



Modeling the Number of Foreign Tourist Visits to Indonesia in 2020 Using GWPR Method

Muhammad Zidni Subarkah^{a,1}, Rizki Wahyuningtia^{a,2}, Martina Hildha^{a,3}, Winita Sulandari^{a,4,*}

^a Department of Statistics, Universitas Sebelas Maret, Surakarta, Indonesia

¹ mzidni100701@student.uns.ac.id; ² rizkiwahyuningtia@student.uns.ac.id; ³ mhilda.mh@student.uns.ac.id; ⁴ winita@mipa.uns.ac.id*

* Corresponding author

ARTICLE INFO

ABSTRACT

Keywords
tourists
Poisson
spatial
GWPR
AIC

In December 2020, the number of foreign tourists visiting Indonesia experienced a sharp decline of 88.08% compared to the number of visits in December 2019. However, compared to the previous month, November 2020, this number increased by 13.58%. Modeling the number of foreign tourists visiting Indonesia in 2020 using the Geographically Weighted Poisson Regression (GWPR) method is needed to elaborate on the Indonesian government's policy decisions, especially in the tourism sector. The results showed that the GWPR model with the Kernel fixed Gaussian weighted function had an AIC value of 1,521,240.873, deviance of 1,521,196.695, and deviance-R2 of 0.741 or 74.1%. This model produced two different clusters of characteristics of foreign tourists' country of origin based on the variable's significance. Cluster one consisted of Finland and Qatar and the rest were in cluster two. The characteristics of cluster two were influenced by the rupiah exchange rate variable, short stay visa free (*Bebas Visa Kunjungan Singkat*, BVKS), Consumer Price Index (CPI), economic growth, total imports, and the distance of CGK to the international airport. Meanwhile, cluster one had almost the same characteristics as cluster two but was not influenced by the BVKS factor variables.

1. Introduction

Indonesia is one of the countries with attractive tourist destinations that draw tourists, particularly foreign tourists, to enjoy its appealing attraction. Foreign tourists are those visiting other countries outside their home country, motivated by various needs, without any intentions to earn income in the destination country and their stay does not exceed 12 months [1]. Indonesia was ranked 40 out of 140 countries in the world in the 2019 Travel & Tourism Competitiveness Index (TTCI) [2]. This ranking is certainly quite encouraging for Indonesia, necessitating the elaboration of government policies on Indonesian tourism.

In 2020, the number of foreign tourist visits amounted to 4.05 million [1]. This number has certainly fallen far compared to foreign tourist visits in 2019 which amounted to 16.11 million. This visit, without a doubt, affect Indonesian economy since tourism sector plays a major role in the economy of Indonesia. The number of foreign tourists who come to Indonesia has a positive

relationship with economic growth in Indonesia [3]. If the number of foreign tourists coming to Indonesia is high, then economic growth will also increase. One way to increase the number of foreign tourist visits is by looking at what factors influence tourists from certain country to visit Indonesia. Previous research used the panel data method with the variable Gross Domestic Product (GDP), the number of foreign tourist arrivals, the number of foreign tourists spending, the duration of residence of foreign tourists, the exchange rate, and short stay visa free (Bebas Visa Kunjungan Singkat, BVKS) [3]. This study used a different method, and factors were mostly taken from the country of origin of tourists.

This study took several factors from each country to be tested and seen which factors affected the number of foreign tourists to visit Indonesia. Several independent factors were used, namely the exchange rate of the foreign tourist's country of origin against the rupiah, GDP per capita (US\$), population density per km², BVKS, Consumer Price Index (CPI), life expectancy (years), economic growth (%), imports (US\$), and the distance from Soekarno-Hatta international airport to the international airport of the foreign country of origin (km). While the dependent factor used was the number of foreign tourist visits from their country of origin to Indonesia, where this data follows the Poisson distribution. Poisson regression method is the right regression method used to analyze data containing a Poisson distribution [4]. The application of Poisson regression has been widely published by many researchers, such as [5]–[8].

In addition, Geographically Weighted Poisson Regression (GWPR) is a statistical method developed for the analysis of data with a Poisson distribution by taking into account the spatial factor [5]. Unfortunately, the Poisson regression model assumes that the location of the observations does not affect the model [9]. Therefore, the GWPR method is used to show the regression model from each observation location. GWPR is a statistical method developed from the Poisson regression method. It pays attention to the weights in the form of the location of the observations so that the dependent variable of the GWPR model is influenced by the independent variable, whose regression coefficient is influenced by geographical location [9]

The GWPR method can show the regression model of each area being tested. In this study, the Poisson and GWPR regression methods were used to analyze the factors that influence the number of foreign tourists visiting Indonesia from various countries. The best model was determined by looking at the Akaike Information Criterion (AIC) [10] value generated from each model. The best model was determined based on the smallest AIC value among other model AIC values. This study is expected to show a regression model of the factors influencing tourists from various countries to visit Indonesia. In addition, it is expected to provide information for the Indonesian government to elaborate on government policies in the tourism sector.

2. Material and Method

2.1. Data

The data used were secondary data derived from a variety of survey institutions such as Statistics Indonesia (Badan Pusat Statistik, BPS), Bank Indonesia (BI), and the World Bank as well as other data attached to the Presidential Decree No. 69 of 2015 and Presidential Decree No. 21 of 2026. This data is data on foreign tourist arrivals in 2020, consisting of 31 countries: ASEAN region (Singapore, Malaysia, Vietnam, Burma, Brunei Darussalam and Thailand), the Asian region (Pakistan, Bangladesh, South Korea, China, India, and Hong Kong), European region (Denmark, Austria, Finland, France, Germany, Belgium, Netherlands, Norway, Russia, Italy, Portugal, Sweden, Switzerland, and Spain), Middle East area (United Arab Emirates, Kuwait, Saudi Arabia, and Qatar), the African region (Yemen), as shown in Table 1.

Table 1. Data Description

Variable	Description	Unit	Data Source
Y ₁	Number of foreign tourists visiting (tourists)	People	BPS
X ₁	Rupiah exchange rate (rupiah exchange rate)	IDR	BI

X ₂	GDP per capita (GDP)	(Thousand) USD	World Bank
X ₃	Population density	Person per km ²	World Bank
X ₄	BVKS Factor (short stay visa free)	Index	World Bank
X ₅	Consumer Price Index (CPI)	Year	World Bank
X ₆	Life expectancy	Percent	World Bank
X ₇	Economic growth	USD	World Bank
X ₈	Total imports (Imports)	1 : have 0 : have not	Presidential Decree No. 69 of 2015 and Presidential Decree No. 21 of 2016
X ₉	CGK distance to international airport	km	Airmilescalculator

2.2. Method

2.2.1. Descriptive Statistics

The purpose of descriptive statistics analysis is to describe the data for each variable. The description may be viewed from the mean, standard deviation, maximum, minimum, sum, range, kurtosis, and skewness (distribution of the distribution).

2.2.2. Poisson Regression

Poisson regression analysis is a statistical method used to analyze the relationship between a random variable. Y response is a discrete variable (count) and p predictor variables $X_1, X_2, X_3, X_4, \dots, X_p$. The response variable Y is assumed to have a Poisson distribution, while the predictor variables are continuous, category and discrete data. Poisson regression model is expressed by (1).

$$\hat{Y} = \exp(X_i^T \beta); i=1,2,\dots,n \quad (1)$$

a. Autocorrelation

The autocorrelation test is a test aimed at determining if the linear regression model has a correlation between the confusing error of period t and the confusing error of period $t-1$.

b. Multicollinearity

To detect or determine the presence or absence of multicollinearity in the model of regression, it may be shown in the value of Variance Inflation Factors (VIF) [11], namely in (2).

$$VIF_j = \frac{1}{(1-R_j^2)} \quad (2)$$

with R_j^2 denotes the value of the determination coefficient obtained from regressing the j th independent variable (x_j) with other independent variables in (3).

$$R^2 = \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (3)$$

With the i th response (y_i), the average (\bar{y}), and the estimated i th response (\hat{y}_i) observed. If the value of $< ,10$ then there is no multicollinearity [11].

2.2.3. Geographically Weighted Poisson Regression (GWPR)

GWPR is a modification of the regression model that is a combination of panel data and GWR. The idea of GWPR is the same as cross-sectional GWR analysis. However, GWPR assumes that the time series of observations is a smooth spatiotemporal process. This process follows a distribution in which the closest observations (either geographic location or time) are more related than distant observations. GWPR analysis aims to combine the overall location (cross sectional) and observations. The general equation of GWPR with a fixed effect model is in (4).

$$\mu_i = \exp \left(\beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i)x_{ij} \right) \quad (4)$$

where μ_i is the corrected average value of the response variable at the i th observation, x_{ij} is the corrected average value of the j th explanatory variable at the i th observation, $\beta_j(u_i, v_i)$ is the value of the j th parameter at location (u_i, v_i) .

a. **GWPR Parameter Estimation**

The parameter estimation of the GWPR model was carried out using the Maximum Likelihood Estimation (MLE) method, aiming to determine the estimator of the parameters by maximizing the likelihood function. Suppose that n random samples are given from the random variable $Y_i \sim P(\lambda(u_i))$. The i th location likelihood function can be written in (5).

$$L(\lambda(u_i), \lambda(u_i) \dots, \lambda(u_i)) = \prod_{j=1}^n \frac{e^{-\lambda(u_i)} \lambda^{y_j}(u_i)}{y_j!} \tag{5}$$

b. **GWPR Hypothesis Testing**

The GWPR model hypothesis testing consists of the GWPR model similarity test and global Poisson regression, simultaneous parameter testing, and partial parameter testing.

2.2.4. Overdispersion

The Poisson regression model assumes that the similarity between the mean and the variance is called equidispersion, which must be met [12]. However, in statistical data analysis, it is often found that the variance is smaller or larger than the average. This situation is called underdispersion or overdispersion [13].

2.2.5. Spatial Heterogeneity

Spatial heterogeneity is a requirement in using point and area approaches [14]. To calculate the spatial heterogeneity, the Breusch-Pagan test can be used with the formula in (6).

$$BP = \left(\frac{1}{2}\right) f^T Z (Z^T Z)^{-1} Z^T f \sim \chi^2 \tag{6}$$

2.2.6. Akaike Information Criteria (AIC)

The AIC method is a method which allows the selection of the best regression model found by Akaike and Schwarz [15]. Equation (7) can be used to calculate AIC value.

$$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \tag{7}$$

The value of k corresponds to the number of parameters estimated in the regression model, n corresponds to the number of findings, and u corresponds to the estimated residual.

2.2.7. Research Method

The author used GWPR method. Following are steps taken. Step 1 was data collection. The data consisted of 2020 data collected from various survey institutions such as BPS, BI and the World Bank as well as other data attached to Presidential Decree No. 69 of 2015 and Presidential Decree No. 21 of 2026. Step 2 was data exploration, which was conducted using descriptive statistics with values of mean, minimum, median, and maximum. Step 3 was testing the assumptions of autocorrelation and multicollinearity. Prior to the Poisson regression modeling, the assumptions were tested, namely the autocorrelation assumption with the Durbin-Watson test and the multicollinearity test by looking at the VIF value. Step 4 was Poisson regression modeling. In this stage, the data was then analyzed by classical regression using Geoda, R, and Minitab software by looking at the significance of the variables on the model through parameter estimation, simultaneous testing, and partial Poisson regression models. Step 5 was overdispersion and spatial heterogeneity test. After obtaining the best regression model, the next step involved conducting an overdispersion test to show the goodness of the Poisson regression model and continue to test the assumption of spatial heterogeneity with the Breusch-Pagan test. Step 6 was GWPR Model. This study employed point spatial regression analysis, namely the GWPR model. In this stage, the model similarity test, partial test, and mapping of significant variables were also carried out. Step 7 was selecting the best model. In this case, the best model was chosen based on the smallest AIC and considered those with the largest R^2 deviance. The research method flowchart is illustrated in Fig. 1.

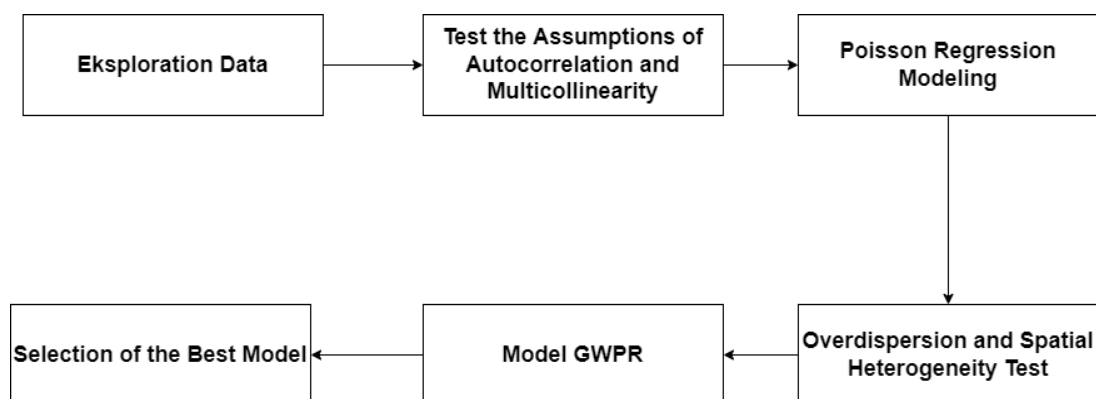


Fig. 1 Research method flowchart.

3. Results and Discussion

3.1. Method

The variable number of foreign tourists visiting Indonesia in 2020 is 2,098,438 foreign tourists spread throughout foreign regions. The lowest number of foreign tourists was from the Qatar region with 255 tourists. Meanwhile, the highest number was from the Malaysian, totaling 980,118 tourists. Data exploration through descriptive statistical methods is presented in Table 2.

Table 2. Descriptive Statistics of International Tourists

Variable	Mean	Minimum	Median
Y ₁	69,948	225	12,768
X ₁	8,643	1	3,916
X ₂	32,783	1,194	31,583
X ₃	703	9	147
X ₄	0.9333	0.0000	1.0000
X ₅	129.35	98.82	116.58
X ₆	79.242	67.780	81.640
X ₇	-3.661	-10.839	-3.705
X ₈	3,283,008,909	65,822,808	757,555,674
X ₉	6,345	522	7,042

3.2. Poisson Regression Modeling

3.2.1. Autocorrelation Test

The autocorrelation test was completed using the Durbin-Watson method with the help of R software, as shown in Fig. 2. The results obtained from the DW value = 1.9207 > dL value = 0.7822 and DW value = 1.9027 < dU = 2.2508 or p-value was 0.4749 > 0.05. The results showed that there was no autocorrelation symptom on data.

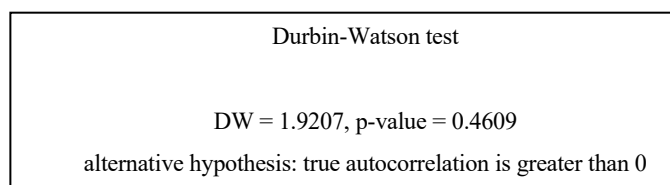


Fig. 2 Durbin-Watson test output.

3.2.2. Multicollinearity Test

The multicollinearity test was performed using the VIF value using the software R (Table 3). All variables had a VIF value < 10, except for the X₃, X₅, X₆, and X₉ variables, meaning that the variables needed to be removed except X₉ because they were still close to the value of 10. VIF of variables

$X_1, X_4, X_5, X_7, X_8,$ and $X_9 < 0$ indicates that the predictor variables are not correlated with each other so that it can be used for the formation of Poisson and GWPR regression models.

Table 3. Poisson Regression Model VIF Value

VIF Value (Variance Inflation Factors)								
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
2.953	2.953	2.953	2.953	2.953	2.953	2.953	2.953	2.953

3.3. Poisson Regression Model Test

3.3.1. Poisson Regression Model Parameter Estimation

The Poisson regression model obtained for foreign tourists to Indonesia in 2020 is as follows,

$$\hat{\mu} = \exp (8,85 - 1.16 \times 10^{-5}\beta_1 + 3.08\beta_4 - 1.36 \times 10^{-3}\beta_5 - 0.20\beta_7 + 6.97 \times 10^{-11}\beta_8 - 2.72 \times 10^{-4} \beta_9)$$

The results of the Poisson regression model parameter estimation test are presented in Fig. 3.

Coefficients:			
value	Estimate	Std. Error	z
	Pr(> z)		
(Intercept)	8.348e+00	1.054e-02	
792.12	<2e-16 ***		
X1	-1.156e-05	1.478e-07	-
78.20	<2e-16 ***		
X4	3.077e+00	8.062e-03	
381.66	<2e-16 ***		
X5	-1.355e-03	3.649e-05	-
37.14	<2e-16 ***		
X7	-2.019e-01	2.724e-04	-
741.05	<2e-16 ***		
X8	6.971e-11	8.003e-14	
071.04	<2e-16 ***		

Fig. 3 Estimation of Poisson regression parameters.

3.3.2. Concurrent Test of Poisson Regression Parameters

Then, the simultaneous test was carried out, the residual deviation value = 2288969 > $\chi^2_{(25;0.05)} = 37.652$, meaning that there is at least one parameter that has a significant effect on the model.

3.3.3. Poisson's Regression Parameter Partial Test

All parameters were greater than $Z_{0.025}=1.96$, indicating that they had a significant effect on the model for all parameters.

3.4. Overdispersion Test

Then the overdispersion test in Fig. 4 was carried out, the output was obtained with p-value = 0.03488 < 0.05, indicating that there was overdispersion in the model, so the Poisson regression model was not good.

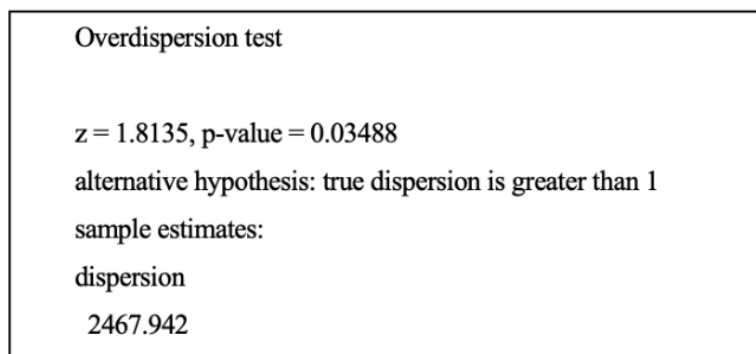


Fig. 4 Overdispersion test output.

3.5. Breusch-Pagan Test

To determine whether there is spatial heterogeneity, the Breusch-Pagan test was performed. The results obtained from the Breusch-Pagan test output (Fig. 5) was $BP = 36.5489 > 12.59$, and p-value $0.00001 < 0.05$, indicating that there was spatial heterogeneity between regions of foreign tourists visiting Indonesia.

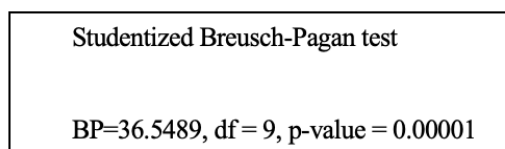


Fig. 5 Breuch-Pagan test output.

3.6. Modeling Geographically Weighted Poisson Regression (GWPR)

In obtaining the GWPR model, first determine the weighting matrix by determining the optimum bandwidth using the golden section search in the range of 0 until 30 and AIC as the selection criteria. The comparison of the weights obtained is as follows.

Table 4. Functional Weighting Comparison

Kernel	AIC
Adaptive Bi-Square	2,166,650.598
Fixed Gaussian	1,521,240.873

Table 4 shows that the GWPR model with the kernel fixed Gaussian weighting function has a lower AIC than the Kernel Adaptive Bi-Square weighting function. In this analysis, the kernel fixed Gaussian function with an optimum bandwidth of 11,771 was used as a weight in the GWPR modeling.

3.7. GWPR Model Parameter Partial Test

To find out the factors that influence the number of foreign tourists visiting Indonesia in 2020, a partial test was conducted and compared the t_{count} value with the $t_{(0.025,25)}$ value = 2.06. H_0 is rejected if $|t_{count}| > 2.056$, it can be concluded that the j th parameter and i th location have a significant effect on the model. If GWPR modeling was carried out with kernel-fixed Gaussian weighting, two groups were obtained according to the significant similarity of variables. For example, the regression model for the number of foreign tourists from Bangladesh is obtained as follows.

$$\hat{\mu} = \exp(1.36 + 0.000007\beta_1 + 4.73855\beta_4 + 0.036954\beta_5 - 0.32144\beta_7 - 0.001099\beta_9)$$

One interpretation of the above model is that every X_5 increases by one unit with other variables constant, the number of foreign tourists will be multiplied by $\exp(1.396954)$.

3.8. Variable Significance Group

Two divisions of groups were obtained according to the similarity of significant variables (Table 5). All variables that had a significant effect were found in almost all countries of origin of foreign tourists.

Table 5. Variable Significance

Tourist Origin	Significant Variable
<i>Finland dan Qatar</i>	X_1, X_5, X_7, X_8 dan X_9
Bangladesh, Burma, Brunei Darussalam, China, Denmark, Austria, France, Germany, India, Italy, Korea Selatan, Kuwait, Malaysia, Belgium, Hong Kong, Norway, Pakistan, Netherlands, Portugal, Saudi Arabia, Spain, Singapore, Sweden, Russia, Thailand, Switzerland, United Arab Emirates, dan Vietnam.	X_1, X_4, X_5, X_7, X_8 dan X_9

3.9. Model Comparison

Table 6 shows that the GWPR model is better used to analyze the number of foreign tourists visiting Indonesia because it has a smaller AIC and deviance value and a larger deviance- R^2 than the Poisson regression model, which is 0.741.

Table 6. Model Comparison

Model	AIC	Deviance	Deviance- R^2
Poisson Regression	2,288,982.793	2,288,968.794	0.611
GWPR	1,521,240.873	1,521,196.695	0.741

4. Conclusion

Modeling the number of foreign tourists coming to Indonesia in 2020 using GWPR with a kernel fixed Gaussian weighting function had an AIC value of 1,521,240.873, deviance of 1,521,196.695, and deviance- R^2 of 0.741 or 74.1%. This model produced two different clusters of characteristics of foreign tourists' country of origin based on the variable's significance. Cluster one consisted of Finland and Qatar, while the rest were in cluster two. The characteristics of cluster two were influenced by the rupiah exchange rate, BVKS factor, CPI, economic growth, total imports, and the distance of CGK to the international airport. Cluster one had almost the same characteristics as cluster two, but it was not influenced by the BVKS factor variable.

References

- [1] H. Wulandari, R. Indriani, R. Untari, and S.S. Bethagustav, *Statistik Kunjungan Wisatawan Mancanegara 2020*. Jakarta, Indonesia: Badan Pusat Statistik Indonesia, 2021.
- [2] L.U. Calderwood and M. Soshkin, "The Travel & Tourism Competitiveness Report 2019: Travel and Tourism at a Tipping Point," World Economic Forum, Geneva, Switzerland, Insight Rep., 2019.
- [3] M. Azizurrohman, R.B. Hartarto, Y-M. Lin, and F.H. Nahar, "The Role of Foreign Tourists in Economic Growth: Evidence from Indonesia," *Jurnal Ekonomi & Studi Pembangunan*, Vol. 22, No. 2, pp. 313–322, Oct. 2021, doi: 10.18196/jesp.v22i2.11591.
- [4] R. Berk and J.M. MacDonald, "Overdispersion and Poisson Regression," *Journal of Quantitative Criminology*, Vol. 24, pp. 264–286, Sep. 2008, doi: 10.1007/s10940-008-9048-4.
- [5] M.J. Hayat, and M. Higgins, "Understanding Poisson Regression," *Journal of Nursing Education*, Vol. 53, No. 4, pp. 207–215, 2014 doi: 10.3928/01484834-20140325-04.
- [6] D.W. Osgood, "Poisson-Based Regression Analysis of Aggregate Crime Rates," *Journal of Quantitative Criminology*, Vol. 16, pp. 21–43, Mar. 2000, doi: 10.1023/A:1007521427059.
- [7] R. Paternoster And R. Brame, "Multiple Routes to Delinquency? A Test of Developmental and General Theories of Crime," *Criminology*, Vol. 35, No. 1, pp. 45–84, Feb. 1997, doi: 10.1111/j.1745-9125.1997.tb00870.x.
- [8] R.J. Sampson and J.H. Laub, "Socioeconomic Achievement in the Life Course of Disadvantaged Men: Military Service as a Turning Point, Circa 1940–1965," *American Sociological Review*, Vol. 61, No. 3, pp. 347–367, Jun. 1996.

- [9] I.M. Fadlilah, Sugiman, and Sunarmi, “Estimasi Parameter Model Regresi Spasial dengan Metode Geographically Weighted Poisson Regression,” *UNNES Journal of Mathematics*, Vol. 8, No. 2, pp. 21–31, Nov. 2019, doi: 10.15294/ujm.v8i2.23796.
- [10] H. Akaike, “A New Look at the Statistical Model Identification,” *IEEE Transactions Automatic Control*, Vol. 19, No. 6, pp. 716–723, Dec. 1974, doi: 10.1109/TAC.1974.1100705.
- [11] S. Candraningtyas, “Regresi Robust Mm-Estimator untuk Penanganan Pencilan Regresi Linier Berganda,” *Jurnal Gaussian*, Vol. 2, No. 4, pp. 395–404, Oct. 2013, doi: 10.14710/j.gauss.2.4.395–404.
- [12] Darnah, “Menentukan Model Terbaik dalam Regresi Poisson dengan Menggunakan Koefisien Determinasi,” *Jurnal Matematika, Statistika, dan Komputasi*, Vol. 6, No. 2, pp. 59–71, Jan. 2010.
- [13] W. Kusuma, D. Komalasari, and M. Hadijati, “Model Regresi Zero Inflated Poisson pada Data Overdispersion,” *Jurnal Matematika*, Vol. 3, No. 2, pp. 71–85, Dec. 2013, doi: 10.24843/JMAT.2013.v03.i02.p37.
- [14] L. Anselin, “An Introduction to Spatial Autocorrelation Analysis with GeoDa.” Accessed: May 20, 2022. [Online]. Available: <https://www.dpi.inpe.br/gilberto/tutorials/software/geoda/tutorials/spauto.pdf>
- [15] M. Fathurahman, “Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike’s Information Criterion dan Schwarz Information Criterion,” *Jurnal Informatika Mulawarman*, Vol. 4, No. 3, pp. 37–41, Sep. 2009, doi: 10.30872/jim.v4i3.41.