Comparison of Geographically Weighted Artificial Neural Network and Geographically Weighted Generalized Poisson Regression on Crime Cases in East Java Indonesia

Dewi Sinta Nur Fajarini [1], Yuliani Setia Dewi [1], Mohamat Fatekurohman [1]

[1] *University of Jember, Jember, Indonesia*

Abstract **– Crime is behavior that violates laws and social norms. One area of Indonesia with a relatively high number of crimes is East Java. Data on the number of crimes in East Java in 2020 showed overdispersion and multicollinearity. This study aims to model the number of crimes reported in East Java by considering the spatial effects in the data. The methods used to analyze the data are GPR, GWANN and GWGPR. Moreover, we also determine the results of comparing the three methods using R^2 and RMSE. The study results show that GWANN provides better results to model the number of reported crimes compared to GWGPR and GPR. The results show that the GWANN model results in eight groups using the three highest value of the variable importance.**

Keywords **– crime, GWANN, GWGPR, variable importance.**

1. Introduction

Indonesia is a country consisting of 34 provinces, one of which is East Java. Astronomically, East Java Province has an area of 47,803.49 km^2.

DOI: 10.18421/SAR71-03
https://doi.org/10.18421/SAR71-03

Corresponding author: Yuliani Setia Dewi,
University of Jember, Indonesia
Email: yulidewi.fmipa@unej.ac.id

Received: 12 February 2024.
Revised: 14 March 2024.
Accepted: 20 March 2024.
Published: 27 March 2024.

It is located between 112° 36′ – 112° 54′ EL and 7°9′- 7°21′ SL, consists of nine cities and 29 regencies with a population of 40.666 million in 2020.

Criminology is an academic discipline that system atically studies making laws, breaking laws, and resp onding to law violations [1], [2], [3]. In a formal juridical sense, crime is a behavior contrary to human morality, harms society, is asocial, and violates the law and criminal law. According to BPS-Statistics, Indonesia, the Polri registration data noted that the number of crimes in Indonesia in 2018–2020 tended to decrease. Data on the number of reported crimes shows that the East Java Regional Police ranks third in Indonesia after the North Sumatra Regional Police and the Metro Jaya Regional Police. The crimes documented in East Java Province amounted to 17,642 incidents spread across 29 districts and nine cities.

The research about panel regression of the factors that influence the amount of crime in East Java [4] finds that economic growth, number of poor people, and population density influence the crime rate in East Java. The research was carried out by [5] regarding the effect of education, income gap, and poverty on crime in Indonesia. Results showed that extreme poverty has a positive impact on crime in Indonesia. Research conducted by [6] on the relationship between unemployment and noneconomic factors with crime rates in Indonesia states that unemployment has no effect on murder, and non-economic factors are assertive in explaining murder and violent crimes.

The spread of crime cases in East Java is different for each regency/city. The number of crimes reported in East Java Province is counted data and can be analyzed using Poisson regression. Poisson regression has an assumption that is the mean and variance are the same, but sometimes the variance is greater than the mean. It is known as the overdispersion case.

Overdispersion cases can be modeled using Generalized Poisson Regression (GPR). Research conducted by [7] concerns modeling the crime rate in East Java province, Indonesia, using the GPR method. This study shows that there are overdispersion symptoms in the data, and the significant variables that related to the crime rate are the percentage of poor people, expenditure, unemployment rate, and number of households. The study did not pay attention to spatial effects, so the pattern of crime cases was not detected in every area in East Java province.

One of the statistical methods that can be used to perform spatial heterogeneity analysis is Geographically Weighted Generalized Poisson Regression (GWGPR). It is an expansion method of GPR. Previous research uses the GWGPR method [8]. This study uses the GWGPR and GWNBR methods to measure the number of postpartum maternal mortality cases. It results that GWGPR with a fixed bisquare kernel has the best performance.

The Poisson models have a good performance, when there is no multicollinearity in the data. Multicollinearity can cause the parameter estimates obtained to be biased. It causes instability in the Poisson model formed [9].

Another method used to analyze spatial heterogeneity is research conducted by Hagenauer and Helbich [10] using the Geographically Weighted Artificial Neural Network method. This research shows that GWANN is better than GWR for nonlinear data and has a high spatial variance. GWANN is a variant of ANN that combines ANN and geographic weighting to model spatial heterogeneity data.

This research investigates the regression methods by considering the existence overdispersion, the spatial heterogeneity, and multicollinearity in the data to find the existence location groups of the number of crime in Esat Java based on some explanatory variables. The techniques used are GPR, GWGPR and GWANN.

2. Method

We use the crime data from the East Java BPS statistics, Indonesia. The response variable (Y) is the number of reported crimes (total crime), and the explanatory variables (X_1-X_5) are described in Table 1. There are 38 locations in East Java, consisting of nine cities and 29 regencies.

Table 1. Research Variables

Variable	Description
Y	Number of Crimes Reported
X_1	Percentage of Poor Population
X_2	Poverty Depth
X_3	Index Sex Ratio
X_4	Average School Years
X_5	Human Development Index

The steps of this research are given as follows:
1. Performing a descriptive analysis;
2. Conducting a multicollinearity test based on the VIF value [11];
3. Doing Poisson regression modeling [12];
4. Checking data for overdispersion [13];
5. Performing spatial heterogeneity testing to model GWGPR and GWANN using the Breusch-Pagan test [14];
6. Performing GWGPR modeling [14] with the following steps;
   a) Testing simultaneously the significance of parameters;
   b) Testing partially the significance of parameters;
   c) Determining the value of R^2 and RMSE [15].
7. Performing GWANN modeling [10], [16], with the following steps;
   a) Defining data in matrix form;
   b) Setting hyperparameters;
   c) Determining the value of R^2 and RMSE
8. Comparing the GWGPR and GWANN models using R^2 [17] and RMSE [18];
9. Concluding.

3. Results

In this section, we give results of the research, explain the characteristics of the data, compare the methods and present more detail the best method.

3.1. Descriptive Statistics

Descriptive statistics contain the average, variance, minimum, and maximum values of the response and explanatory variables (Table 2).

Table 2. Descriptive Statistics of the Data

Variable	Average	Variance	Min	Max
Y	600	187,571.5	70	1,850
X_1	11.021	20.86483	3.890	22.780
X_2	1.692	0.8343144	0.590	4.330
X_3	97.20	6.49394	90.75	101.20
X_4	7.942	2.340792	4.850	11.140
X_5	71.87	25.46523	62.70	82.23

Based on Table 2, the number of reported crimes (Y) has an average value of 600 and a variance value of 187,571.5. The lowest number of crime cases is in Batu City, with 70 crimes, while the highest number was in Malang Regency, with 1,850 crimes. The Figure 1 is a map of the distribution of reported crimes in East Java.

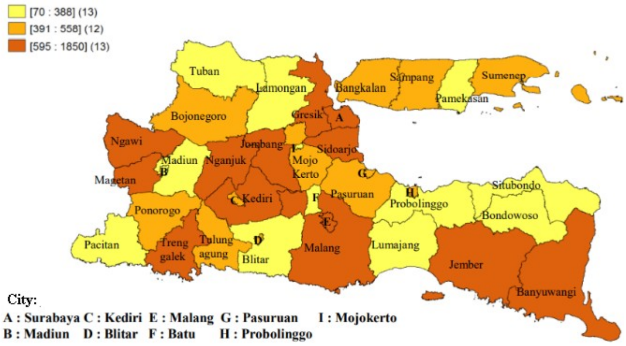

Figure 1. Distribution of the number of reported crimes in East Java

3.2. Multicollinearity Test

We conducted multicollinearity test to find the existence correlation between the explanatory variables. We used variance inflation factor (VIF) to measure the presence of multicollinearity in the data. Table 3 presents the VIF value of the explanatory variables.

Table 3. VIF Values

Predictor	VIF
X_1	12.826694
X_2	8.108023
X_3	1.496214
X_4	25.015707
X_5	25.337702

Based on Table, most of the VIF value exceed 5. It indicates there are multicollinearity cases in the data.

3.3. Overdispersion

In the crime data, the variance is greater than the mean value. Therefore, in this case there is overdispersion in the data. In addition, overdispersion can also be detected by the deviation value divided by the degrees of freedom of Poisson regression, ie.

$$\theta = \frac{deviance}{df} = \frac{6486.3}{32} = 202.69$$

The quotient value is more than one, showing the existence of overdispersion in the data.

3.4. Spatial Heterogeneity Test

The heterogeneity test results using Breusch-Pagan give a value of 12.04 and a p-value of 0.03424.

It means that the BP value is greater than $\chi^2_{(0.05;5)}$, namely 12.04 > 11.07 and the p-value is less than α of 0.05, therefore the decision to reject H_0 or the variances between different locations differ. This result means that there are differences in characteristics at each observation location.

3.5. GPR, GWGPR, and GWANN Models

Generalized Poisson Regression (GPR) is a method to handle the existence overdispersion or underdispersion in the Poisson data. It is a global model and development of Poisson regression. The GPR model ignores spatial effects, so we cannot detect the characteristic differences between locations. In this research, the GPR model results that the average length of school variable (X_4) and the human development index variable (X_5) significantly related to the number of crimes in East Java, Indonesia (Table 4).

Table 4. Estimated Parameters of GPR model

Coefficients	Estimate	Std.Error	Z value
$\hat{\beta}_0$	-12.27901	6.28480	-1.954
$\hat{\beta}_1$	-0.07469	0.07494	-0.997
$\hat{\beta}_2$	0.28330	0.29302	0.967
$\hat{\beta}_3$	0.08911	0.05098	1.747
$\hat{\beta}_4$	-0.76943	0.32019	-2.403*
$\hat{\beta}_5$	0.22812	0.09407	2.425*
Deviance	11.52252		

*Note: * Significant at α = 0.05*

Regression considering spatial effects can provide an overview of distribution patterns for each region resulting in a more accurate model. GWGPR estimates the local parameter of each location. The optimum bandwitdth for this model is 0.9999371 (CV = 7231412) using an adaptive bisquare kernel. By using significance level (α = 0.05), the research on the GWGPR model results in four groups locations (Table 5).

Table 5. The Grouping Locations Based on Significant Variables Using GWGPR

Group	Significant variable	Regency/City
1	X_4	Jember
3	X_5	Batu
4	X_4 and X_5	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun.
4	no significant variable	Other areas (12 locations)

GWANN is the development of an Artificial Neural Network (ANN) that uses geographic weights. As GWGPR, GWANN build the model of each location. Using GWANN method we get eight groups of locations using the three most important variable (Table 7).

3.6. Comparing the Methods

We compare the models by looking at the R^2 and RMSE value. The comparison results can be seen in the Table 6.

Table 6. Best Model Selection

Method	R^2	RMSE
GWANN	0.9291596	113.7452
GPR	0.2765377	363.4967
GWGPR	0.3209243	352.1694

Based on the best model results, we find that GWANN is better than GPR and GWGPR in modeling reported East Java's crime cases. The value of R^2 of the GWANN model is closer to 1 than the the other models, which means that the GWANN model has better results of R^2. The RMSE value is better if the value is smaller, which means that the RMSE value of GWANN is better than the RMSE of GPR and GWGPR.

3.7. GWANN Model

In this research. GWANN model is obtained by training the model using different hidden neurons between 5 and 40. The batch sizes used are 5, 25, 50, and 100. and the three optimizers used are momentum, sgd, and Nesterov. The kernels used in modeling GWANN is adaptive bisquare.

The best model with the smallest RMSE value is obtained when the hidden neuron value is 40, batch size is equal to 5, momentum optimizer and the optimum bandwidth of 3. In these conditions, we get an RMSE value of 113.7452.

The variable importance is used to see whether the research variable has a role in this modeling. Regency/city grouping is related to the three highest values of the variable importance.

Table 7. Location Groups Based on Three Variable Importance of the GWANN model

Important Variable	Regency/City
$X_1X_2X_3$	Mojokerto, Bojonegoro, Sampang
$X_1X_2X_4$	Ngawi, Lamongan, Kota Probolinggo, Batu
$X_1X_2X_5$	Pacitan, Ponorogo, Magetan
$X_1X_3X_5$	Blitar, Sidoarjo, Surabaya
$X_1X_4X_5$	Lumajang, Jember, Jombang, Nganjuk, Pamekasan, Sumenep, Kota Kediri, Kota Blitar
$X_2X_3X_4$	Bangkalan, Kota Madiun
$X_2X_3X_5$	Trenggalek, Kediri, Bondowoso, Situbondo, Probolinggo, Madiun, Tuban, Kota Malang, Kota Pasuruan, Kota Mojokerto
$X_3X_4X_5$	Tulungagung, Malang, Banyuwangi, Pasuruan, Gresik

Based on Table 7, we obtain eight groups using the three most important variables. The grouping of locations can be seen in the following Figure 2.

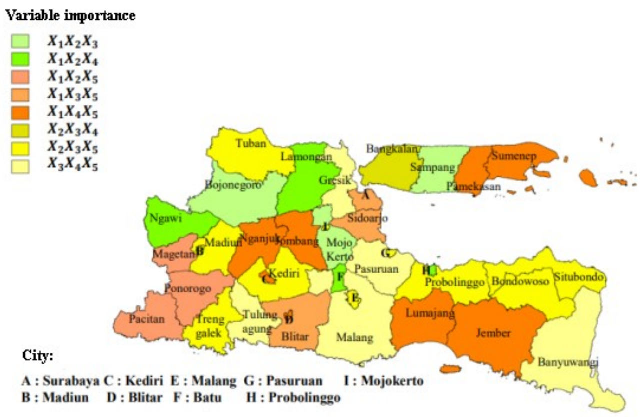

Figure 2. Locations Groups of GWANN

As an example, the following is the value of the variable importance to the number of crimes in Lumajang Regency.

Table 8. Value of the importance variable in Lumajang Regency

Variable	Variable Importance
X_1	126485.4
X_2	126482.6
X_3	126484.0
X_4	126494.6
X_5	126491.0

Table 8 shows that the three most influential variables in Lumajang Regency are the percentage of poor people, the average length of schooling, and the human development index.

4. Conclusion

This research models criminal cases in East Java using the GWANN, GWGPR, and GPR methods on data containing multicollinearity, overdispersion, and spatial heterogeneity. A comparison of the GWANN, GWPR, and GPR methods in modeling the number of reported crimes shows that GWANN is better than GPR and GWPR based on the highest R^2 value and the smallest RMSE. Using the three highest-importance variables. GWANN model gives eight groups of locations.

References:

[1]. Sutherland, E. H., Cressey, D. R., & Luckenbill, D. F. (1992). *Principles of criminology*. Altamira Press.

[2]. Elmas, M. S. (2021). Perceived risk of terrorism, indirect victimization, and individual-level determinants of fear of terrorism. *Security Journal*, *34*(3), 498-524.

[3]. Dooley, B. D. (2021). Establishing a profession through boundary drawing: Defining criminology's autonomy vis-à-vis six competing disciplines. *American journal of criminal justice*, 1-24.

[4]. Hariyantia, F. N., Purtomo, R., & Wilantari, R. N. (2021). The Effect of Economic Growth and Demographic Condition on Criminality in East Java Province. *Jurnal Ekonomi dan Bisnis Airlangga Volume*, *31*(1).

[5]. Sugiharti, L., Purwono, R., Esquivias, M. A., & Rohmawati, H. (2023). The nexus between crime rates, poverty, and income inequality: A case study of Indonesia. *Economies*, *11*(2), 62.

[6]. Sugiharti, L., Esquivias, M. A., Shaari, M. S., Agustin, L., & Rohmawati, H. (2022). Criminality and Income Inequality in Indonesia. *Social Sciences*, *11*(3), 142.

[7]. Handayani, D., Artari, A. F., Safitri, W., Rahayu, W., & Santi, V. M. (2021, November). Count Regression Models for Analyzing Crime Rates in The East Java Province. In *Journal of Physics: Conference Series*, 2123(1), 012028. IOP Publishing.

[8]. Tyas, S. W., & Puspitasari, L. A. (2023). Geographically weighted generalized poisson regression model with the best kernel function in the case of the number of postpartum maternal mortality in east java. *MethodsX*, *10*, 102002.

[9]. Lukman, A. F., Adewuyi, E., Månsson, K., & Kibria, B. G. (2021). A new estimator for the multicollinear Poisson regression model: simulation and application. *Scientific Reports*, *11*(1), 3732.

[10]. Hagenauer, J., & Helbich, M. (2022). A geographically weighted artificial neural network. *International Journal of Geographical Information Science*, *36*(2), 215-235.

[11]. Daoud, J. I. (2017). Multicollinearity and regression analysis. In *Journal of Physics: Conference Series*, *949*(1), 012009). IOP Publishing.

[12]. Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*, *53*. Cambridge university press.

[13]. Sundari, M., & Sihombing, P. R. (2021). PENANGANAN OVERDISPERSI PADA REGRESI POISSON:(Studi Kasus: Pengaruh Faktor Iklim Terhadap Jumlah Penderita Penyakit Demam Berdarah di Kota Bogor). *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *2*(1), 1-9.

[14]. Wulandari, S. P., & Nurul'Aini, M. (2023). Modeling Factors Affecting the Number of Tuberculosis Cases in Indonesia Using Geographically Weighted Generalized Poisson Regression. *Res Militaris*, *13*(2), 5329-5349.

[15]. Almasi, S. A., Behnood, H. R., & Arvin, R. (2021). Pedestrian crash exposure analysis using alternative geographically weighted regression models. *Journal of advanced transportation*, *2021*, 1-13.

[16]. As'adi, N. A., Dewi, Y. S., & Tirta, I. M. (2023). Spatial Heterogeneity Models of Undernourished Toddlers in Java Island, Indonesia. *SAR Journal (2619-9955)*, *6*(2).

[17]. YÜZBAŞI, B., & GÖRÜR, Ç. (2021). EXAMINING SPATIAL VARIABILITY IN THE ASSOCIATION BETWEEN SOME FACTORS AND NUMBER OF PRISONERS IN TURKEY: A GWR ANALYSIS. *Mugla Journal of Science and Technology*, *7*(1), 141-148.

[18]. Sutawinaya, I. P., Astawa, I. N. G. A., & Hariyanti, N. K. D. (2017). Perbandingan Metode Jaringan Saraf Tiruan pada Peramalan Curah Hujan. *Logic: Jurnal Rancang Bangun dan Teknologi*, *17*(2), 92-97.