International Journal of Advanced Research and Publications
ISSN: 2456-9992

Poverty Data Analysis In Riau Province Using Geographically Weighted Regression Model With Exponential And Tricube Adaptive Kernels

Mia Amiati, Arisman Adnan, Rado Yendra

University of Riau, Department of Mathematics,
Pekanbaru, Indonesia, 28293
mia_amiati13@yahoo.com

University of Riau, Department of Mathematics,
Pekanbaru, Indonesia, 28293

State Islamic University, Department of Mathematics,
Pekanbaru, Indonesia, 28293

Abstract: one of the serious problems in Riau Province is the poverty rate. To overcome the problem the government has made various efforts, in the hope that poverty alleviation will become more directed, but these efforts have do not get effective results so a new aproach is needed to look specifically at poverty cases in each location. The regression analysis approach has often been used in predicting poverty rates, but still global and enforced at all observed locations without involving geographical location based on earth's longitude and latitude. One model could accommodate geographical effects on data is the Geographically Weighted Regression (GWR) model. The data used in this study is the poverty in Riau Province ($\boldsymbol{y}$) and three independent factors ($\boldsymbol{x}$) which will be modeled using the GWR model. The parameters of the model are calculated at each location, so it observation location has a local regression parameter value. The method for estimating the parameters of the GWR model is the Weighted Least Square (WLS) method. The weighti functions used are exponential and tricube adaptive kernels. The selection of optimum bandwidth use the Cross Validation (CV) method. The best selection criteria used are $\boldsymbol{R^2}$, AIC and RMSE. The study show the GWR model with tricube adaptive kernel weighting function is better than the GWR model with an exponential adaptive kernel.

Keywords: Adaptive, Eksponential, Geographically Weighted Regression, Riau Poverty, Tricube.

1. Introduction

Geographically Weighted Regression (GWR) modeling has recently been widely applied in various fields, with the aim of exploring a problem or case in the form of data involving spatial references. Spatial reference is information about the state of an area on earth which involves a spatial analysis approach. Spatial analysis is a technique used in data processing that includes geographical location. Geographic location in general is the location of an area seen from reality on earth or the position of that region on the earth with a size. The size of the location can be the location of an area, if this is involved in a data then the type of data is called spatial data. Spatial data is geographically oriented data that has a certain coordinate system as the basis of its reference such as location information relating to a coordinate (latitude and longitude). Data in spatial form are completed by a method that can accommodate its spatial characteristics, like the proposed GWR model by Fotheringham et al [9]. The GWR model is a development of the global regression model into a local regression model that pays attention to the spatial effect on the data. The GWR model produces parameter estimators that are local to each location point by providing a weighting function. The weighting function in the GWR model is a different parameter estimator at each observation location point in the data. One example of implementing GWR is GWR Modeling Using the Gaussian Kernel Weighting Function [6]. The GWR model is also widely applied to regional problems or cases, such as poverty. Sinaga [17] has discussed Poverty Data Modeling in North Sumatra Province Using the GWR Model, and Deller has discussed the use of GWR in Tourism and Poverty data [5]. Poverty is a different case in each region. The definition of poverty according to CSA is a condition of the economic inability to meet the basic needs of food and non food measured from the expenditure side [2]. Poverty can be influenced by many factors. The level of poverty of an area and the factors that influence it may be different in each region [8]. The factors that influence the poverty used in this study are the percentage of open unemployment rate [16], percentage of human development index (HDI) [18] and percentage of population growth [13].

2. Materials and Methods

2.1 Objectives of the Study

This study discusses the GWR model on poverty data in Riau Province by involving several influencing factors. The data in this study secondary data from CSA of Riau Province in 2017, with the regions studied were 12 districts / cities in Riau Province. The response variable or dependent variable in this study is the percentage of poverty in each district / city. While the explanatory variables or independent variables in this study are the factors that have been raised by economists in measuring the poverty obtained from the Central Statistics Agency [2], as shown in Table 1.
.

Table 1:*Research Variables*

No	*Variable*	*Unit*
1	*Poverty* (y)	*Percentage (%)*
2	*HDI* (x_1)	*Percentage (%)*
3	*Unemployment* (x_2)	*Percentage (%)*
4	*Population growth* (x_3)	*Percentage (%)*

5	*Latitude* (u_i)	*Km*
6	*Longitude* (v_i)	*Km*

Areas or research points in Riau Province can be seen in the picture

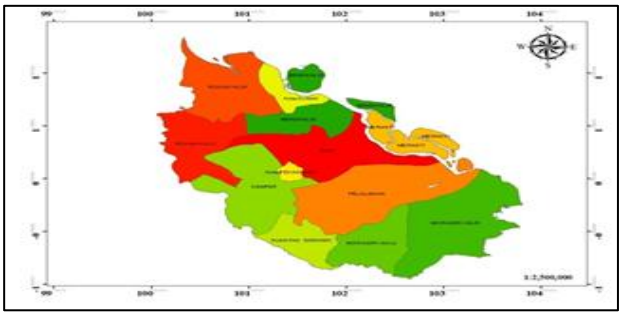

Figure 1:*Riau Province Region*

2.1 GeographicallyWeighted Regression (GWR)

The Geographically Weighted Regression (GWR) model is a statistical method that extends the linear regression framework into a location regression that estimates parameters locally [8]. The GWR model is a spatial regression model based on the location point approach. The point of location of an observation is expressed by geographical coordinates (latitude and longitude). In estimating parameters at a location, all observations are given different weights. The model can be stated as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (1)$$

Where y_i is the dependent variable, x_{ik} is the independent variable, p the number of independent variables, ε is the residual of the model, and β_k are the regression coefficients which are functions of observation (u_i, v_i). GWR is a weighted regression so that the estimated parameters in the GWR model are the WLS (Weighted Least Square) method, namely by providing a different weighting for each location. So the estimated parameter β at point $i$formulated as follows:

$$\hat{\beta}(u_i, v_i) = (X^T(u_i, v_i)X)^{-1}X^T W(u_i, v_i)Y \quad (2)$$

$W(u_i, v_i)$ is a diagonal sized matrix $n \times n$.

2.2 Spatial Weighting Functions

Spatial weighting is a matrix (w_i) that describes the relationship between regions or locations. The weighting calculated for a location depends on the distance between one region and another [20]. The purpose of using a weighting matrix is to determine or estimate different parameters at each point of observation. In the GWR model, the coefficient estimation is performed using WLS, which is the least squares method by giving different weight values at each observation location. The weighting is a diagonal matrix in which the diagonal elements are a weighting function from the point of observation. In this study two weighting functions are used in the form of an Adaptive Kernel. Adaptive Kernel weighting function is a weighting function that has a different bandwidth for each observation location. The two Adaptive kernel weighting functions used in this study are Exponential and Tricube Adaptive Kernel.

1. Eksponensial Adaptive Kernel [6]

$$w_j(u_i, v_i) = \begin{cases} \exp\left(\frac{-d_{ij}}{h_i}\right) & ; d_{ij} \le h_i \\ 0; \ d_{ij} > h_i \end{cases} \quad (3)$$

2. Tricube Adaptive Kernel [15]

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i}\right)^3\right)^3 & ; d_{ij} \le h_i \\ 0; \ d_{ij} > h_i \end{cases} \quad (4)$$

2.3 Bandwidth

Bandwidth is the radius of a circle so that the point of observation within a circle's radius is considered influential in forming parameters at the i. location. Therefore, observations arelocated within radius h is still considered to affect the model at that location so that it will be given a weight that will depend on the function used. According to Fotheringham et al [8], choosing the optimum bandwidth size is one of the important things because it will affect the accuracy of the regression results. There are several methods used to choose the optimum bandwidth, one of which is using a Cross Validation (CV) and can be stated as:

$$CV = \sum_{i=1}^{n}(y_i - \hat{y}_{\neq i}(h))^2 \quad (5)$$

2.4 Research procedure

1. Test spatial heterogeneity with the Pagan Breusch test.
2. Analyzing the GWR model with the following steps:
   a. Calculates the Eucliden's distance between the i-th location located at coordinates (u_i, v_i) and the j-location located at coordinates (u_i, v_i)
   b. Determine the optimum Adaptive bandwidth using the Cross Validation (CV) method
   c. Calculates each Adaptive kernel weighting matrix
   d. Get the parameter estimator of the GWR model
   e. Hypothesis testing with t test
   f. Local Multicollinearity Testing
   g. Selection of the best model

3. Make conclusions.

2.5 Data Implementation Results

GWR model with the exponential kernel Adaptive and Adaptive Tricube weighting in Pekanbaru City:

$$\hat{y}_{Pekanbaru} = 32.87 - 0.41x_1 + 0.21x_2 + 3.83x_3$$

$$\hat{y}_{Pekanbaru} = 33.53 - 0.45x_1 + 0.50x_2 + 4.53x_3$$

With significant variables in each region can be seen in the picture:

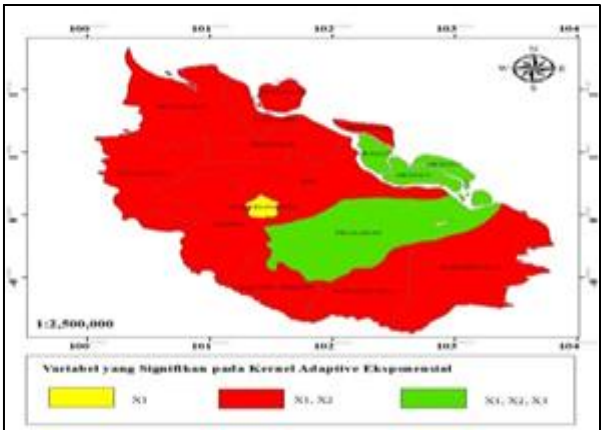

Figure 2*: Exponential Adaptive Kernel*

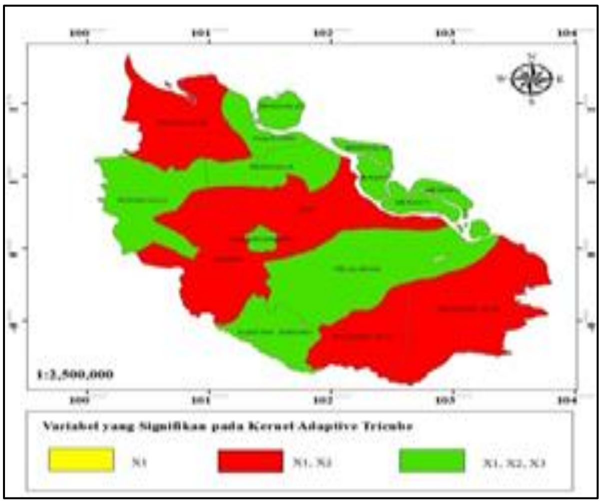

Figure 3:*Adaptive Tricube kernel*

2.6 Criteria for Selecting the Best Model

The model selection criteria are used to get the best model from the two weighting functions used. Model selection criteria are based on tests R^2, Akaike Information Criterion (AIC) and Root Mean Square Error (RMSE). The values for each model can be seen in table 2:

Table 2:*Value R^2, AIC, RMSE Model GWR*

Model	R^2	*AIC*	*RMSE*
Eksponential Adpt	*0.9114*	*58.3943*	*1.8859*
Tricube Adpt	*0.9965*	*21.3689*	*1.1536*

The best model between two exponential and tricube adaptive kernel weighting functions,can be determined by values R^2, AIC and RMSE. The best model is one that has value $R^2$the highest, the lowest AIC and RMSE values. Value $R^2$the highest is the tricube adaptive kernel weighting function 0.9965 This means that the GWR model with the tricube adaptive kernel weighting function is able to explain that 99.65% the percentage of poor population is influenced by the independent variable and the rest by 0.35% influenced by other variables outside the model. AIC and RMSE values in the tricube adaptive kernel weighting function Table 2 shows that the modeling of poverty cases in Riau Province using the tricube adaptive kernel weighting function gives better results than the exponential adaptive kernel weighting function.

3. Conclusion

From the results of the analysis and discussion of GWR with exponential adaptive kernel weighting and tricube adaptive kernel, it can be concluded that the GWR model with tricube adaptive kernel weighting is the best model for modeling poverty data in Riau Province. Because the GWR model with the tricube adaptive kernel weighting has the smallest AIC and RMSE values of 21.3689 and 1.1536 with the largest R^2 value of 0.9965. Based on the GWR model with the tricube adaptive kernel weighting, the average poverty in each location is influenced by the percentage of Human Development Index (HDI) and Open Percentage Unemployment Rate. This GWR model is appropriate for the poverty because it can explain the factors that influence the poverty in each location.

4 References

[1] L. Anselin, “Lagrange Multiplier Test for Spatial Dependence and Spatial Heterogeneity”, Journal of Geographically Analysis, 20 (1) , pp. 1–17, 1988.
[2] Central Statistics Agency, Riau in Figures 2018, CSA of Riau Province, 2017.
[3] C. Brundson, A. S. Fotheringham and M. E. Charlton, Geographically Weighted Regression: a Method Exploring Spatial Nonstationarity, Geographically Analysis, 28 (4) (1996), 281–298.
[4] W. S. Cleveland, “Robust Locally Weighted Regression and Smoothing Scatterplots”, Journal of the American Statistical Association, 74 (368), pp. 829–836, 1979.
[5] S. Deller, “Rural Poverty, Tourism and Spatial heterogeneity”, Annals of Tourism Research, 37 (1), pp. 180–205, 2010.
[6] M. F. Dziauddin and Z. Idris, “Use of Geographically Weighted Regression (GWR) Method to Estimate the a Effect of Location Attributes on the Resindential Property Values”, Indonesian Journal of Geography, 49 (1), pp. 97-110, 2017.
[7] N. Edayu and Syerrina, “A Statistical Analysis for Geographical Weighted Regression”, IOP Conference Series: Earth and Environmental Science, pp. 1–9. 2018.
[8] J. Foster, J. Greer and E. Thorbecke, “A Class Decomposable Poverty Measures”, The Econometric Society, 52 (3), pp. 761–766, 1984.
[9] S. Fotheringham, C. Brundson, and M. E. Charlton “Geographically Weighted Regression”, American Journal of Agricultural Economics, 28 (4), pp. 281–297, 2002.
[10] Graif and R. J. Sampson, “Spatial Heterogeneity in Effects of Immigration and Diversity on Neighborhood Homicide rates”,Homicide Stud, 13, pp. 242-260, 2009.
[11] R. Haining, Spatial Data Analysis: Theory and Practice, Cambridge, 2003.
[12] M. Helbich, W. Brunauer, E. Vaz and P. Nijkamp, “Spatial Heterogeneity in Hedonic House Price Model: the Case of Austria”, Tinbergen Institute Discussion Paper, 51, pp. 390–411, 2014.

[13] M.L. Jhingan, The Economics of Development and Planning, VrindaPublications (P) Ltd, Delhi, 2010.
[14] J. P. Lesage, "A family of Geographically Weighted Regression", Department of Economics University of Toledo, pp. 1–36, 2001.
[15] Levitan, Program in Aid of the Poor for the 1980's, Policy Studies In Employment And Welfare, The Johns Hopkins University Press, 1980.
[16] P. A. Samuelson and W. Nordhaus, Economics, Fourteen Edition, 1993.
[17] K. P. Sinaga, "Poverty Data Modeling in North Sumatera Province Using Geographically Weighted Regression (GWR) Method", International Journal of Science and Research (IJSR), 4 (2), pp. 1738–1742, 2005.
[18] M.P.Todaro and S.C. Smith, Economic Development, Twelfth Edition, Pearson plc, London, 2015.