ENTHUSIASTIC
INTERNATIONAL JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 2, Issue 1, April 2022, pp. 26-36

Geographically Weighted Regression with The Best Kernel Function on Open Unemployment Rate Data in East Java Province

Robiansyah Putra [a,1,*], Sischa Wahyuning Tyas [a,2], Muhammad Ghani Fadhlurrahman [a,3]

[a] Departement of Mathematics, Universitas Gadjah Mada, Sekip Utara BLS 21, Yogyakarta, 55281, Indonesia
[1] robian23calculus@mail.ugm.ac.id; [2] sischa.wahyuning@mail.ugm.ac.id; [3] mghanif1109@mail.ugm.ac.id
* corresponding author

ARTICLE INFO

Article history
Received February 26, 2022
Revised April 28, 2022
Accepted April 30, 2022

Keywords
Open Unemployment
Geographically Weighted Regression
Spatial Analysis
Bandwidth
Kernel

ABSTRACT

Unemployment is one of the problems that hinder employment development programs. Based on East Java BPS data, the Open Unemployment Rate in East Java in 2019 is about 3.92 percent. In 2020, unemployment increased by 466.02 thousand people. On the other hand, OUR increased by 2.02 percent to 5.84 percent in August 2020. In addition to the indicators affecting OUR, each observation location has different characteristics, so multiple linear regression modeling is inappropriate. Geographically Weighted Regression is one of the spatial analyses developed from multiple linear regression for data containing spatial heterogeneity effects. The weighting functions used for this GWR model are Kernel Fixed and Adaptive functions (Gaussian, Bi-Square, Tricube, and Exponential). The analytical step carried out in estimating the parameters is to use WLS. The best weighting was obtained in the test, namely the Adaptive Tricube. Based on the study results, the GWR model with Adaptive Tricube weighted resulted in the value of R-Squared = 87.02%. However, the best model is obtained from the GWR model with exponential weighting with the smallest Akaike Information Criterion (AIC) value compared to the others, AIC = 95.77804 with R-Squared = 77.41.

1. Introduction

Unemployment is one of the problems that hinder employment development programs. The indicator commonly used to measure the unemployment rate following the employment concept is the Open Unemployment Rate (OUR).

Based on East Java BPS (Badan Pusat Statistik) data, the Open Unemployment Rate in East Java in 2019 is about 3.92 percent. In 2020, unemployment increased by 466.02 thousand people. At the same time, OUR increased by 2.02 percent to 5.84 percent in August 2020. The high unemployment rate was influenced by the percentage of population density, average length of schooling, average income of informal workers, per capita expenditure on foodstuffs, rate of gross regional domestic

 p-ISSN 2798-253X
e-ISSN 2798-3153

product, regional minimum wage, percentage of the labor force to working age, unemployment with the last education being primary school, harvested area agriculture (ha).

In addition to the indicators mentioned above, each observation location has different characteristics. Therefore, a multiple linear regression modeling is not appropriate because a spatial effect on the Open Unemployment Rate (OUR) may result in the possibility of spatial diversity.

GWR analysis is an extension of the multiple linear regression model that considers location or geographical and fulfills the assumption of spatial aspect. Unlike the multiple linear regression models, a global regression model, the GWR model allows local parameters to be estimated rather than global parameters. One research on the OUR is the Spatial Regression Approach in Modeling the Open Unemployment Rate [1]. Furthermore, some research uses the GWR method: in 2018, research entitled "How Does Social Development Influence Life Expectancy? A Geographically Weighted Regression Analysis in China" [2] and in 2019 research entitled "GWR Modeling with Kernel Gaussian and Bi-square Functions" [3].

The analysis results with the GWR model are expected to explain the indicators that affect the OUR in East Java. It can be seen by comparing the best model between the multiple linear regression model and the GWR model.

2. Method

2.1 Data and Data Sources

The data that used in this research is secondary data related to the Open Unemployment Rate (OUR) in East Java Province in 2020 from BPS East Java which can be accessed through the website https://jatim.bps.go.id/ [4]. Observations in this research involved 39 regencies/cities in East Java, consisting of 29 regencies and 9 cities. The variables are Open Unemployment Rate (Y), amount of population (X1), average length of schooling (X2), per capita expenditure on foodstuffs (X3), rate of gross regional domestic product (X4), regional minimum wage (X5), percentage of the labor force to working age (X6), percentage of poor people (X7), and Residents aged 15 and over who worked during the past week Agriculture, Forestry, Fisheries/Agriculture, Forestry, and Fishing (X8). The steps of data analysis are defined as follows:

1. Create descriptive statistics.
2. Testing the multicollinearity of the independent variables.
3. Normality Assumption Test
4. Multiple Linear Regression
5. Spatial Heterogeneity Test
6. Geographically Weighted Regression (GWR) Model
7. Parameters Estimation GWR Model
8. Goodness of Fit Test
9. Simultaneous Test Parameter
10. Parameter Partial Test
11. Coefficient of Determination of Model and AIC Model
12. Conclusion

2.2 Multiple Linear Regression

Analysis regression is an analytical technique in statistics to examine the relationship and mathematical model between the response variable (Y) and one or more predictor variables (X).

Regression models that involve the relationship of the response variable with *p* predictor variable are called multiple regression models and can be expressed as follows [5]:

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+...+\beta_pX_{ip}+\varepsilon_i,\ \forall i=1,2,\dots,n \quad (1)$$

with β_0 and $\beta_1,\beta_2,...,\beta_p$ is a regression coefficient parameter and the error in equation (1) is assumed to be the average or $E(\varepsilon_i)=0$;i=1,2,…, n , variance or $var(\varepsilon_i)=\sigma^2$;i=1,2,…, n, covariance or $cov(\varepsilon_i,\varepsilon_j)=0, \forall i\neq j$, where *n* is the number of observations and *j* represents the observations for the *j*-th location. The linear regression model in equation (1) can be written as follows:

p-ISSN 2798-253X
e-ISSN 2798-3153

$$Y_1=\beta_0+\beta_1X_{11}+\beta_2X_{12}+...+\beta_pX_{1p}+\varepsilon_1$$
$$Y_2=\beta_0+\beta_1X_{21}+\beta_2X_{22}+...+\beta_pX_{2p}+\varepsilon_2 \quad (2)$$
$$\vdots$$
$$Y_n=\beta_0+\beta_1X_{n1}+\beta_2X_{n2}+...+\beta_pX_{np}+\varepsilon_n$$

In matrix form it can be written as follows:

$$Y=X\beta+\varepsilon \quad (3)$$

where

$$Y=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix};\ X=\begin{bmatrix}1 & X_{11} & X_{12} & \cdots & X_{1p}\\1 & X_{21} & X_{22} & \cdots & X_{2p}\\\vdots & \vdots & \vdots & \ddots & \vdots\\1 & X_{n1} & X_{n2} & \cdots & X_{np}\end{bmatrix};\ \beta=\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_p\end{bmatrix};\ \varepsilon=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix} \quad (4)$$

2.3 Geographically Weighted Regression

We will carry out spatial analysis if the research data fulfill the spatial aspect, which has spatial heterogeneity. Spatial heterogeneity is a condition where the same predictor variable has unequal influence at different locations within the research area. The Breusch-Pagan (BP) test is employed to detect the presence of spatial heterogeneity by having the decision criterion reject the null hypothesis H_0 if $BP_{test}>\chi^2_{(a,p)}$ [6].
The test statistics for Breusch-Pagan (BP) are:

$$BP=\left(\frac{1}{2}\right)f^TZ\left(Z^TZ\right)^{-1}Z^Tf \quad (5)$$

with the element vector f is:

$$f_i=\frac{e_i^2}{\hat{\sigma}^2}-1 \quad (6)$$

where e_i is the least squares error for i-th observation, σ^2 is the variance of e_i , and Z is a matrix of size $n\times(k+1)$which contains a vector that has been standardized for each observation.

The GWR model is a spatially varying coefficient regression approach to explore the spatial nonstationarity of the regression relationship for spatial data [7]. The GWR model extends the multiple linear regression model by allowing local parameters to be estimated rather than global parameters [9]. The formula of the GWR model is as follows:

$$y_i=\beta_0(u_i,v_i)+\beta_1(u_i,v_i)x_{i1}+\beta_2(u_i,v_i)x_{i2}+\ldots+\beta_k(u_i,v_i)x_{ik}+\varepsilon_i \quad (7)$$

where y_i is the dependent variable, i is the number of observation locations, x_{ik} is the independent variable, β_0 is the model intercept, β_k is the estimated coefficient parameter, (u_i,v_i) is the coordinates of the location of the i-th observation, and ε_i is the error in the observation i which is assumed to be independent, identical, and normally distributed.

Estimation parameter $\beta(u_i,v_i)$ at the i–th location, can be done using the weighted least squares or WLS method [8]. In estimating the parameters at a point location, the WLS method assigns unequal weighting to all observations. The amount of weighting is based on the distance between the observed locations. The closer the distance to the parameter estimated observations, the greater the weight in the estimate $\beta(u_i,v_i)$. The parameter estimator of the GWR model is obtained as follows:

$$\hat{\beta}(u_i,v_i)=\left(X'W(u_i,v_i)X\right)^{-1}X'W(u_i,v_i)Y \quad (8)$$

where $W_i=\text{diag}\,[w_1(u_i,v_i),w_2(u_i,v_i),\ldots,w_n(u_i,v_i)]$ is the weighted diagonal that varies from each predicted parameter at the i-th location.

p-ISSN 2798-253X
e-ISSN 2798-3153

GWR model parameter testing is done by partially testing the parameters. Tests are carried out to find out which parameters significantly affect the response variable. The form of the hypothesis is as follows:

H_0: $\beta_k(u_i,v_i)=0$, for each $k=1,2,\ldots,p$, $i=1,2,\ldots,n$ (There is no one predictor variable that affects the dependent variable)
H_1: There is at least one k between $1,2,\ldots,p$ with $\beta_k(u_i,v_i)\neq 0$, $i=1,2,\ldots,n$ (At least one predictor variable that affects the dependent variable)

The test statistics used are:

$$t_{hit}=\frac{\hat{\beta}_k(u_i,v_i)}{SE[\hat{\beta}_k(u_i,v_i)]}\ , \tag{9}$$

where SE is the standard error of $\hat{\beta}_k(u_i,v_i)$. The test criteria used are $|t_{hit}| > t_{1-\frac{\alpha}{2},(n-p-1)}$, if H_0 is rejected.

2.5 Weighting Model

Spatial weighting for parameter estimation using a spatial weighting matrix **(W)**.

$$\mathbf{W}_{ij}=\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{10}$$

The weighting matrix **(W)** is matrix that obtained based on the distance information from neighbors or the distance between one observation location and another observation location. Observation points closer to the regression point are given greater weight than observation points farther away. The selection of spatial weights for parameter estimation is very important. One of the weights used in GWR is the kernel function. The kernel function is used to estimate the parameters in the GWR model if the distance function is a continuous and monotonically descending function [9]. The weights formed by using this kernel function are the Gaussian kernel function, the Exponential kernel function, the Bi-square kernel function, and the Tricube kernel function. The weighting function form is as follows [10]:

a. Exponential

$$w_{ij}=\sqrt{\exp\left(-\left(\frac{d_{ij}}{h}\right)^2\right)}$$

b. Gaussian

$$w_{ij}=\exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right)$$

c. Bi Square

$$w_{ij}=\begin{cases} \left[\left(1-\left(\frac{d_{ij}}{h}\right)^2\right)\right]^2 & , d_{ij}<h \\ 0, & \text{others} \end{cases}$$

d. Tricube

$$w_{ij}=\begin{cases} \left[\left(1-\left(\frac{d_{ij}}{h}\right)^3\right)\right]^3 & , d_{ij}<h \\ 0, & \text{others} \end{cases}$$

where $d_{ij}=\sqrt{(u_i-u_j)^2+(v_i-v_j)^2}$ and h is a non-negative parameter commonly called bandwidth (smoothing parameter). The kernels above are *fixed* or have the same bandwidth at each observation location. The kernel function will be declared as *adaptive* when it has different bandwidth at each observation location. The bandwidth for adaptive kernel function symbolized as h_i.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.6 Optimum Bandwidth Selection

Bandwidth selection is one of the important things because it will affect the accuracy of the estimation results. One of the methods that can be used for bandwidth selection is Cross Validation at all locations, with the formula:

$$CV=\sum_{i=1}^{n}\left(y_i-\hat{y}_{\neq i}(h)\right) \quad (11)$$

$\hat{y}_{\neq i}$ is the estimated value at the observation of the i-th location with a certain bandwidth removed from the prediction process. The optimum bandwidth will be obtained by an iterative process until the minimum CV is obtained [11].

The next step after searching for bandwidth is to choose the best model among the kernel functions used by looking at the AIC value whose form is shown in the following equation:

$$AIC=2n\ln\hat{\sigma}+n\ln 2\pi+n+tr(S) \quad (12)$$

The model with the smallest AIC value is the best model [7].

3. Results and Discussion

3.1 Mapping the Open Unemployment Rate in East Java

East Java Province is a province in the east Java Island, Indonesia. The capital city of East Java in the city of Surabaya. Its area is 47,803.49 km², with a population of 40,665,696 people (2020) and a population density of 851 people/km2. East Java has the largest area among six provinces on the island of Java and has the second-largest population in Indonesia after West Java. Java Sea borders east Java in the north, Bali strait in the east, the Indian Ocean in the south, and Province Central Java in the west. The East Java region includes Madura Island, Bawean Island, Kangean Island, and several small islands in the Java Sea: Masa Lembu Island, Indian Ocean, Sempu Island, and Nusa Barung. Based on data from the Ministry of Home Affairs, in East Java, there are 38 regencies/cities. With details, as many as 29 are regencies, and 9 are cities. The mapping of the Open Unemployment Rate in East Java in 2020 is as follows:

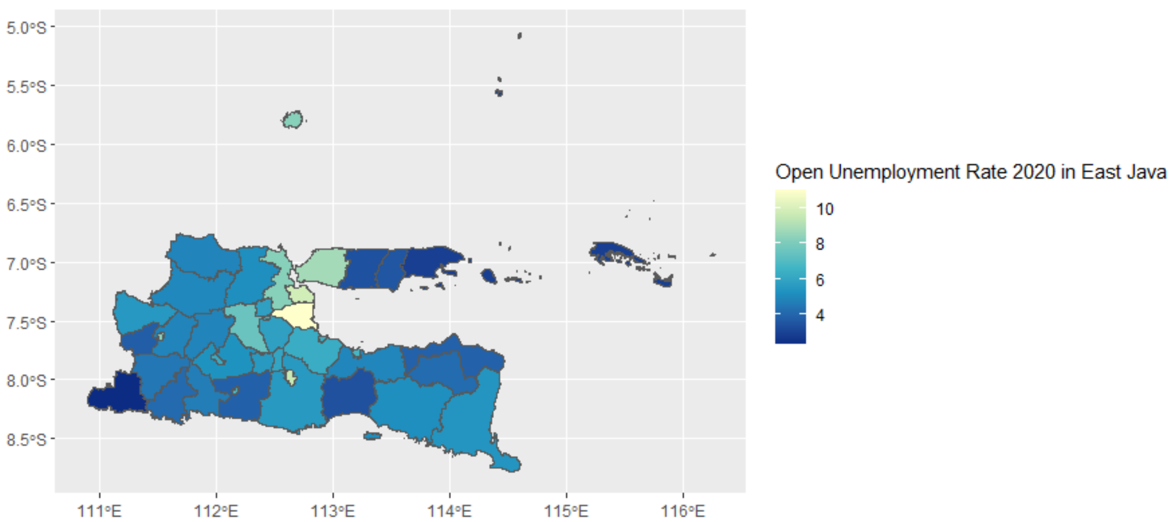

Fig. 1. Map of the OUR in East Java 2020

The description of the East Java Open Unemployment Rate data in 2020 is presented in

Table 1.

p-ISSN 2798-253X
e-ISSN 2798-3153

Table 1. Description of Research Variables

	Min	Max	Mean	Std. Deviation
Y	2.28	10.97	5.623947	2.003467993
X1	132	2874	1070.105	668.7220078
X2	4.85	11.14	7.941579	1.529964717
X3	37.04	62.9	50.25316	5.995050709
X4	-6.46	-0.29	-2.77263	1.522527565
X5	1913322	4200479	2446161	759360.0808
X6	65.75	80.36	70.61974	3.159913176
X7	3.89	22.78	11.02132	4.567803997
X8	1905	510626	182091.2	132203.791

Table 1. shows the highest standard deviation for the predictor variable X5, namely the Regional Minimum Wage (UMR) predictor. It means that the population of East Java has a regional minimum wage in the district/city. A relatively small standard deviation is found in the X5 variable, which is the predictor of the GRDP Rate, which means that the GRDP rate in East Java is evenly distributed in each district/city.

3.2 Open Unemployment Rate Modeling with Multiple Linear Regression

Multiple Linear Regression Modeling aims to determine the variables that affect the Open Unemployment Rate in East Java without considering the spatial aspect. Multiple Linear Regression is a global regression of the Geographical Weighted Regression model. The first step in Multiple Linear Regression analysis is to perform multicollinearity test and classical assumption test (identical, independent and normal distribution).

a. Multicollinearity Test

Multicollinearity test is seen from the value of Variance Inflation Factors (VIF).
The multicollinearity assumption is met if the VIF value is less than 10.

Table 2. Multicollinearity Test

Predictor	VIF
X1	4.303047
X2	8.606343
X3	5.744430
X4	1.315459
X5	2.663739
X6	1.466995
X7	3.840486
X8	7.3788

Table 2 shows that each predictor variable has a VIF value of less than 10, so it can be said that there is no multicollinearity between the predictor variables.

b. Normality test

Assuming normality test using Shapiro-Wilks, obtained p-value > 0.01706 then it is stated that H0 is accepted so that it can be concluded that the multiple linear regression model error follows a normal distribution.

Normality test is performed to determine whether the error is normally distributed or not. The hypothesis used is

p-ISSN 2798-253X
e-ISSN 2798-3153

H_0: Normal distribution error
H_1: Error is not normally distributed

with test statistics

$$T_3=\frac{1}{D}\left[\sum_{i=1}^{n} a_i(x_{n-i+1}-x_i)^2\right]$$

where

$$D=\sum_{i=1}^{n} a_i(x_i-\bar{x})^2$$

a_i = Shapiro Wilk's test coefficient
$x_{n-i+1} = n-i+1$ data
x_i = i-th data
$\bar{x}$ = data average

Based on the calculation, it was obtained that the *p-value* = 0.4513 > 0.05. Thus failed to reject H0 which means that the error in the data is normally distributed.

c. OUR Modeling with Multiple Linear Regression

OUR modeling and the factors that are thought to influence it using the Ordinary Least Square (OLS) parameter estimation method aims to determine the variables that are significant to OUR globally.
The following is the result of OUR modeling with multiple linear regression.

$$Y=17.22+0.001362X_1+0.3737X_1-0.0932X_1-0.1553X_1+1.006\times10^{-7}X_1-0.1778X_1+0.1735X_1-7.528\times10^{-6}X_1$$

3.3 Open Unemployment Rate Modeling with Geographically Weighted Regression

a. Test the Spatial Aspect of the Data

Spatial heterogeneity test was carried out to find out whether there were characteristics or uniqueness in each observation location. This test uses the Breusch Pagan test statistic. The hypothesis of the Breusch-Pagan test is as follows.

H_0: There is no effect of spatial heterogeneity (homoscedasticity)
H_1: There is an effect of spatial heterogeneity (heteroscedasticity)
Test Statistics

In this research, the hypotheses used in the Breusch-Pagan test:

H_0 : There is no difference in variance between regions.
H_1 : There are differences in variance between regions.

Test statistics:

$$BP=\frac{1}{2}f^TZ(Z^TZ)^{-1}Z^Tf$$

where

$f=(f_1,f_2,\ldots,f_n)^T$ with , $f=\left(\frac{\varepsilon_i^2}{\sigma^2}-1\right)$ $\varepsilon_i=y_i-\hat{y}_i$

Z is a matrix containing vectors that have been standardized for each observation.
Reject H0 if or with is the number of predictors $BP>\chi^2_{(p)}$p-value< αp (Anselin, 1988).

The test results at a significance level of 0.05 resulted in the BP test statistic value of 18.322 and p-value = 0.01894 <0.05. Then it was decided to reject H_0 which means the variance at each location is different (heterogeneous). Because one of the tests of the spatial aspect was fulfilled, further analysis was carried out using the Geographically Weighted Regression method.

b. Geographically Weighted Regression Models

To determine the optimum bandwidth, the Eucledian distance is determined first by calculating the distance lattitude and longitude (u_i, v_i). Latitude and longitude distance values in this study were carried out with the help of R software. The following is the Eucledian distance for several districts/cities in East Java.

Table 3. Eucledian distance for several districts/cities in East Java

	Pacitan	Ponorogo	…	Surabaya	Stone
Pacitan	0	0.2954657	…	0.7942921	0.4876474
Ponorogo	0.2954657	0	…	1.0841587	0.7810890
⋮	⋮	⋮	⋱	⋮	⋮
Surabaya	0.7942921	1.0841587	…	0	0.31064449
Stone	0.4876474	0.7810890	…	0.31064449	0

One of the steps that can be used to determine the optimum bandwidth is to use the AIC (Akaike Information Criterion) method. According to the AIC method, the best model is the model that has the smallest AIC value.

Table 4. Eucledian distance for several districts/cities in East Java

Bandwith		Bandwidth Value (h)	CV
Adaptive Kernels	Gaussian	28	71.34173
	Bi-Square	35	66.10363
	Tricubes	37	62.94622
	Exponential	30	68.31555
Fixed Kernels	Gaussian	1.075556	70.67708
	Bi-Square	2.68074	66.09748
	Tricubes	2.757565	65.91585
	Exponential	1.276672	68.65603

Overall, the kernel function that produces the optimum bandwidth is the adaptive Tricube weighting function because it has the smallest AIC value compared to other weighting functions as shown in the table above.

c. Parameters Estimation GWR Model

To determine the estimated parameters of the geographic weighted regression model, a weighted matrix is needed. The spatial weighting matrix is defined based on the optimum bandwidth value and the best kernel function, in this case the adaptive Gaussian weighting function. The estimation results of the geographic weighted regression model parameters are as follows.

Table 5. Estimation of Weighted Regression Model Parameters

County/City	Intercept	x1	x2	…	x8
Pacitan	20.60509	0.00139	0.138415	…	-8.04E-06
Ponorogo	20.63265	0.001426	0.137306	…	-8.23E-06
Trenggalek	14,73513	0.001028	0.571205	…	-3.94E-06
⋮	⋮	⋮	⋮	⋱	⋮
Surabaya	20.48311	0.001035	0.192997	…	-5.10E-06
Stone	20.06272	0.001241	0.169314	…	-7.00E-06

p-ISSN 2798-253X
e-ISSN 2798-3153

d. Goodness of Fit Test

This test was conducted to determine whether the geographically weighted regression model was significantly better in modeling the data than the global model (multiple linear regression model). The hypotheses used are:

H_0: $\beta_k(u_i,v_i)=\beta_k$, for each k=1,2,…,8
(There is no difference between the GWR model and the Multiple Linear Regression model)
H_1: There is at least one k between $1,2,\dots,8$ with $\beta_k(u_i,v_i) \neq \beta_k$
(GWR model explains data better than Multiple Linear Regression model)

The test statistics used are:

$$F=\frac{RSS(H_1)/\delta_1}{RSS(H_0)/(n-k-1)}=1.1188$$

Based on table F with a significance level of $\alpha = 0.10$ it is known that the F-table with $df_1 = 29$ and $df_2 = 27$ is 0.6137 so that $F_{hitung}=1.1188 > F_{tabel}=0.6137$. Therefore, it can be concluded that the GWR model is better than Multiple Linear Regression in explaining the 2020 East Java OUR (Open Unemployment) data.

e. Simultaneous Parameter Test

The results of the beta parameter estimator from OUR are used to test the suitability of the beta parameter coefficient simultaneously between multiple linear regression models and geographically weighted models with the following hypothesis

H_0 : $\beta_k(u_i,v_i)=0$, k=1,2,…,8
H_1 : $\beta_k(u_i,v_i)\neq 0$

Simultaneous testing of the case examples above, the statistical value of the F test of the geographically weighted regression model is 1.7009 which is greater than $F(0.10,27.1066,1.8934) = 0.3447$. This shows that there are predictor variables that have a significant effect on the response variables in each district or city.

f. Parameter Partial Test

This test was conducted to determine the parameters that have a significant effect on the response variable at the i-th predictor variables. The form of the hypothesis is as follows:

H_0: $\beta_k(u_i,v_i)=0$, for each k=1,2,…,p
H_1: There is at least one k between 1,2,…,p with $\beta_k(u_i,v_i)\neq 0$
Test statistics:

$$t_k=\frac{\hat{\beta}_k(u_i,v_i)}{\hat{\sigma}\sqrt{c_{ikk}}}$$

Based on this test, districts/cities in East Java are grouped into four groups based on independent variables that are significant to the Open Unemployment Rate in East Java.

Table 6. Significant Variables

Group	Area	Significant Variable
1	Ponorogo, Pamekasan	X_1, X_4, X_6, X_7, X_8

2	Pacitan, Blitar, Bondowoso, Probolinggo, Pasuruan, Bojonegoro, Bangkalan, Kota Blitar, Kota Pasuruan	X_1, X_6, X_7, X_8
3	Banyuwangi, Gresik, Kota Malang, Batu	X_1, X_6, X_8
4	Situbondo, Nganjuk, Sumenep, Kota Kediri, Kota Mojokerto, Surabaya	X_1, X_3, X_6
5	Trenggalek, Malang, Jember, Sidoarjo, Mojokerto, Ngawi, Tuban, Sampang, Kota Probolinggo	X_1, X_8
6	Tulungagung, Kediri, Lumajang, Jombang, Madiun, Magetan, Lamongan, Kota Madiun	X_1

Based on the results of the grouping of significant variables, it is possible to form a GWR model in each regency/city. One of the GWR models formed is the Open Unemployment Rate in Ponorogo Regency, with the following equation:

$$\hat{y}=20.63+0.001426X_1-0.24255X_4+0.18913X_6-0.16479X_7-0.00000823X_8$$

From the model formed, the Open Unemployment Rate in Pacitan Regency is influenced by amount of population (X_1), rate of gross regional domestic product (X_4), percentage of the labor force to working age (X_6), percentage of poor people (X_7), and Residents aged 15 and over who worked during the past week Agriculture, Forestry, Fisheries/Agriculture, Forestry, and Fishing (X_8).

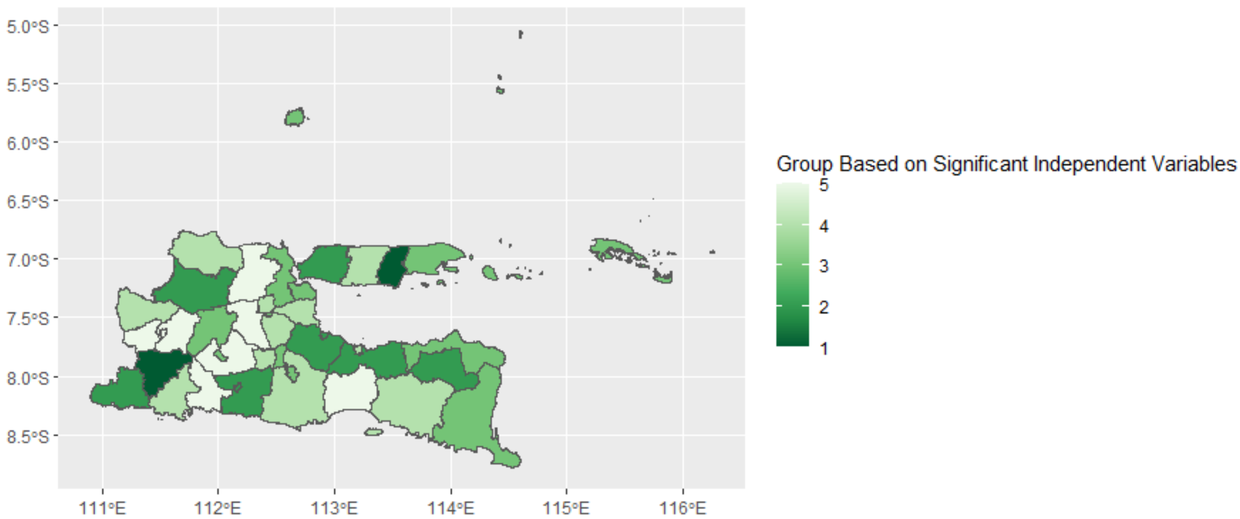

Fig. 2. Map of the significance variables

3.4 GWR Model Comparison and Multiple Linear Regression

The selection of the best model for the case of the Open Unemployment Rate in East Java in 2020 uses the AIC value. The best model is the model that has the smallest AIC value. **Table 5** shows that the GWR model with Adaptive Tricube Kernel Function Bandwidth has the smallest AIC value compared to Multiple Linear Regression model.

Table 7. Selection of the Best Model

	Regression Model	R-Square	Adjusted R-Square	AIC
	Multiple linear regression	0.8044	0.7505	117.6262
GWR	Adaptive Tricube	0.8702	0.7741	95.77804

p-ISSN 2798-253X
e-ISSN 2798-3153

4. Conclusion

The GWR Model with Adaptive Tricube Kernel Function Bandwidth is the best model for data on the Open Unemployment Rate (OUR) in East Java in 2020, with an AIC of 95.77804, an R-square of 87.02%, and an Adjusted R-Square of 77.41%. The significant factors that affected OUR in East Java in 2020 are the amount of population (X1), rate of gross regional domestic product (X4), percentage of the labor force to working age (X6), percentage of poor people (X7), and Residents aged 15 or more who worked during the past week Agriculture, Forestry, Fisheries/Agriculture, Forestry, and Fishing (X8).

References

[1] Mariana, "Pendekatan Regresi Spasial dalam Pemodelan Tingkat Pengangguran Terbuka", Jurnal Matematika dan Pembelajarannya, IAIN Ambon, Vol. 1, No. 1, pp. 42-63, 2013.

[2] J. Jiang, L. Luo, P. Xu, and P. Wang, "How Does Social Development Influence Life Expectancy? A Geographically Weighted Regression Analysis in China", Int J Public Health, Vol. 163, pp. 95-104, October 2018.

[3] N. Lutfiani, S. Sugiman, S. Mariani, "Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-Square", UNNES Journal of Mathematics, Vol. 8, No. 1, pp. 82-91, 2019.

[4] BPS, Badan Pusat Statistika, [Online]. Available: https://jatim.bps.go.id/.

[5] N. R. Draper, and H. Smith, Applied Regression Analysis, New York: John Wiley and Sons Inc, 1998.

[6] A. R. Tizona, R. Goejantoro, and Wasono, "Pemodelan Geographically Weighted Regression dengan Fungsi Pembobot Adaptive Kernel Bi-Square untuk Angka Kesakitan Demam Berdarah di Kalimantan Timur Tahun 2015", Jurnal Eksponensial, Vol. 8, No. 1, pp. 87-94, 2017.

[7] A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Chicester: John Wiley and Sons, 2002.

[8] C. Brunsdon, A. S. Fotheringham, and M. Charlton, "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity", Geographical Analysis, Vol. 28, No. 4, pp. 281-298, 1996.

[9] C. Chasco, I. Garcia, J. Vicens, "Modeling Spastial Variations in Household Disposible Income with Geographically Weighted Regression", Munich Personal RePEc Archive (MPRA), [Online]. Available: https://mpra.ub.uni-muenchen.de/1682/.

[10] H. Yasin, "Pemilihan Variabel pada Model Geographically Weighted Regression", Media Statistika, Universitas Diponegoro, Vol. 4, No. 2, pp. 63-72, 2011.

[11] I. Nurhuda, and I. G. N. M. Jaya, "Pemodelan Kriminal di Jawa Timur dengan Metode Geograhically Weighted Regression (GWR)", Jurnal Matematika MANTIK, UIN Sunan Ampel Surabaya, Vol. 4, No. 2, pp. 150-158, 2018.

 p-ISSN 2798-253X
e-ISSN 2798-3153