Evaluation of Factors Affecting Increased Unemployment in East Java Using NGWR-TS Method

by Sifriyani M.si

Submission date: 26-Sep-2019 09:14AM (UTC+0700)
Submission ID: 1180203187
File name: IJSBAR_Vol_45_1_2019.pdf (1.57M)
Word count: 6527
Character count: 32145

International Journal of Sciences:
Basic and Applied Research
(IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Evaluation of Factors Affecting Increased Unemployment in East Java Using NGWR-TS Method

Sifriyani[a*], Budiantara I. N.[b], Kartiko S. H.[c], Gunardi[d]

[a]*Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Samarinda Indonesia*

[b]*Departement of Statistics, Faculty of Mathematics and Natural Sciences, Institut Tekhnologi Sepuluh Nopember, Surabaya, Indonesia*

[c,d]*Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta Indonesia*

Abstract

Objectives: In this study, an experimental investigation of how the factors that affect TPT in East Java Indonesia from a spatiotemporal perspective. Study Design: Data were taken from 38 Regencies/Cities in East Java obtained from the Central Bureau of Statistics (BPS). Variables used in this research consisted of two types, namely respond variables and eight predictor variables. The amount of data used in the research is as much as 382. Method: Nonparametric Truncated Spline in the Geographically Weighted Regression method (NGWR-TS) was performed, a total of 38 Area were treated as units of our analysis. Result: Based on the results of the study, it shows that the unemployment rate has a geographical influence spatial heterogeneity. The unemployment rate has an unknown regression curve so that the NGWR-TS method is feasible to be used for modeling of the Unemployment rate. The NGWR-TS method has a model goodness of 80.42%. Significant factors that influence the unemployment rate are the Percentage of the poor population, Percentage of Low-Educated or elementary school dropouts workforce, economic growth rate, Investment ratio workforce number, Regional minimum wage, Ratio of the amount of Large-Medium Industry workforce number, Percentage of people working in the agricultural sector and Area of agricultural land.

----

* Corresponding author.

Conclusion: Each area in 38 Regencies/Cities has different significant variables so that each area has a different model from other areas. This is important for the Regional Government to take measures to reduce unemployment rates in East Java Province.

Keywords: Multivariate; Nonparametric geographically weighted Regression; Spatial; Truncated Spline; Unemployment Rate.

1. Introduction

Unemployment issues is still one of the biggest emphasis in development in Indonesia. The indicator for measuring the high unemployment rate is by looking at the open unemployment rate. According to the Central Bureau of Statistics or BPS open unemployment is someone who is looking for a job, preparing a business, not looking for a job, because they are impossible to get a job, or already have a job, but has not started work [3]. BPS survey results in 2017 stated that the number of open unemployment in Indonesia is as much as 6,18 percent of the total population. Unemployment has a negative impact on all sectors, including: the wasting of human resources, the burden of families and communities, the main source of poverty, promoting social and criminal unrest, and hampering the development of a region [3].

The imbalance between the growth of the labor force and the creation of employment opportunities has affected the migration of labor both spatially between village to city and in sector. This is in line with the statement of BPS Indonesia [4] which explains that the occurrence of migration caused by high wages or income that can be obtained in the destination. The presence of spatial effects is a common occurrence between one region and another, this means that one area affects the other. The availability of facilities allows one to move from one region to another, especially in the area of East Java, which has a proximity area between districts / cities to one another. Therefore, the existence of spatial dependence because of the location of adjacent areas and has the same characteristics allows unemployment in a region affected by unemployment in the surrounding area. Therefore, spatial modeling is needed to determine factors affecting open unemployment rate in East Java.

Unemployment is caused by three factors: population, education and economy. The population aspect that affects unemployment is rapid population growth. In education aspect, the education of working age population is still low in quality. While from the economic aspect, economic growth factor that has not been based on investment causes the economic sector has not been able to absorb adequate labor [4]. Based on these three aspects, in this research will discuss the factors that affect the TPT that is the education aspect includes the percentage of low-educated or elementary school dropout's workforce. For population aspect such as percentage of poor people, percentage of population working in agriculture sector and economic aspect such as economic growth rate, investment ratio per workforce number, regional minimum wage, ratio of large mid industrial quantities per workforce number and Area of agricultural land.

The problem of the open unemployment rate in East Java cannot be solved by using the GWR method. This is due to the relationship between the pattern of open unemployment and several independent variables and demographics do not form a linear and nonparametric pattern so that it cannot be solved by the GWR method.

Completion of the open unemployment rate is solved by the nonparametric truncated spline method in the weighted regression NGWR-TS [7, 8, 9]. The truncated spline is an approach used to solve the problem of changing relationships in several sub-sub intervals [6]. The results obtained using the NGWR-TS method have the goodness of the 80.42% model with the smallest MSE and AIC values.

The paper is organised as follows. Section 2 presents the statistical methodology employed, the reason for its choice and presents the theoretical model adopted and the variables selected to carry out the analysis. The results are discussed in Section 3 and the conclusions are presented in Section 4.

2. Literature Reviews

2.1. Method NGWR-TS

Nonparametric truncated spline in the geographically weighted regression NGWR-TS is the development of nonparametric regression for spatial data with parameter estimators local to each location of observation. Truncated spline approach is used to solve spatial analysis problems which its regression curve is unknown [6]. The assumption of the regression model used is the normal distributed error with mean zero and variance $\sigma^2(u_i, v_i)$ at each location (u_i, v_i). Location coordinates (u_i, v_i) is an important factor in determining the weights used to estimate the parameters of the model. Given data $(x_{1i}, x_{2i}, \dots, x_{li}, y_i)$ and relationship between $(x_{1i}, x_{2i}, \dots, x_{li})$ and y_i is assumed to follow multivariate nonparametric regression model as follows [8]:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{li}) + \varepsilon_i, \quad i = 1,2, \dots, n \tag{1}$$

With y_i as response variable and $f(x_{1i}, x_{2i}, \dots, x_{li})$ is unknown regression curve and assumed to be additive. If $f(x_{1i}, x_{2i}, \dots, x_{li})$ is approached with a truncated spline function. Mathematically, the relation between response variable y_i and the predictor variable $(x_{1i}, x_{2i}, \dots, x_{li})$ at *i*-th location for the multivariate nonparametric truncated spline regression model can be expressed as follows [7]:

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^{l}\sum_{k=1}^{m} \beta_{pk}(u_i, v_i)x_{pi}^k + \sum_{p=1}^{l}\sum_{h=1}^{r} \delta_{p,m+h}(u_i, v_i)\left(x_{pi} - K_{ph}\right)_+^m + \varepsilon_i\,, \tag{2}$$

with truncated function:

$$\left(x_{pi} - K_{ph}\right)_+^m = \begin{cases} \left(x_{pi} - K_{ph}\right), x_{pi} \geq K_{ph} \\ 0, \qquad\qquad x_{pi} < K_{ph} \end{cases}, \tag{3}$$

equation (2) is a multivariate nonparametric truncated spline regression model of degree *m* with *n* area. The components in Equation (2) and (3) are described as follows:

y_i is a response variable at *i*-th location, where $i = 1,2, \dots, n$,

x_{pi} is a *p*-th predictor variable at *i*-th location with $p = 1,2, \dots, l$,

K_{ph} is an h-th knot point in p-th predictor variable component with $h = 1,2, \ldots, r$,

$\beta_{pk}(u_i, v_i)$ is a polynomial component parameter of a multivariate nonparametric truncated spline regression. $\beta_{pk}(u_i, v_i)$ is a k-th parameter from p-th predictor variable at i-th location. $\delta_{p,m+h}(u_i, v_i)$ is a truncated component from multivariate nonparametric truncated spline regression. $\delta_{p,m+h}(u_i, v_i)$ is an $l + h$-th parameter in h-th knot point and p-th predictor variable at i-th location.

3. Data and Methodology

3.1. Data source and study area

Data were taken from 38 Regencies/Cities in East Java obtained from the Central Bureau of Statistics or BPS Institution. Variables used in this research consisted of two types, namely respond variables and eight predictor variables. The amount of data used in the research is as much as 382. Here's an explanation of each variable.

Dependent Variable

Percentage of Open Unemployment Rate (y)

Open unemployment rate (TPT) is the number of unemployment in regencies / cities divided by the number of labor force in regencies / cities multiplied by 100 percent [3]. The distribution and mapping of TPT in 2017 in 38 regencies / cities in East Java Province can be shown from Figure 2 where the percentage of TPT is divided into five categories, namely TPT value is very high, TPT value is high, TPT value is medium, TPT value is low and TPT value is very low.

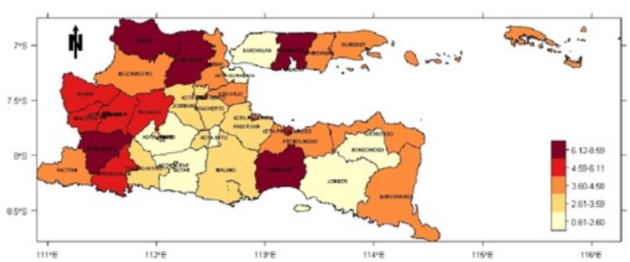

Figure 1: Mapping of Unemployment Rate (TPT) of 2017 in East Java

Based on Figure 1, it can be seen that the percentage of TPT of 2017 in 38 regencies / cities in East Java tend to be in the medium category i.e. there are 10 districts / cities (area) in the category.

Mapping Area of Independent Variables

1. Percentage of the Poor Population (x_1)

The poor population is those whose expenditure per capita is below the poverty line [1]. The poverty line consists of two components: the food poverty line and the non-food poverty line. The calculation of the poverty line is done separately for urban and rural areas. The poor population is people who have average expenditure per capita per month below the poverty line. Percentage of the Poor Population is presented in Figure 2.

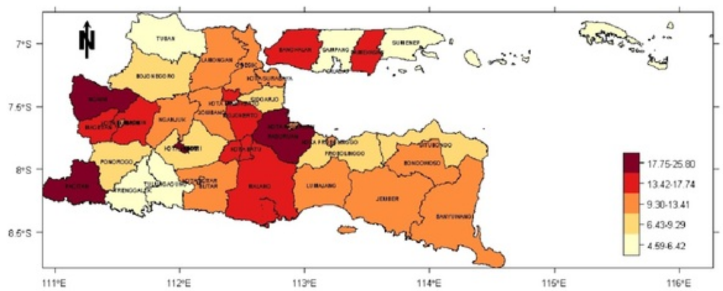

Figure 2: Mapping of Percentage of the poor population of 2017 in East Java

2. Percentage of Low-Educated or Elementary School Dropouts Work Force (x_2)

The percentage of people who are poorly educated or elementary school dropouts is the number of low-educated or unemployed primary workforce divided by the total number of workforce multiplied by100 percent [1]. Percentage of of Low-Educated or Elementary School Dropouts Work Force is presented in Figure 3.

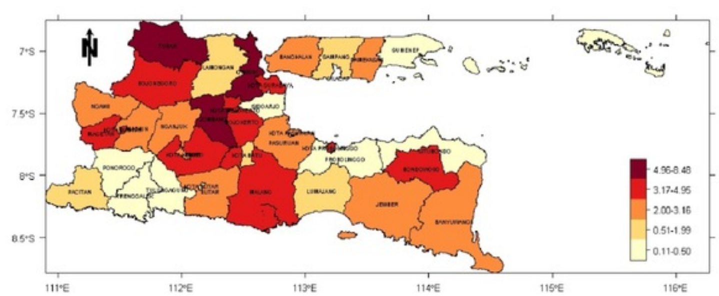

Figure 3: Mapping of Percentage of Low-Educated or Elementary School Dropouts Work Force of 2017 in East Java

3. Economic Growth Rate (x_3)

Economic Growth Rate is presented in Figure 4. The economic growth rate of regencies/municipalities derived from an increase in Gross Regional Domestic Product (GRDP) at constant prices (ADHK) from one year to the previous year is multiplied by 100 percent [2].

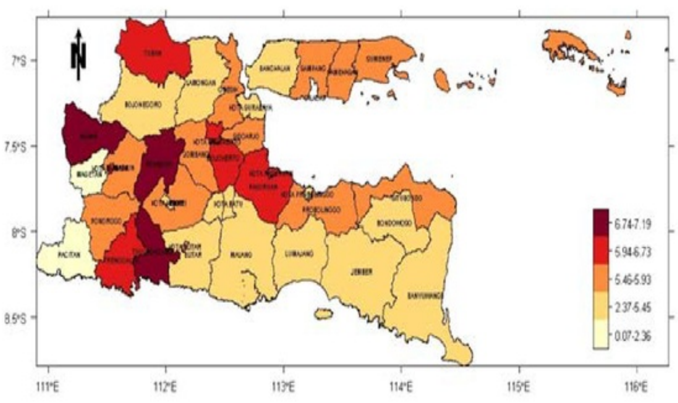

Figure 4: Mapping of Percentage of Economic Growth Rate of 2017 in East Java

4. Investment Ratio Workforce Number (x_4)

The investment ratio per Workforce Number is measured from the amount of investment in the Regency/City divided by the total workforce in the regency/city. Investment is derived from the amount of realized domestic investment (domestic investment) plus the realization of foreign investment (foreign investment). Investment Ratio Workforce Number is presented in Figure 5.

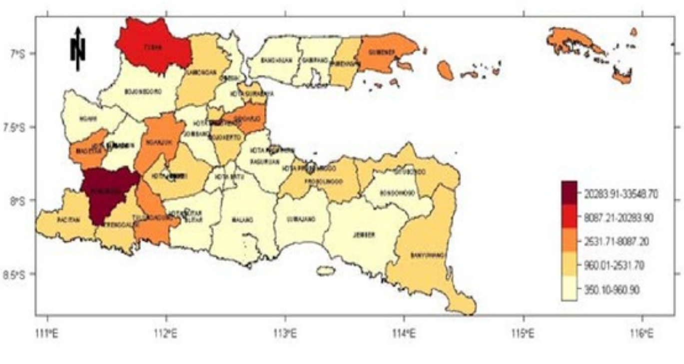

Figure 5: Mapping of Investment Ratio Workforce Number of 2017 in East Java

5. Regional Minimum Wage (x_5)

The regional minimum wage (UMR) is a minimum wage standard for a regency / city determined by the regional government in order to maintain the welfare of the workforce in their respective regions depending on the economic conditions [3]. Regional Minimum Wage is presented in Figure 6.

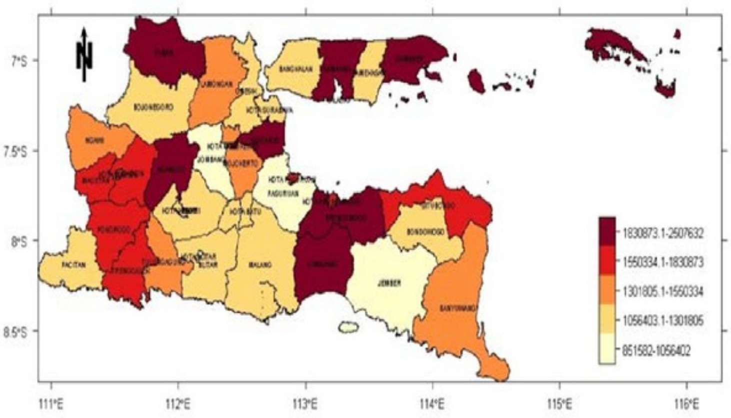

Figure 6: Mapping of Regional Minimum Wage of 2017 in East Java

6. Ratio of Amount of Large-Medium Industry Workforce Number (x_6)

The ratio of the number of large-medium industries is measured by the number of larga-medium industry in the District/City divided by the number of workforce. Large-medium industry is the processing of raw materials or semi-finished goods into finished goods that have added value to get the advantage that the amount of labor amounted to between 20 or more. Assembling and reparation are part of the industry. Industrial output is not only in the form of goods, but also in the form of services [2]. Ratio of Amount of Large-Medium Industry Workforce Number is presented in Figure 7.

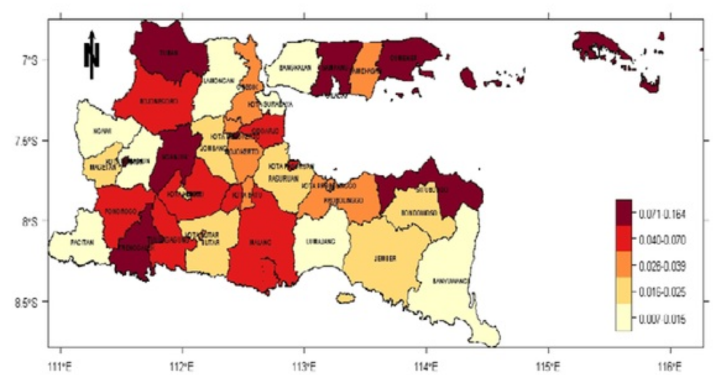

Figure 7: Mapping of Ratio of Amount of Large-Medium Industry Workforce Number of 2017

7. Percentage of People Working in the Agricultural Sector (x_7)

The percentage of the population working in agricultural sector is the percentage of the workforce with the status of free worker in agriculture divided by the workforce multiplied by 100 percent [1]. Percentage of People Working in the Agricultural Sector is presented in Figure 8.

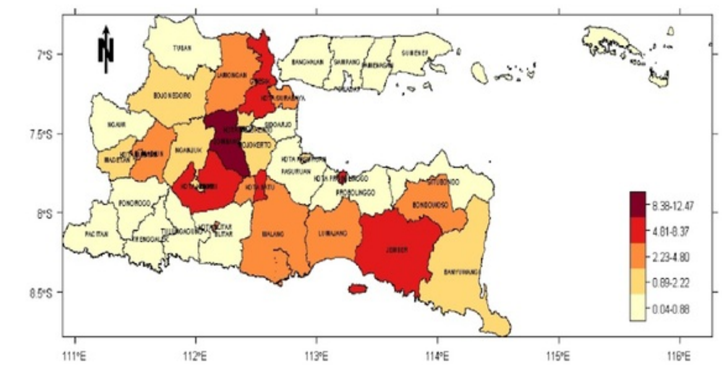

Figure 8: Mapping of Percentage of People Working in the Agricultural Sector of 2017 in East Java

8. Area of Agricultural Land (x_8)

The area of agricultural land is the area of irrigated and non-irrigated rice field by regency / city in East Java [1]. Area of Agricultural Land is presented in Figure 9.

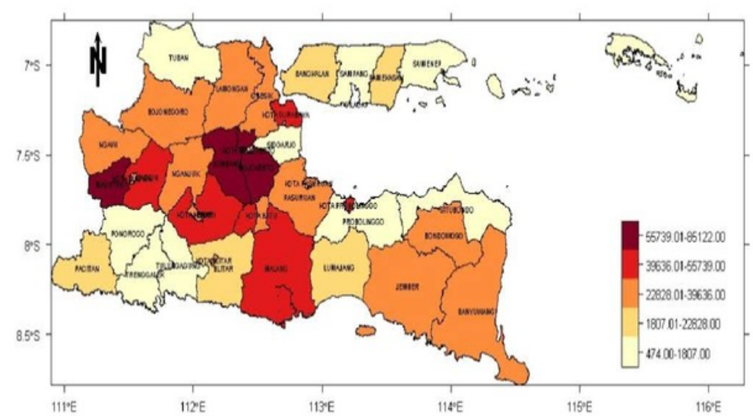

Figure 9: Mapping of Area of Agricultural Land of 2017 in East Java

3.2. Methodology

Steps to apply the nonparametric truncated spline in the geographically weighted regression (NGWR-TS) model on unemployment rate TPT data.

Step 1. Identifying response variables and predictor variables.

Step 2. Creating descriptive statistical analysis of each variable.

Step 3. Creating a scatter plot between the Y response variables with each predictor variable that is suspected to have the effect of knowing the data pattern.

Step 4. Testing of spatial heterogeneity with Breusch-Pagan method.

Step 5. Testing the model conformity hypothesis between the multivariable nonparametric truncated spline in the geographically weighted regression model versus a multivariable spline truncated nonparametric regression model.

Step 6. Calculating the Eucliden distance between the location i located at the coordinates (u_i, v_i) of the location j located at the coordinates(u_j, v_j).

Step 7. Determining the optimum knot point.

Step 8. Determining the weights for the multivariable nonparametric truncated spline in the geographically weighted regression model.

Step 9. Obtaining an estimator of a multivariable nonparametric truncated spline in the geographically weighted regression model.

Step 10. Interpretation of multivariable nonparametric truncated spline in the geographically weighted regression model.

Step 11. Testing the simultaneous hypothesis for multivariable nonparametric truncated spline in the geographically weighted regression model.

Step 12. Partial hypothesis test is the test of location influence partially on each predictor variable.

Step 13. Mapping 38 areas in East Java based on significant predictor variables.

Step 14. Drawing a conclusion.

4. Result and Discussion

In this chapter we will present result and discussion consisting of multivariable Nonparametric truncated spline in the geographically weighted regression NGWR-TS, optimal knot point selection, hypothesis testing simultaneously and partial hypothesis test.

4.1. Spatial Heterogeneity Test

The Breusch-Pagan Test in Table 1 is used to see the spatial heterogeneity of each location.

Table 1: *Breusch-Pagan Test*

Test	Significance Value
Breusch-Pagan	0,002414

Because spatial aspect testing is fulfilled i.e there are spatial heterogeneity effects, the case can be solved by using the point approach. Furthermore, the analysis will be done by using multivariable NGWR-TS model. Prior to modeling in advance, will be tested the model suitability as follows:

4.2. Model Suitability Test

Hypothesis for model conformity testing between multivariable NGWR-TS model versus multivariable spline truncated (global) nonparametric regression model using hypothesis formulation as follows:

$H_0: \beta_{pk}(u_i, v_i) = \beta_{pk}$ and $\delta_{p,m+h}(u_i, v_i) = \delta_{p,m+h}$,

$p = 1,2, \dots ,8;\ k = 1;\ h = 1,2,3;\ i = 1,2, \dots ,38$

H_1: Paling tidak ada satu $\beta_{pk}(u_i, v_i) \neq \beta_{pk}$ atau $\delta_{pm+h}(u_i, v_i) \neq \delta_{p,m+h}$,

$p = 1,2, \dots ,8;\ k = 1;\ h = 1,2,3;\ i = 1,2, \dots ,38$

Statistic test for fit model test between multivariable NGWR-TS model versus nonparametric multivariable spline truncated regression model [8] is given as follows:

$$V = \frac{\dfrac{\tilde{Y}^{T} S\, \tilde{Y}}{n - lm - 1}}{\dfrac{\tilde{Y}^{T}\, D(u_i, v_i)\, \tilde{Y}}{\mathrm{tr}\left((I - \xi)^{T} (I - \xi)\right)}} \tag{5}$$

Matrix is a matrix constructed from a multivariable spline truncated nonparametric model. Matrix is a matrix constructed from a multivariable NGWR-TS model. The numerator of equation (5) is given by:

$$\frac{\tilde{Y}^{T} S\, \tilde{Y}}{n - lm - 1} = 3,8 \tag{6}$$

with free degrees $n - lm - 1 = 29$. While the denominator of equation (5) is given by the:

$$\frac{\tilde{Y}^{T}\, D(u_i, v_i)\, \tilde{Y}}{\mathrm{tr}\left((I - \xi)^{T} (I - \xi)\right)} = 1,8 \tag{7}$$

With free degrees of $\mathrm{tr}\left((\mathrm{I}-\xi)^{\mathrm{T}}(\mathrm{I}-\xi)\right)=27{,}08$.= 1,88. Then the statistic test of $V=2{,}06$ is obtained. By using the level of significance of   0, 05, The conclusion of H_0 is rejected due to $V > F_{(0{,}05;\,29{,}27)} = 1{,}88$. Therefore, it can be concluded that there is a significant difference between multivariable NGWR-TS. Because of the influence of geographical factors on the model, the appropriate model used is multivariable NGWR-TS. Before making modeling, the best geographic weight is first sought.

4.3. Weighting of Geographic

The weighting function used for multivariable NGWR-TS model is the Gaussian kernel function and the Bisquare kernel function. The Gaussian kernel function [5] is given as follows:

$$w_j(u_i, v_i) = \emptyset\left(\frac{d_{ij}}{\sigma\, b}\right), j = 1,2, \dots, n, \tag{8}$$

where $\emptyset$ is the standard normal density function and shows the deviation standard of the distance vector $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$ and is the bandwidth value. The Bisquare kernel function [5] is given as follows:

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & for\ d_{ij} \le b \\ 0, & for\ d_{ij} > b \end{cases} \tag{9}$$

Obtain the best geographical weights, the optimum bandwidth value is required. In this research, the determination of optimum bandwidth value using cross validation (CV) method [5] is given in equation as follows:

$$CV = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_{\neq i}(b)\right)^2 \tag{10}$$

where $\hat{Y}_{\neq i}(b)$ is the estimated value on observations at the locations omitted from the estimation process. Here's the calculation of bandwidth and CV values presented on Table 2.

Table 2: CV Value for Weighted Functions.

		Cross Validation
Gaussian Kernel Function	1,69	126,97
Bisquare Kernel Function	3,03	132,27

From the two weighted functions are obtained the best model is to use Gaussian kernel function weighing with CV value smaller than the CV value of bisquare kernel function.

4.4. Selection of Optimum Knots Point

The next step for a multivariable NGWR-TS model is to determine knot points. The knot point is the point where the data pattern changes. Selection of this knot point uses GCV method. The smaller the GCV value the more optimum the selected knot point.

Selection of optimum knot point with one knot point

Selection of optimum knot points with one knot point for each predictor variable to TPT is expected to be found the smallest GCV that produces the best Spline model. Here is a multivariable NGWR-TS model with a single knot point.

$$\begin{aligned}\hat{y}_i = {} & \hat{\beta}_0(u_i,v_i) \quad \hat{\beta}_{11}(u_i,v_i)x_{1i} \quad \hat{\beta}_{21}(u_i,v_i)x_{2i} \quad \hat{\beta}_{31}(u_i,v_i)x_{3i} \\ & +\hat{\beta}_{41}(u_i,v_i)x_{4i} + \hat{\beta}_{51}(u_i,v_i)x_{5i} + \hat{\beta}_{61}(u_i,v_i)x_{6i} + \hat{\beta}_{71}(u_i,v_i)x_{7i} + \\ & +\hat{\beta}_{81}(u_i,v_i)x_{8i} + \hat{\delta}_{11}(u_i,v_i)(x_{1i}-K_{11})_+ + \hat{\delta}_{21}(u_i,v_i)(x_{1i}-K_{21})_+ + \\ & +\cdots+\hat{\delta}_{81}(u_i,v_i)(x_{1i}-K_{81})_+ \end{aligned} \qquad (12)$$

GCV values for the one-knot linear spline model are given in Table 3. The smaller the GCV value the more optimum the chosen knot points will be so that the multivariable NGWR-TS model is the best model.

Table 3: Selection of Optimum Knot Points with One Knot Point

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	GCV
1.	25,38	7,50	7,05	32878,43	2474511	0,16	12,22	83429,04	2,27*
2.	25,16	7,42	6,98	32543,29	2457950,5	0,16	12,10	82582,56	2,28
3.	24,95	7,35	6,91	32208,16	2441390	0,16	11,97	81736,08	2,30
4.	24,74	7,27	6,83	31873,02	2424829,5	0,16	11,85	80889,60	2,31
5.	24,53	7,20	6,76	31537,88	2408269	0,15	11,72	80043,12	2,32
6.	23,68	6,90	6,48	30197,34	2342027	0,15	11,23	76657,20	2,33
7.	23,89	6,98	6,55	30532,48	2358587,5	0,15	11.35	77503,68	2,34
8.	24,32	7,12	6,69	31202,75	2391708,5	0,15	11,60	79196,64	2,36
9.	24,10	7,05	6,62	30867,61	2375148	0,15	11,48	78350,16	2,36
10.	19,22	5,34	4,98	23159,48	1994256,5	0,12	8,62	58881,12	2,42

Based on Table 3 we get the optimum knot points with the smallest GCV value of 2,27, the knot point is:

$$K_{11} = 25{,}38;\ K_{21} = 7{,}50;\ K_{31} = 7{,}05;\ K_{41} = 32878{,}43;$$

$$K_{51} = 2474511;\ K_{61} = 0{,}16;\ K_{71} = 12{,}22;\ K_8 = 83429{,}04$$

The smallest GCV values for the multivariable NGWR-TS model with one knot point will be compared with the GCV model with two knots and three knots to find the best model.

Selection of optimum knots with two knots

After the selection of a knot point with a knot point, the next step is to select the knot point with two knots to look for the smallest GCV. Based on Table 5 in the Appendix, the optimum knots are obtained with the smallest GCV values of 2,7261 is as follows:

$K_{11} = 6{,}2868;\ K_{12} = 22{,}8306;\ K_{21} = 0{,}80526;\ K_{22} = 6{,}603715;$

$K_{31} = 0{,}6396;\ K_{32} = 6{,}1932;\ K_{41} = 2716{,}188;\ K_{42} = 28856{,}8;$

$K_{51} = 984066;\ K_{52} = 2275785;\ K_{61} = 0{,}019961;\ K_{62} = 0{,}142157;$

$K_{71} = 1{,}034963;\ K_{72} = 10{,}73027;\ K_{81} = 7245{,}84;\ K_{82} = 73271{,}28$

The GCV value at knot point selection with two knots is greater than the GCV value at knot point selection with one knot. This reflects that a multivariable NGWR-TS model with one knot point is better than the multivariable NGWR-TS model with two knot points.

Selection of optimum knots with three knots

The smallest GCV value for the multivariable NGWR-TS model with three knots is 2,643. The GCV value is smaller than the GCV value for the regression model with two knot points, but the GCV value is still greater when compared to the GCV value in the regression model with a knot point. Based on Table 6, the optimum knot point for the multivariable NGWR-TS model with three knot points is as follows:

$K_{11} = 6{,}2868;\ K_{12} = 22{,}8306;\ K_{13} = 25{,}3758;$

$K_{21} = 0{,}80526;\ K_{22} = 7{,}272768;\ K_{23} = 7{,}495785;$

$K_{31} = 0{,}6396;\ K_{32} = 6{,}834;\ K_{33} = 7{,}0476;$

$K_{41} = 2716{,}188;\ K_{42} = 31873{,}02;\ K_{43} = 32878{,}43;$

$K_{51} = 984066;\ K_{52} = 2424830;\ K_{53} = 2474511;$

$K_{61} = 0{,}019961;\ K_{62} = 0{,}156256;\ K_{63} = 0{,}160956;$

$K_{71} = 1{,}034963;\ K_{72} = 11{,}84896;\ K_{73} = 12{,}22186;$

$K_{81} = 7245{,}84;\ K_{82} = 80889{,}6;\ K_{83} = 83429{,}04;$

Based on the smallest GCV value on selection of optimum knot point with one knot point, two knot points and three knot points, the decision to use the best model is multivariable NGWR-TS model with one optimum knot point. Furthermore, we find the parameter estimator of the model.

4.5. Parameter Estimator of Multivariable NGWR-TS

Below is parameter estimates from the multivariable NGWR-TS model for the 30th area that is Kediri City that owns TPT from 38 areas in the east java.

$$\begin{aligned} \hat{y}_{30} = {} & 0{,}00938\quad 0{,}{,}04198x_{1;30}\quad 0{,}002691\left(x_{1;30}\quad 25{,}38\right)_{+}\quad 0{,}288484x_{2;30}\quad 0{,}004111\left(x_{2;30}\quad 7{,}50\right)_{+} \\ & -0{,}13702x_{3;30} - 0{,}00034\left(x_{3;30} - 7{,}05\right)_{+} - 0{,}000084x_{4;30} + 0{,}010486\left(x_{4;30} - 32878{,}43\right)_{+} + \\ & -0{,}0000034x_{5;30} - 0{,}000068\left(x_{5;30} - 2474511\right)_{+} - 0{,}00093x_{6;30} - 0{,}00000032\left(x_{6;30} - 0{,}16\right)_{+} + \\ & -0{,}08803x_{7;30} - 0{,}00301\left(x_{7;30} - 12{,}22\right)_{+} - 0{,}000084x_{8;30} - 0{,}00086\left(x_{8;30} - 83429{,}04\right)_{+} \end{aligned} \quad (13)$$

Interpretation of the above model is described as follows. If the variables $x_2, x_3, x_4, x_5, x_6, x_7$ are considered constant, the TPT (y) model for the percentage of the poor is as follows:

$$\hat{y}_{30} = 0{,}04198x_{1;30} + 0{,}002691\left(x_{1;30} - 25{,}38\right)_{+} \quad (14)$$

$$= \begin{cases} 0{,}04198x_{1;30} & x_{1;30} < 25{,}38 \\ 0{,}04467x_{1;30} - 0{,}068298 & x_{1;30} \geq 25{,}38 \end{cases}$$

Based on the above model, it is explained if the percentage of poor people in the Kediri city is less than 25.38 percent so in this condition, for the percentage of poor people that rose 1 percent in this area so TPT will also increase by 0.042 percent. In case the percentage of poor people in the Kediri city is more than or equal to 25,38 percent then in this condition it can be explained, to increase percentage of poor people by 1 percent in Kediri City so TPT will also increase by 0,045 percent.

If the variables $x_1, x_3, x_4, x_5, x_6, x_7$ and x_8 are considered constant, the TPT (y) model for the percentage of low educated or elementary school dropout (x_2) is

$$\hat{y}_{30} = 0{,}288484\ x_{2;30} + 0{,}004111\left(x_{2;30} - 7{,}50\right)_{+} \quad (15)$$

$$= \begin{cases} 0{,}288484\ x_{2;30} & x_{2;30} < 7{,}50 \\ 0{,}2926\ x_{2;30} - 0{,}0308 & x_{2;30} \geq 7{,}50 \end{cases}$$

The interpretation of above model is on the percentage condition of low educated or elementary school dropout workforce in Kediri City less than 7.50 percent hence for any increase in the percentage of low-educated or elementary school dropout which increased 1 percent in this area resulted in TPT will also increase by 0.288 percent. While for percentage condition of low educated or elementary school dropout workforce in Kediri City is more than or equal to 7,50 percent hence to increase percentage of low educated elementary school dropout work force which increased 1 percent in this area result TPT also will increase of 0.293 percent.

If the variables $x_1, x_2, x_4, x_5, x_6, x_7$ dan x_8 are considered constant, the TPT (y) model for economic growth rate (x_3) is

$$\hat{y}_{30} = -0,13702x_{3;30} - 0,00034\left(x_{3;30} - 7,05\right)_+ \tag{16}$$

$$= \begin{cases} -0,13702\ x_{3;30} & x_{3;30} < 7,05 \\ -0,13736\ x_{3;30} + 0,002397 & x_{3;30} \geq 7,05 \end{cases}$$

If the variables $x_1, x_2, x_3, x_5, x_6, x_7$ dan x_8 are considered constant, the TPT (y) model for investment ratio workforce number (x_4) is

$$\hat{y}_{30} = -0,000084x_{4;30} + 0,010486\left(x_{4;30} - 32878,43\right)_+ \tag{17}$$

$$= \begin{cases} -0,000084x_{4;30} & x_{4;30} < 32878,43 \\ 0,010402x_{4;30} - 344,7632 & x_{4;30} \geq 32878,43 \end{cases}$$

If the variables $x_1, x_2, x_3, x_4, x_6, x_7$ dan x_8 are considered constant, the TPT (y) model for regional minimum wage (x_5) is

$$\hat{y}_{30} = -0,0000034\ x_{5;30} - 0,000068\left(x_{5;30} - 2474511\right)_+ \tag{18}$$

$$= \begin{cases} -0,0000034x_{5;30} & x_{5;30} < 2474511 \\ -0,0000714x_{5;30} + 168,26674 & x_{5;30} \geq 2474511 \end{cases}$$

If the variables $x_1, x_2, x_3, x_4, x_5, x_7$ dan x_8 are considered constant, the TPT (y) model for ratio of amount of large-medium industry workforce number (x_6) is

$$\hat{y}_{30} = -0,00093x_{6;30} - 0,0000032\left(x_{6;30} - 0,16\right)_+ \tag{19}$$

$$= \begin{cases} -0,00093x_{6;30} & x_{6;30} < 0,16 \\ -0,0009303x_{6;30} + 5,12\text{E-}08 & x_{6;30} \geq 0,16 \end{cases}$$

If the variables $x_1, x_2, x_3, x_4, x_5, x_6$ dan x_8 are considered constant, the TPT (y) model for percentage of people working in the agricultural sector (x_7) is

$$\hat{y}_{30} = -0,08803\, x_{7;30} - 0,00301\left(x_{7;30} - 12,22\right)_+ \tag{20}$$

$$= \begin{cases} -0,08803\, x_{7;30} & x_{7;30} < 12,22 \\ -0,09104\, x_{7;30} + 0,0367822 & x_{7;30} \geq 12,22 \end{cases}$$

If the variables $x_1, x_2, x_3, x_4, x_5, x_6$ dan x_7 are considered constant, the TPT (y) model for area of agricultural land (x_8) is

$$\hat{y}_{30} = -0,0000084\, x_{8;30} - 0,00086\left(x_{8;30} - 83429,04\right)_+ \tag{21}$$

$$= \begin{cases} -0,0000084 x_{8;30} & x_{8;30} < 83429,04 \\ -0,0008684 x_{8;30} + 71,7489744 & x_{8;30} \geq 83429,04 \end{cases}$$

Multivariable NGWR-TS model with one knot point has R-Square or coefficient of determination equal to 80,42. This indicates that the multivariable NGWR-TS model is able to explain the influence of predictor variables on TPT response variable of 80.42%.

The result of the open unemployment rate estimator variable $(\hat{Y})$ using a multivariable NGWR-TS model is shown in Figure 13. The result of a multivariable NGWR-TS model has a value close to the original data (Y) shown in Figure 13.

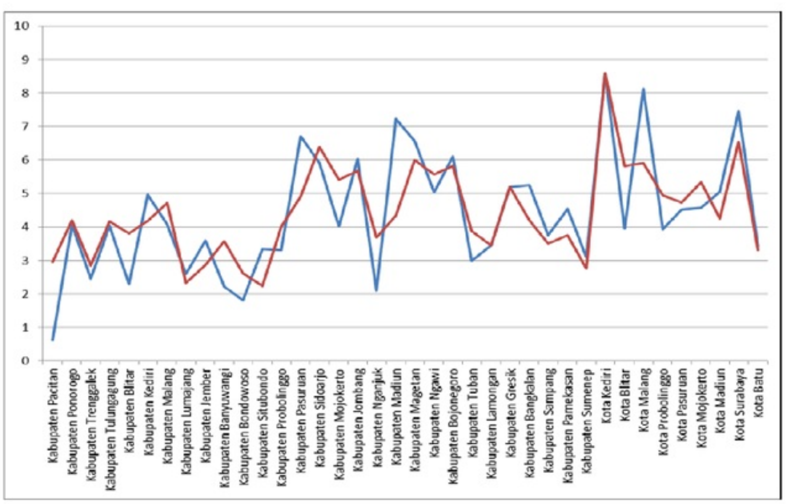

Figure 13: TPT distribution pattern with multivariable NGWR-TS model.

The red line is . $(\tilde{Y})$. or the result of the open unemployment rate estimator variable by using the multivariable NGWR-TS model and the blue line is the open unemployment observation data in east Java in 2017.

4.6. Hypothesis Test of Multivariable NGWR-TS

Hypothesis test is a decision-making method based on data analysis. Some steps in the hypothesis test procedure is to formulate hypotheses of H_0 and H_1, determine test statistics, determine the distribution of test statistics, determine the area of rejection and make a conclusion. Hypothesis test for multivariable NGWR-TS model consisted of two hypothesis test that is simultaneous test for parameters in NGWR-TS model and partial test or individual test for each parametric multivariable NGWR-TS model.

Simultaneous Test NGWR-TS

The hypothesis for testing model parameters simultaneously is as follows:

$H_0: \beta_{11}(u_i, v_i) = \beta_{21}(u_i, v_i) = \cdots = \beta_{81}(u_i, v_i) = \delta_{11}(u_i, v_i) = \cdots = \delta_{81}(u_i, v_i) = 0$

H_1: at least there is one $\beta_{pk}(u_i, v_i) \neq 0$ or $\delta_{p,m+h}(u_i, v_i) \neq 0$,

$p = 1,2, \ldots, 8;\ k = 1\ ;\ h = 1,2,3\ ;\ i = 1,2, \ldots, 38$

The eight predictor variables simultaneously affect the open unemployment rate, the statistic test for simultaneous test of $V^* = 1{,}76$ the multivariable NGWR-TS model (Sifriyani, 2018c) is required given as follows:

$$V^* = \frac{\left(\dfrac{\tilde{Y}^T M(u_i, v_i)\,\tilde{Y}}{tr\left((I - B_\omega)^T (I - B_\omega)\right)}\right)}{\left(\dfrac{\tilde{Y}^T D(u_i, v_i)\,\tilde{Y}}{tr\left((I - \xi)^T (I - \xi)\right)}\right)} \tag{22}$$

The result of the numerator on the equation (22) is given by:

$$\left(\frac{\tilde{Y}^T M(u_i, v_i)\,\tilde{Y}}{tr\left((I - B_\omega)^T (I - B_\omega)\right)}\right) = 3,24 \tag{23}$$

The denominator of equation (22) is as follows:

$$\left(\frac{\tilde{Y}^T D(u_i, v_i)\,\tilde{Y}}{tr\left((I - \xi)^T (I - \xi)\right)}\right) = 1,84 \tag{24}$$

So the calculation results for statistic test is $V^* = 1,76$ By using the significance level of $\alpha = 0,05$ the conclusion of H_0 is rejected due $V^* = 1,76 > F_{(0,05;\ 36,27)} = 1,60$. So it can be concluded that there is at least one parameter in the multivariable NGWR-TS model is significant against the response variable.

Partial Test

The following is a partial test hypothesis of model parameter in multivariable NGWR-TS.

$H_0: \beta_{pk}(u_i, v_i) = 0$ and $\delta_{p,m+h}(u_i, v_i) = 0$

with $p = 1,2, \dots ,8;\ k = 1;\ h = 1,2,3\ ;\ i = 1,2, \dots ,38$

H_1: at least there is one $\beta_{pk}(u_i, v_i) \neq 0$ and $\delta_{p,m+h}(u_i, v_i) \neq 0$

with $p = 1,2, \dots ,8;\ k = 1;\ h = 1,2,3\ ;\ i = 1,2, \dots ,38$,

Statistics Test for Partial Tests on multivariable NGWR-TS using the formula as follows:

$$t = \frac{\hat{\tilde{\eta}}(u_i, v_i)}{SE\left(\hat{\tilde{\eta}}(u_i, v_i)\right)}, \tag{25}$$

where $SE\left(\hat{\tilde{\eta}}(u_i, v_i)\right) = \sqrt{g_{kk}}$

and g_{kk} is the diagonal element to $k+1$ from matrix $\left(Q^{T} W(u_i, v_i) Q\right)^{-1} \hat{\sigma}^2(u_i, v_i)$.

The test statistic for the multivariable spline truncated nonparametric regression model in GWR in equation (25) will follow the t distribution with the free degree $(n - 1)$ and the significance level of α, then the decision taken will reject H_0 if the value is $|t| > t_{\left(\frac{\alpha}{2},(n-1)\right)}$.

To find out which predictor variables have a significant effect partially to the open unemployment rate for each location can be seen from the statistical value of t and the value of P- value. From the t-test results obtained significant predictor variables for each area and obtained 10 clustering at each research location. Grouping of areas based on significant variables on TPT is given in Table 7. TPT mapping in Table 7 can be presented in Figure 10 as follows:

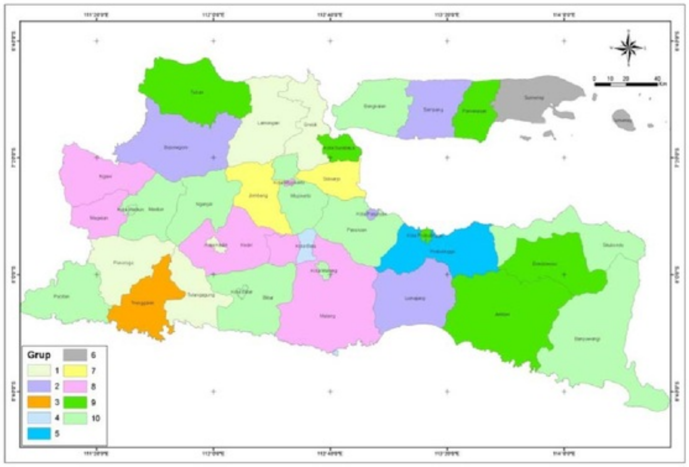

Figure 10: Mapping of TPT Estimator of 2017 in East Java based on the significant variables

Based on Figure 10, Lamongan and Gresik Regencies are located in adjacent areas and these two areas have the same characteristics, all predictor variables are significant. Likewise, some other areas such as Mojokerto and Pasuruan Regencies are located in adjacent areas and have the same characteristics that are variables and significant that define TPT of 2017 on the area is influenced by the rate of economic growth and UMR.

The first law of geography proposed in Tobler's theory in Anselin (1988) said that "Everything is related to everything else, but near things are more related than distant things". Everything is related to one another, but something closer has more influence than something far away. Based on Figure 10 there are several adjacent areas and have similar characteristics, so that the Tobler Theory proposed by Anselin can be applied in multivariable NGWR-TS model.

5. Conclusion

This paper presented results of factors that influence the unemployment rate (TPT) are the Percentage of the poor population, Percentage of Low-Educated or elementary school drop outs work force, economic growth rate, Investment ratio workforce number, Regional minimum wage, Ratio of amount of Large-Medium Industry workforce number, Percentage of people working in the agricultural sector and Area of agricultural land. The NGWR-TS method has a model goodness of 80.42%. In this research, we successfully applied NGWR-TS models. The modeling is used because the relationship formed between the TPT and each predictor variable shown in Figure 12 does not follow a certain pattern and there is a pattern that varies in certain sub-intervals. This change is because each region uses different characteristics. Based on the results of model suitability hypothesis test it is obtained the best model used is NGWR-TS model. Furthermore, the weighting function used for the modeling is the Gaussian kernel function presented in Equation (8) and the Bisquare Kernel Function in Equation (9). The results of the study showed that the appropriate TPT model used was a multivariable NGWR-TS model with the optimum one-knot composition shown in Equation (12) and the best weights used were

Gaussian kernel function weighting because it has the smallest CV value compared to Bisquare Kernel Function. Furthermore, the factors that significantly influence based on partial statistical test results on TPT for each location can be seen in Table 7 and Figure 14 and Modeling results in a coefficient of determination of 80.42%.

Acknowledgment

The author acknowledge The Kementrian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia (Kemenristekdikti RI) Grants and program of improve the quality of scientific (Sandwich-Like 2017) work for generously supporting this project.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] BPS Jawa Timur, Kalimantan Timur in Figures. Badan Pusat Statistika Provinsi Jawa Timur, 2018.

[2] BPS Jawa Timur, Produk Domestik Regional Bruto Kabupaten/Kota Menurut Lapangan Usaha. Badan Pusat Statistika Provinsi Jawa Timur, 2018.

[3] BPS Jawa Timur, Keadaan Angkatan Kerja di Jawa Timur. Badan Pusat Statistika Provinsi Jawa Timur, 2018.

[4] BPS Indonesia, Keadaan Angkatan Kerja di Indonesia 2016. Badan Pusat Statistika Indonesia, 2018.

[5] Fotheringham, A.S, Brundson, C and Charlton, M., Geographically Weighted Regression : The Analysis of Spatially Varying Relationships, John Wiley & Sons Ltd, England, 2002.

[6] Sifriyani, Haryatmi, I.N Budiantara, and Gunardi., "Geographically Weighted Regression with Spline Approach". Far East Journal of Mathematical Sciences, 101(6), 1183-1196, 2017.

[7] Sifriyani, S. H. Kartiko, I. N. Budiantara and Gunardi, "Development Of Nonparametric Geographically Weighted Regression Using Truncated Spline Approach", Songklanakarin Journal of Science And Technology, 40(4), 909-920, 2018.

[8] Sifriyani, I.N. Budiantara, S.H. Kartiko and Gunardi, "A New Method of Hypothesis Test for Truncated Spline Nonparametric Regression Influenced by Spatial Heterogeneity and Application", Abstract and Applied Analysis. 2018, https://doi.org/10.1155/2018/9769150.

[9] Sifriyani, "Multivariable Nonparametric Regression Truncated Spline in The Geographically Weighted Regression Models", Ph.D, Universitas Gadjah Mada, Dept of Mathematics, Indonesia, 2018.

Evaluation of Factors Affecting Increased Unemployment in East Java Using NGWR-TS Method

ORIGINALITY REPORT

9%	5%	4%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ Submitted to Udayana University

Student Paper

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Samarinda, Oktober 2019

Mengetahui,
Ketua Tim Validasi Artikel Ilmiah
FMIPA Unmul,

Dr. Sri Wahyuningsih, M.Si
NIP. 19690413 200012 2 001