to be significant to the model if the value of $t_{cal} > t_{table}(0.025;31)$, from the modeling results, obtained 35 models for each weighted function that shows in Table 9 below.

Table 9. Models that obtained

| No	District (D)/ Municipality (M)	Model	
		Adaptive Kernel Bisquare	Adaptive Kernel Tricube
1	Cilacap (D)	$-221.42 + 41.703X_2$	$-231.241 + 41.823X_2$
2	Banyumas (D)	$-254.91 + 41.916X_2$	$-300.198 + 40.885X_2$
3	Purbalingga (D)	$353.86 + 59.26832X_2$	$353.389 + 59.826X_2$
4	Banjarnegara (D)	$502.53 - 6.538X_1 + 27.412X_2$	$497.482 - 6.477X_1 + 27.355X_2$
5	Kebumen (D)	$242.57 + 61.088X_2$	$239.29 + 60.97X_2$
6	Purworejo (D)	$284.93 + 40.359X_2$	$317.336 + 37.029X_2$
7	Wonosobo (D)	$341 + 29.306X_2$	$328132 + 30.093X_2$
8	Magelang (D)	$628.15 - 8.964X_1 + 45.315X_2$	$636.793 - 9.071X_1 + 45.215X_2$
9	Boyolali (D)	$590.04 - 8.964X_1 + 49.324X_2$	$586.272 - 8.529X_1 + 49.702X_2$
10	Klaten (D)	$-172.4 + 47.009X_2$	$-188.976 + 47.604X_2$
11	Sukoharjo (D)	$635.6 - 8.173X_1 + 28.66X_2$	$637.611 - 8.154X_1 + 27.603X_2$
12	Wonogiri (D)	$10.32 + 41.29X_2$	$18.431 + 41.494X_2$
13	Karanganyar (D)	$623.8 - 8.54X_1 + 38.06X_2$	$624.131 - 8.579X_1 + 38.551X_2$
14	Sragen (D)	$95.12 + 42.53X_2$	$80.7795 + 42.273X_2$
15	Grobogan (D)	$-24.87 + 42.04X_2$	$65.6847 + 41.521X_2$
16	Blora (D)	$251.6 + 60.27X_2$	$265.596 + 59.605X_2$
17	Rembang (D)	$318.5 + 24.85X_2$	$323.469 + 25.019X_2$
18	Pati (D)	$637.7 - 8.967X_1 + 42.68X_2$	$642.573 - 9.019_x1 + 42.618X_2$

No	District (D)/ Municipality (M)	Model Adaptive Kernel Bisquare	Adaptive Kernel Tricube
19	Kudus (D)	$393.7 - 4.899X_1 + 20.4X_2$	$391.743 - 4.878X_1 + 20.445X_2$
20	Jepara (D)	$263.6 + 26.67X_2$	$268.585 + 26.981X_2$
21	Demak (D)	$123.9 - 1.537X_1 + 37.59X_2$	$132.785 + 37.725X_2$
22	Semarang (D)	$567.4 - 6.816X_1$	$604.906 - 7.244X_1$
23	Temanggung (D)	$32.15 + 49.77X_2$	$76.927 + 50.478X_2$
24	Kendal (D)	$287.8 + 30.7X_2$	$283.286 + 31.05X_2$
25	Batang (D)	$335.1 + 23.59X_2$	$367.104 + 22.98X_2$
26	Pekalongang (D)	$155.5 + 27.48X_2$	$165.392 + 27.039X_2$
27	Pemalang (D)	$639.8 - 9.307X_1 + 48.79X_2$	$646.303 - 9.392X_1 + 48.793X_2$
28	Tegal (D)	$-264 + 45.91X_2$	$-273.267 + 47.2X_2$
29	Brebes (D)	$631.4 - 9.385X_1 + 54.53X_2$	$628.786 - 9.373X_1 + 54.919X_2$
30	Magelang (M)	$263.1 + 30.97X_2$	$270.919 + 30.325X_2$
31	Surakarta (M)	$198.5 + 26.06X_2$	$216.921 + 24.195X_2$
33	Salatiga (M)	$474.1 - 5.689X_1$	492.908
33	Semarang (M)	$649.6 - 9.334X_1 + 46.26X_2$	$644.784 - 9.29X_1 + 46.645X_2$
34	Pekalongan (M)	$448 - 5.783X_1 + 26.23X_2$	$441.202 - 5.954X_1 + 24.91_5$
35	Tegal (M)	$281.2 + 30X_2$	$297.498 + 29.12X_2$

From models that were obtained, there were three groups of significant variables in each observation location the first group is the district/ municipality with the HDI variable (X_1), which considerably affects NPP. The second group is the PP variable (X_2) which has a significant effect on NPP, while the third group is the HDI variable (X_1) and JP (X_2), which have significant effects. Table 10 is a breakdown of which district/ municipality is included in each group.

Table 10. The group of significant variables in the GWR model with the Adaptive Kernel Bisquare weighting function.

Group of Significant Variables	District (D)/ Municipality (M)
X_1	Semarang(D), Salatiga (M).
X_2	Cilacap(D), Banyumas(D), Purbalingga(D), Kebumen(D), Purworejo(D), Wonosobo(D), Klaten(D), Wonogiri(D), Sragen(D), Grobogan(D), Blora(D), Rembang(D), Jepara(D), Temanggung(D), Kendal(D), Batang(D), Pekalongan(D), Tegal(D), Magelang(M), Surakarta(M), Tegal(M).
X_1, X_2	Banjarnegara(D), Magelang(D), Boyolali(D), Sukoharjo(D), Karanganyar(D), Pati(D), Kudus(D), Demak(D), Pemalang(D), Brebes(D), Semarang(M), Pekalongan(M).

Based on Table 10, the visualization of the distribution of groups of variables that are significant to the NPP with GWR Adaptive Kernel Bisquare in Central Java Province is presented in Figure 2.
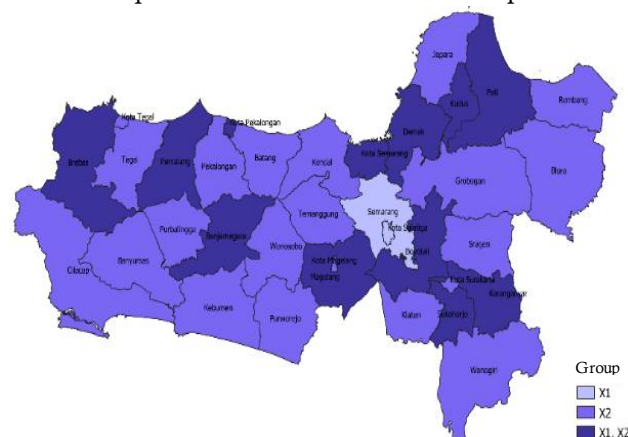
Figure 2. Thematic map of the grouping of influential variables using GWR Adaptive Kernel Bisquare.

Based on table 10 and Figure 2, the following illustrates the GWR model obtained for the Semarang municipality is presented in Equation 9.

$$y_{Semarang\,(M)} = 649.6 - 9.33X_1 + 46.26X_2 \qquad (9)$$

The model above shows that if the HDI (X_1) increases by 1 unit, the NPP (Y) will decrease by 9,33. If PP (X_2) has increased by 1 unit, there will be an increase in PP of 46.26 in Semarang municipality.

Based on Table 7 also, the parameter estimation value of the Adaptive Kernel Tricube function obtained that the HDI variable has an estimated minimum value of -9.39 and a maximum of 4.32. It can be said that HDI can affect NPP in Central Java Province with a range of estimated values between -9.39 to 4.32. Likewise, the 1st and 3rd quartile values, the HDI predictor variable has the 1st quartile value of the estimated GWR model parameter, which is -6.86, and the 3rd quartile value is -1.78, with a median or mean value of the parameter estimate of -4.06. From the results of the goodness of fit test, the F_{cal} value is 3.85. The value of F_{table}, F(α, 32, 14.57) = 2.26, it can be said that F_{cal} = 3.85 > 2.62 = F_{table} then the decision to reject H_0 is obtained, which means that there is a significant difference between the linear regression and the GWR model and the Adaptive Kernel Tricube weighting function.

In contrast to the Adaptive Kernel Bisquare function, the Adaptive Kernel Tricube function obtained four groups of significant variables at each observation location. The first group is the HDI variable (X_1) which has a significant effect on NPP, and the second group is the PP variable (X2) which has a significant impact on NPP. The third group is the variable HDI (X1), and PP (X2) has a significant effect on NPP. In the fourth group, both HDI (X1) and JP (X2) variables have no significant impact on NPP (Y) in the area. The details of the four groups are presented in Table 11 and Figure 3.

Table 11. The group of significant variables in the GWR model with the Adaptive Kernel Bisquare Tricube weighting function.

Group of Significant Variables	District/ municipality
X_1	Semarang(D)
X_2	Cilacap(D), Banyumas(D), Purbalingga(D), Kebumen(D), Purworejo(D), Wonosobo(D), Klaten(D), Wonogiri(D), Sragen(D), Grobogan(D), Blora(D), Rembang(D), Jepara(D), Demak(D), Temanggung(D), Kendal(D), Batang(D), Pekalongan(D), Tegal(D), Magelang(M), Surakarta(M), Tegal(M).
X_1, X_2	Banjarnegara(D), Magelang(D), Boyolali(D), Sukoharjo(D), Karanganyar(D), Pati(D) Kudus(D), Pemalang(D), Brebes(D), Semarang(M), Pekalongan(M).
–	Salatiga(M)
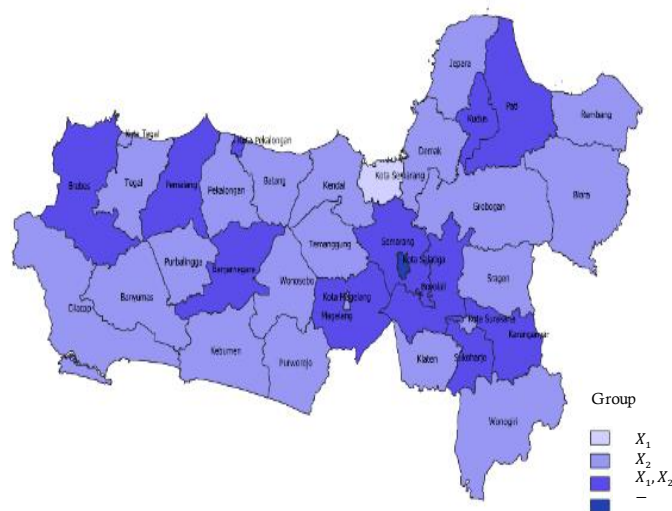
**Figure 3**. Thematic map of the grouping of influential variables using the GWR Adaptive Kernel tricube.

In selecting the best model, KPI values are used in the form of R^2, AIC, MAPE, MSE, and SSE values from linear regression and GWR models. KPIs from the linear regression model and GWR model are presented in Table 12.

Table 12. Determination of the best model

KPI	OLS	GWR Adaptive Kernel Bisquare	GWR Adaptive Kernel Tricube
R^2	79.64%	95.17%	94.71%
AIC	342.66	392.43	382.67
MAPE	35.99%	15.53%	16.36%
MSE	831.91	197.08	215.93
SSE	29118.14	6515.16	7091.72

Based on Table 12, it is found that the Adaptive Kernel Bisquare GWR model has a coefficient of determination or R^2 greater than the linear regression model and the Adaptive Kernel Tricube GWR, which is 95.17%. Thus, from looking at the value of the coefficient of determination or R^2, the GWR Adaptive Kernel Bisquare model is better used than the linear regression model. Furthermore, the model with smaller AIC, MAPE, MSE, and SSE values is the GWR Adaptive Kernel Bisquare model with AIC value of 392.43 and MSE of 197.08, SSE of 6515.16, MAPE of 15.52% (including good category) [35].

Based on this, the GWR model with the Adaptive Kernel Bisquare weighting function is the best model used in modeling the Number of Poor Population (NPP) in Central Java Province in 2020. The visualization of the actual data and the prediction results using the linear regression and GWR model is presented in Figure 4.
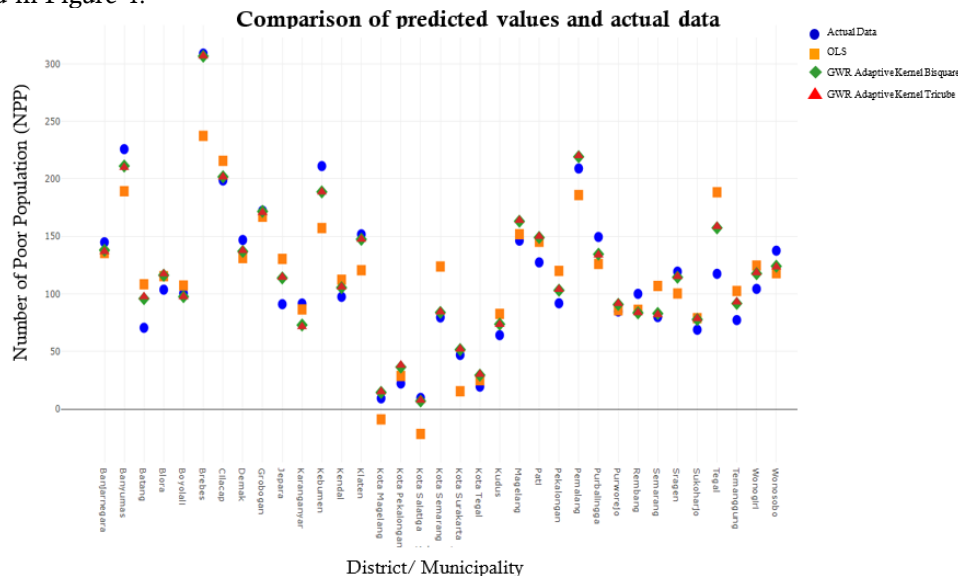
Figure 4. Comparison of original and predicted data using OLS & GWR models.

Figure 4 above shows that the prediction result is good if it has a small error or is close to the original data, which means it has a low difference. Visually, it can be said that the prediction results are said to be the best if the point is closer to the point that states the original data. From Figure 4 above, it can be seen visually that the green dot, namely GWR with Adaptive Kernel Bisquare, is closer to the blue dot, which is the original data, compared to the other points.

Conclusion

Based on the problems and objectives of this study, the area with the lowest NPP value is Magelang municipality with a total of 9.3 thousand inhabitants, and the highest is Brebes Regency with an NPP of 208.8 thousand inhabitants. GWR modeling is suitable for use in this study due to the differences in the nature or character of each location. Based on the values of R^2, AIC, MAPE, MSE, and SSE, it is found that the GWR model with the Adaptive Kernel Bisquare weighting function is the best model to predict the amount of poverty in Central Java province when compared to the OLS model. The example of result that obtained was model of Semarang Municipality shows that if the HDI (X_1) increases by 1 unit, the NPP (Y) will decrease by 9,33. If PP (X_2) has increased by 1 unit, there will be an increase in PP of 46.26.

Acknowledgement

References

[1] Badan Pusat Statistik, Kemiskinan dan ketimpangan, 2021. [Online]. Available: https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html. [Accessed May 28th, 2021]

[2] Badan Pusat Statistik Jawa Tengah, Data dan informasi kemiskinan provinsi Jawa Tengah 2016-2020, 2021.

[3] R. Fitriani and A. Efendi, Ekonometrika spasial terapan dengan R. Malang: UB Press, 2019.

[4] H. Yasin, B. Waryanto, and A. R. Hakim, Regresi spasial: aplikasi dengan R. Ponorogo: WADE Group, 2020.

[5] R. E. Caraka and H. Yasin, Geographically Weighted regression (GWR) sebuah pendekatan regresi geografis. MOBIUS, 2017.

[6] D. N. Gujarati, Ekonomi dasar. Jakarta: Erlangga, 1999.

[7] A. Efendi, N. W. S. Wardhani, R. Fitriani, and E. Sumarminingsih, Analisis regresi: teori dan aplikasi dengan R. Malang: UB Press, 2020.

[8] N. Hanifah, N. Herrhyanto, and F. Agustina, Penerapan metode weigthed least square untuk mengatasi heteroskedastisitas pada analisis regresi linear, Jurnal EurekaMatika 3(1) (2015) 105–114.

[9] A. L. Andrietya, A. Pujiati, and A. Setyadharma, Determinants of Poverty in Central Java Province 2013-2018, Journal of Economic Education 9(1) (2020) 81–88.

[10] S. Haryanto and G. A. Andriani, Pemodelan jumlah penduduk miskin di Jawa Tengah menggunakan Geographically Weighted Regression (GWR), Jurnal Litbang Sukowati Media Penelitian dan Pengembangan 4(2) (2019) 10.

[11] A. S. Fotheringham, H. Yue, and Z. Li, Examining the influences of air quality in China's cities using multi-scale Geographically Weighted Regression, Transaction in GIS 23(6) (2019) 1444–1464.

[12] E. H. Koh, E. Lee, and K. K. Lee, Application of Geographically Weighted Regression models to predict spatial characteristics of nitrate contamination: Implications for an effective groundwater management strategy, Journal of Environment Management 268 (2020) 110646.

[13] R. Zhao, M. Yao, L. Yang, H. Qi, X. Meng, and F. Zhou, Using Geographically Weighted Regression to predict the spatial distribution of frozen ground temperature: A case in the Qinghai–Tibet Plateau, Environmental Research Letter 16(2) (2021).

[14] Z. Ma, Y. Xue, and G. Hu, Geographically Weighted Regression analysis for spatial economics data: A Bayesian Recourse, International Regional Science Review 44(5) (2021) 582–604.

[15] A. Muche, M. S. Melaku, E. T. Amsalu, and M. Adane, Using Geographically Weighted Regression analysis to cluster under-nutrition and its predictors among under-five children in Ethiopia: Evidence from demographic and health survey, PLOS One 16 (5) (2021).

[16] Suyono, Analisis regresi untuk penelitian. Yogyakarta: Deepublish, 2018.

[17] E. Dwiningsih, Seemingly Unrelated Regression (SUR), Universitas Negeri Yogyakarta, 2012.

[18] N. Sari, H. Yasin, and A. Prahutama, Geographically Weighted Regression Principal Component Analysis (GWRPCA) pada pemodelan pendapatan asli daerah di Jawa Tengah, Jurnal Gaussian 5 (4) (2016) 717–726.

[19] M. Vebiriyana, M. Y. Darsyah, and I. M. Nur, Pemodelan geographically weighted regression dengan fungsi kernel bisquare terhadap faktor-faktor yang mempengaruhi tingkat kemiskinan di kabupaten Demak, Statistika 3 (1) (2015) 34–39.

[20] N. Lutfiani and S. Mariani, Pemodelan geographically weighted regression (GWR), UNNES Journal of Mathematics 5(1) (2017) 82–91.

[21] W. Jetz, C. Rahbek, and J. W. Lichstein, Local and global approaches to spatial data analysis in ecology, Global Ecology and Biogeography 14(1) (2015) 97–98.

[22] A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression: the analysis of spatially varying relationships, John Wiley & Sons Ltd 3 (2) (2002) 54–67.

[23] T. Wuryandari, A. Hoyyi, D. S. Kusumawardani, and D. Rahmawati, Identifikasi autokorelasi spasial pada jumlah pengangguran di Jawa Tengah menggunakan indeks moran, Media Statistika 7(1) (2014) 1–10.

[24] D. S. Ningtyas, Pemodelan Geographically Weighted Regression (GWR) dengan fungsi pembobot Adaptive Gaussian Kernel, Adaptive Bisquare Kernel dan Adaptive Tricube Kernel, 2019.

[25] W. Nurpadilah, I. M. Sumertajaya, and M. N. Aidi, Geographically Weighted Regression with kernel weighted function on Poverty Cases in West Java Province: Regresi Terboboti Geografis dengan fungsi pembobot Kernel pada data kemiskinan di provinsi Jawa Barat, Indonesian Journal of Statistics and Its Applications 5(1) (2021) 173–181.

[26] T. Hendrawati, A. H. Wigena, I. M. Sumertajaya, and B. Sartono, Performance evaluation of AIC and BIC in time series clustering with Piccolo method, in The 1st International Conference on Statistics and Analytics, 2020.

[27] C. Sammut and G. I. Webb, Eds., Mean Absolute Error, in Encyclopedia of Machine Learning, Boston, MA: Springer US, 2010, pp. 652.

[28] C. Sammut and G. I. Webb, Eds., Mean Squared Error, in Encyclopedia of Machine Learning, Boston, MA: Springer US, 2010, pp. 653.

[29] P. M. Swamidass, Ed., MAPE (mean absolute percentage error), in Encyclopedia of Production and Manufacturing Management, Boston, MA: Springer US, 2000, pp. 462.

[30] A. A. Grasa, Econometric model selection : a new approach advanced studies in theoretical and applied econometrics. Kluwer Academic, 1989.

[31] A. Widarjono, EKONOMETRIKA: teori dan aplikasi untuk ekonomi dan bisnis. 2005.

[32] J. M. Flenniken, S. Stuglik, and B. V. Iannone, Quantum GIS (QGIS): An introduction to a free alternative to more costly GIS platforms, Edis 2020 (2) (2020)

[33] D. Rosadi, Analisis statistika dengan R, 1st ed. Yogyakarta: Gadjah Mada University Press, 2016.

[34] D. T. Utari, Analisis regresi terapan dengan R. Yogyakarta: Universitas Islam Indonesia, 2019.

[35] P. C. Chang, Y. W. Wang, and C. H. Liu, The development of a weighted evolving fuzzy neural network for PCB sales forecasting, Expert Systems With Applications 32(1) (2007) 86–96.