Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 6, No. 2, Desember 2022
pp. 180-187

Analysis of Factors Influencing Drug Abuse Cases Using Models *Geographically Weighted Regression* (GWR) in Indonesia

Endah Nurfebriyanti[1], Hamidah Nasution[2], Hendra Cipta[3]
[1,3]Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[2]Department of Mathematics, Universitas Negeri Medan, Indonesia

Article Info

Article history:

Received mm dd, yyyy
Revised mm dd, yyyy
Accepted mm dd, yyyy

Keywords:

Drug Abuse
Spatial Heterogeneity
Geographically Weighted Regression
Gaussian Kernel

ABSTRACT

The number of drug abuse cases in Indonesia from 2015-2020 tends to fluctuate. Factors influencing drug abuse cases in each region are thought to vary according to geographical location. This geographic effect is known as spatial heterogeneity. Spatial heterogeneity was analyzed usingGeographically Weighted Regression(GWR). This study aims to model the factors that influence drug abuse in every province in Indonesia, namely Economic Situation (X_1), Association/Environment (X_2), Convenience (X_3) and Lack of Supervision (X_4) using a Gaussian kernel. The results showed that the GWR with the Gaussian kernel is better at estimating the model because it has a higher value, namely R^2 with 90.94% and the AIC value equals 598.798979. Factors that significantly affect the number of cases of drug abuse in Indonesia are Economic Conditions, Association/Environment, Convenience and Lack of Supervision.

Corresponding Author:

Endah Nurfebriyanti,
Department of Mathematics,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: endahnurfebriyanti@gmail.com

1. INTRODUCTION

Regression analysis is an analysis that functions to obtain the relationship of a variable. Regression modeling can be implemented in various fields which include social, economic, health and environmental fields. Often spatial data is found in these fields. Spatial data is a data type that contains information in the form of attributes and locations that are interdependent on one another. Spatial data has different characteristics at each observation point but has a very close relationship to observation points that are close together [9].

The classical regression method is used by researchers to see the relationship between drug abuse cases and the variables that influence them. However, this method does not show the level of drug abusers regionally. The statistical technique used to determine the dependent variable model with independent variables based on area or area is Geographically Weighted Regression [4]. Compared to modeling with the classical regression method, Geographically Weighted Regression can provide regional or regional models and can determine the factors of the variables that influence the spatial aspect [15].

*Geographically Weighted Regression*is the development of multiple linear regression and a spatial model that contains geographic information in the form of coordinate points, namely longitude and latitude which have geographic weights [8]. According to data from the Police Headquarters, the number of drug-

related crimes in Indonesia in 2015-2020 tends to increase. In 2015 there were 62694 cases, in 2016 there were 67250 cases, in 2017 there were 73168 cases, in 2018 there were 97835 cases. Whereas in 2019 there were 113647 cases of drug abuse and the peak in 2020 was 128716 cases [5].

Based on the records owned by BNN in 2020 North Sumatra occupies the first position with more than 1 million drug abusers [8]. Even though Indonesia is being hit by the Covid-19 pandemic, drug trafficking still exists.With the increasing number of drug abuse in Indonesia, we need a way to prevent this problem.

2. RESEARCH METHODS

The procedures contained in this study function to find out what factors can influence drug abuse cases in Indonesia using the Geographically Weighted Regression (GWR) analysis method, which are as follows:

1. Collect data collection from each source
2. Conduct descriptive analysis on the data to find out the initial picture of drug abuse in Indonesia
3. Perform data assumption testing
4. Analyzing data using the basic approach of the Regression model
5. Analyzing data using the Geographically Weighted Regression (GWR) model approach generated through the GWR4 software application
6. Comparing the results of the Regression model with the Geographically Weighted Regression (GWR) model
7. Make a thematic map of the spread of drug abuse in Indonesia based on the estimation results from the Regression model, the Geographically Weighted Regression (GWR) model
8. Conclusion

3. RESULT AND ANALYSIS

Types of independent variables and dependent variables that will be used in this study. The dependent variable is often referred to as the dependent variable, in this study the dependent variable is the number of cases of drug abuse in Indonesia in 2020. The independent variable is often called the independent variable, where this variable will affect other variables, in this study the independent variable is symbolized by the letter X and there are 4 independent variables that are thought to affect the dependent variable[1].

The variables used in this study are:

Table 1. Table of Research Variables

No.	Variable	Variable Definitions
1	Drug Abuse (Y)	The number of cases of drug abuse in Indonesia in 2020
2	Economic Situation ()X_1	This variable defines the number of poor people in Indonesia in 2020
2	Association/Environment ()X_2	This variable defines the number of dropouts that occurred in Indonesia in 2020
3	convenience()X_3	This variable defines the number of location points in drug-prone areas in Indonesia in 2020
4	Lack of supervision ()X_4	This variable defines the number of broken families in Indonesia in 2020

Descriptive Analysis

Descriptive statistical analysis is needed to be able to see an overview of drug abuse cases in Indonesia in 2020. The following are the results of descriptive statistics that can be seen through the size of the concentration and distribution of data obtained through the SPSS program.

Table 2. Table of Descriptive Statistics

	N	*Minimum*	*Maximum*	*mean*	*std. Deviation*	*variances*
Drug Abuse ()*Y*	33	73	20261	3900.48	4595,930	21122575,758
Economic Situation ()X_1	33	36.37	4419.10	797.6442	1117.55974	12448939.772
Association / Environment ()X_2	33	221	6030	1791.79	1474,522	2174214.797
convenience()X_3	33	2	117	19.00	23,373	546,313
Lack of supervision ()X_4	33	0	65755	8838.70	15801.384	249683735.218
Valid N (listwise)	33					

Through Table 2 it can be seen the minimum, maximum, average, standard deviation and coefficient of variance values. The average case of drug abuse in Indonesia in 2020 is quite high, namely 3900 cases. North Sumatra Province is the province with the highest rate of drug abuse, namely 20,261 cases. Meanwhile, East Nusa Tenggara Province is the province with the lowest level of drug abuse, namely 73 cases.

Multicollinearity Test

Multicollinearity testing was carried out to ensure that there is a linear relationship between the independent variables and the dependent variable in the regression model. The relationship between variables can be seen through VIF and tolerance values. The VIF (Varian Inflating Factor) value is approximately 1 and does not exceed 10. The tolerance value is close to 1 with a tolerance$\frac{1}{VIF}$[10].

Table 3.table*of Collinearity Statistics*

Variables	*tolerance*	*VIF*
Economic Situation ()X_1	0.110	9.109
Association / Environment ()X_2	0.252	3,974
convenience()X_3	0.753	1,328
Lack of supervision ()X_4	0.195	5.136

Table 3 above shows that the VIF value range for each variable is still between 1 to 10 and the tolerance value is close to 1, so it can be concluded that there are no multicollinearity problems in the model.

Linear Regression Analysis Model

The linear regression analysis aims to identify the variables that influence data on drug abuse in Indonesia in 2020 without paying attention to their spatial impact. Based on the results of simultaneous testing with a confidence level of 95% it is known that the Economic Condition variable (X_1),Association/Environment (X_2),convenience (X_3), AndLack of supervision (X_4) effect simultaneously (together) on cases of drug abuse (). Therefore, based on Table 4, the linear regression model for drug abuse cases in Indonesia in 2020 is:*Y*

$$Y = -576.720 - 2.892X_1 + 1.827X_2 + 99.746X_3 + 0.183X_4$$

To obtain a linear regression model from Table 4, the estimated results of the linear regression model for drug abuse in Indonesia in 2020. In addition, the results in Table 4 are compared with a significance level of 5%.$t_{hitung} t_{tabel}$

Table 4. Parameter Estimation Results of Linear Regression Model

Variables	*Estimates*	*SE*	t_{hitung}
Intercepts	-576,720	937,939	-0.615
Economic Situation ()X_1	-2,892	2,289	-1,263
Association / Environment ()X_2	1,827	0.813	2,249
convenience()X_3	99,746	24,062	4.145
Lack of supervision ()X_4	0.183	0.122	1,492

The regression model obtained shows that cases of drug abuse in Indonesia can increase with high association/environment () of 1.827% provided that other variables are constant. Cases of drug abuse in

Indonesia can also increase with the high ease of obtaining drugs by 99.746% provided that other variables are constant.X_2X_3

Simultaneous Test

Simultaneous test or F test is used as a determinant of all independent variables which together affect the dependent variable. Below are the results of testing simultaneously with SPSS 26 software.

Table 5. Simultaneous Test

Model	*Sum of Squares*	*df*	*R Square*	*F*
Regression	422821100718	4	0.626	11,694
residual	253101323,524	28		
Total	675922424242	32		

By using a confidence level of 95%, the results of the analysis show that there is at least one independent variable that influences the decision. This means that the regression model used is appropriate for describing the variables*dependent*which has a relationship to the variable*independent*.The value of the correlation coefficient () is equal to, which means that the independent variable explaining drug abuse cases is equal to . The rest is explained by other variables outside the model.$R^2$62.6%62.6%37.4%

Partial Test

After knowing that the independent variables can influence drug abuse cases in Indonesia in 2020 with a simultaneous test, then a partial test is then used to test the effect of each independent variable [2] on drug abuse cases in Indonesia in 2020. The following are the results of the test Partial:

Table 6. Partial Tests

Variables	t_{hitung}	*Significant*
Intercepts	-0.615	0.544
Economic Situation ()X_1	-1,263	0.217
Association / Environment ()X_2	2,249	0.033
convenience()X_3	4.145	0
Lack of supervision ()X_4	1,492	0.147

From the results and in the table above, it can be concluded that there is an influence of independent variables influencing drug abuse cases in Indonesia in 2020.$T_{hitung}p - value$

Geographically Weighted Regression (GWR) Model Analysis

The information used in the GWR model is a spatial model for each geographic region. Knowing geographic information in the form of longitude and latitude points in each province in Indonesia is the first step to get a GWR model with GWR4 software. After knowing the geographic information, the next step is to determine the optimum bandwidth by using the Cross Validation (CV) method. This aims to provide coverage to an area with other areas that are still related to each other. To get optimum bandwidth, you can use GWR4 software. In this study, fixed Gaussian kernel weights were used to determine the weight matrix [6]. To determine the results, you can use the GWR4 software or manually.

Bandwidth Determination

The first thing that must be done is to determine the geographical position by knowing the longitude and latitude coordinate points in each province, namely by determining the Euclidean distance [11]. Furthermore, with Cross Validation (CV) to obtain the optimum bandwidth value, shown in Table 7 below:

Table 7. Optimum Weight

Function weights	Optimum Bandwidth	Minimum CV
Fixed Gaussian	5,525	7484337.451

The Optimum Bandwidth value is 5,525 and*Cross Validation (CV)*the minimum is 7484337.451 The function of the Fixed Gaussian weighting in this study is a function that will describe for all observation locations will have the same bandwidth value. The bandwidth value will be used for statistical calculations in equation 2.21 which is symbolized as h.

$$w_{ij}(u_i, v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{11}}{h}\right)^2\right] become\, w_{ij}(u_i, v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{11}}{5.525}\right)^2\right]$$

Geographically Weighted Regression (GWR) Model Weighting

To determine the model weighting*Geographically Weighted Regression*Gaussian Karnel function weighting is required. With a minimum value of Cross Validation (CV) of 7484337.451, the weighting value of the Gaussian Karnel function produces a bandwidth value of 5.525. After getting the results of the most optimum bandwidth value using the Gaussian Karnel function, the weighting matrix will be determined from the resulting matrix for each location to form a model, so that it will be different at each location [14].

$$d_{11} = \sqrt{(u_1 - u_1)^2 + (v_1 - v_1)^2}$$
$$= \sqrt{(96.7494 - 96.7494)^2 + (4.695135 - 4.695135)^2}$$
$$= 0$$
$$d_{12} = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}$$
$$= \sqrt{(96.7494 - 99.5451)^2 + (4.695135 - 2.115355)^2}$$
$$= \sqrt{(-2.7957)^2 + (2.57978)^2}$$
$$= \sqrt{7.81593849 + 6.655264848}$$
$$= \sqrt{14.47120334}$$
$$= 3.804103487$$
…
$$d_{1n} = \sqrt{(u_1 - u_n)^2 + (v_1 - v_n)^2}$$
$$= \sqrt{(96.7494 - 138.0804)^2 + (4.695135 - (-4.26993))^2}$$
$$= \sqrt{(-41.331)^2 + (8.965065)^2}$$
$$= \sqrt{1708.251561 + 80.37239045}$$
$$= \sqrt{1788.623951}$$
$$= 42.29212635$$

After knowing the value of the Euclidean distance at the first location, then using the Gaussian Karnel function to get the weight matrix at the first location (). With the value of the bandwidth used for$W_1$5,525, the weight matrix is obtained at location 1 as follows.

$$w_{11} = exp\left[-\frac{1}{2}\left(\frac{d_{11}}{h}\right)^2\right] = exp\left[-\frac{1}{2}\left(\frac{0}{5.525}\right)^2\right] = 1$$

$$w_{12} = exp\left[-\frac{1}{2}\left(\frac{d_{12}}{h}\right)^2\right] = exp\left[-\frac{1}{2}\left(\frac{3.804103487}{5.525}\right)^2\right] = 0.788964704$$

…

$$w_{1n} = exp\left[-\frac{1}{2}\left(\frac{d_{1n}}{h}\right)^2\right] = exp\left[-\frac{1}{2}\left(\frac{42.29212635}{5.525}\right)^2\right] = 1.88989$$

In matrix form as follows

$$W_1 = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{12} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{1n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0.788964704 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1.88989 \end{bmatrix}$$

The following is the result of the weighting matrix usingGaussian Karnel function at each observation location.

$$W = \begin{bmatrix} 1 & 0.78896704 & 0.471138221 & \cdots & 1.88989 \\ 0.78896704 & 1 & 0.852712292 & \cdots & 1.40113 \\ 0.471138221 & 0.852712292 & 1 & \cdots & 1.05845 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1.88989 & 1.40113 & 1.05845 & \cdots & 1 \end{bmatrix}$$

Parameter estimation of GWR is determined by a weighting matrix for each observation. The results of parameter estimation with the GWR model can be seen using the GWR4 software.

GWR Model Suitability Testing

To test the suitability of the model in this study an F or Goodness of Fit test will be carried out to be able to determine the effect of weighting on parameter estimation [7]. The results of the analysis obtained through the GWR4 software are shown in Table 8 below:

Table 8. Geographically Weighted Regression Model Suitability (GWR) Test

Source	SS	*DF*	*Ms*	*F*
Global Residuals	253101323,524	28,000		
GWR Improvement	191904180.249	11,098	17291566,474	
GWR Residuals	61197143.275	16,902	3620733,682	4.775708

Through Table 8 above, it can be seen that the rejection model in the GWR model suitability test by making a decision to reject if the value is at a confidence level of 95%. With that is equal to 2.701 and that is equal to 4.775708.H_0 $F_{hitung} > F_{tabel} F_{tabel} F_{hitung}$

GWR Model Significance Variable Mapping

Mapping provinces in Indonesia are grouped based on variables that significantly influence data on drug abuse cases with a confidence level of 95% as shown in Figure 1, there are 8 groups based on variables that significantly influence drug abuse cases. The mapping of drug abuse case data for each province in Indonesia from the local model to the GWR model using ArcMap 10.4.1 software is as follows:

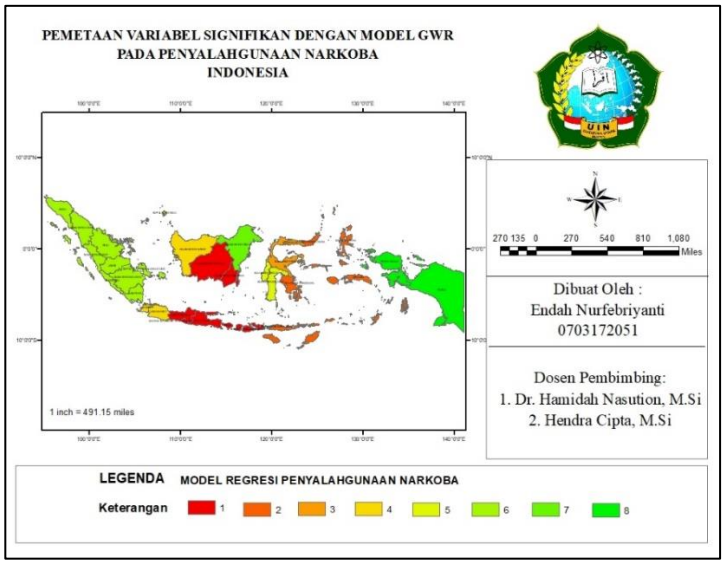

Figure 1. Mapping of Drug Abuse Cases with the Local GWR Model

The model produced with the Geographically Weighted Regression model for each province is different. As an example, the GWR model for North Sumatra Province is, but after doing a comparison with the GWR model, it is significantly relevant. $\hat{y}_{Sumut} = -2270.14 - 7.64269X_1 + 6.16875X_2 + 72.88953X_3 + 0.158625X_4 t_{hitung} t_{tabel} \hat{y}_{Sumut} = -2270.14 - 7.64269X_1 + 6.16875X_2 + 72.88953X_3$

4. CONCLUSION

With the Geographically Weighted Regression (GWR) model it is proven to be able to show the factors that influence drug abuse cases in Indonesia in a spatial aspect. This is evidenced by the magnitude of the valueR^2on models*Geographically Weighted Regression*(GWR)of 0.909461 or equal to 90.9461%. Compared to the global regression model which has a value of 0.625547 or equal to 62.5547%. With so on the modelR^2Geographically Weighted Regression (GWR)factorEconomic Condition, Association/Environment, Convenience and Lack of Supervision simultaneously (together) influence drug abuse cases in Indonesia by 90.9461%. While the rest is influenced by other factors.A model is also said to be good if it has value*Classic AIC*lower. The Geographically Weighted Regression (GWR) model is also proven to have a Classic AIC valueof 598.798979 where the value is smaller awhen compared with the global regression which has a Classic AIC value of 628.792109. So that the Geographically Weighted Regression (GWR) model is considered the best model for modeling Drug Abuse Cases in Indonesia.

As for suggestions that can be used for further research, namely the addition of independent variables to determine the effect on drug abuse inIndonesia.

Reference

[1] Anshori, Muslich. 2017. Quantitative Research Methodology. Surabaya: Airlangga University Press
[2] Ardianti, D., Pramoedyo, H., & Nurjannah, N. 2021. Distance weight of GWR-Kriging model for stunting cases in East Java. In Journal of Physics: Conference Series (Vol. 1968, No. 1, p. 012028). IOP Publishing.
[3] Ardianti, D., Pramoedyo, H., & Nurjannah, N. 2021. Distance weight of GWR-Kriging model for stunting cases in East Java. In Journal of Physics: Conference Series (Vol. 1968, No. 1, p. 012028). IOP Publishing.
[4] Cakra, Rezzy Eko and Hasbi Yasin. 2017. Geographically Weighted Regression (GWR) A Geographical Regression Approach. Yogyakarta: Mobius
[5] Central Bureau of Statistics. 202 0 . Criminal Statistics 2020 . Jakarta : BPS RI
[6] Edayu, ZN, & Syerrina, Z. 2018. A statistical analysis for geographical weighted regression. In IOP Conference Series: Earth and Environmental Science (Vol. 169, No. 1, p. 012105)
[7] Feng, X. 2017. Research on spatial correlation between air quality and land use based on GWR Models. Nature Environment and Pollution Technology, 16(1), 155.

[8] Fotheringham, A. Stewart, Chris Brunsdon and Martin Chalton. 2002. Geographically Weighted Regression The Analysis of Spatial Variation Relationships. USA : John Wiley
[9] Inayah, Ulfa rest. 2020. Geographically Weighted Logistic Regression Model with Function Bisquare Adaptive Weighting . Samarinda : Mulawarman University
[10] Li, Yange, et al. 2020. Spatial proximity-based geographically weighted regression model for landslide susceptibility assessment: a case study of Qingchuan area. China : Applied Sciences 10.3 : 1107.
[11] Liu, J., Yang, Y., Xu, S., Zhao, Y., Wang, Y., & Zhang, F. 2016. A geographically temporally weighted regression approach with travel distance for house price estimation. Entropy, 18(8), 303.
[12] National Narcotics Agency. 2020. Indonesia Drugs Report 2020. Jakarta: BNN
[13] Ristea, A., Kounadi, O., & Leitner, M. 2018. Geosocial Media Data as Predictors in a GWR Application to Forecast Crime Hotspots (Short Paper). In 10th International Conference on Geographic Information Science : Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[14] Soemartojo, SM, Ghaisani, RD, Siswantining, T., Shahab, MR, & Ariyanto, MM 2018. Parameter estimation of geographically weighted regression (GWR) model using weighted least square and its application. In AIP Conference Proceedings (Vol. 2014, No. 1, p. 020081)
[15] Yusuf, Dessy WS, et al. 2020. Geographically Weighted Regression (GWR) Modeling on the Percentage of Crime in East Java Province in 2017. Journal of Statistics. 4(1) : 156-163