ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

INDONESIAN JOURNAL OF GEOGRAPHY

RESEARCH ARTICLE

Spatial Autoregressive Modeling of Diarrhea in East Java Province

Andi Muhammad Hakam[1*], Reza Hardiansyah[2], Ridwan[3]

[123]Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, 90245, Indonesia

Submit:
Received:
Accepted:

Key words: Diarrhea, East Java Province, Spatial Autoregressive, Spatial Effect, Regression.

Correspondent email:
andimuhammadhakam@gmail.com

Abstract. Diarrhea is one of the health problems in developing countries, including Indonesia. East Java Province is one of the provinces in Indonesia that experiences a significant incidence of diarrhea with a prevalence rate of 0.0115%. The high number of diarrhea cases in this province is suspected to be influenced by geographical location factors, and this case may have a spatial effect. Therefore, an analysis was conducted using the Spatial Autoregressive (SAR) method to provide a more comprehensive understanding of the factors contributing to diarrhea cases. Based on the analysis results, it was found that the factors significantly influencing the number of individuals affected by diarrhea in East Java Province are the number of drinking water providers in regencies/cities in East Java Province, the population density in regencies/cities in East Java Province, and the number of households experiencing water contamination in regencies/cities in East Java Province.

1. Introduction

Diarrhea is one of the global problems in various countries, especially developing countries like Indonesia. Diarrhea is one of the causes of illness and death among children worldwide [1]. The morbidity rate caused by diarrhea in every region of Indonesia is relatively high, indicating that diarrhea is an endemic disease and a potential outbreak (Kejadian Luar Biasa/KLB) disease [2]. KLB refers to a significant epidemiological occurrence of illness or death within a population group during a specific period or the occurrence of a communicable disease that has increased two-fold or more compared to the previous period [3]. East Java Province is one of the provinces in Indonesia that experiences a significant incidence of diarrhea [4].

East Java Province has recorded an increase in diarrhea cases that require attention. According to data from Social Security Agency on Health (Badan Penyelenggara Jaminan Sosial/BPJS Kesehatan), the prevalence of diarrhea cases in East Java Province was 0.0115% in 2018 [5]. Its impact is not limited to public health aspects but also affects the education, economy, and social sectors in the region. Therefore, a deeper understanding of the contributing factors to diarrhea cases in this province is needed [6][7].

The high number of diarrhea cases is suspected to be influenced by the geographical location factor in East Java. This is because

ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

INDONESIAN JOURNAL OF GEOGRAPHY

neighboring regencies/cities have similar characteristics. Based on this, the diarrhea case model in East Java Province may have a spatial effect. Therefore, an analysis is conducted using spatial regression analysis that considers the location aspect. This approach combines linear regression elements with spatial elements to account for spatial interdependencies between the analysis units, determine the relationships between variables, and measure their influence while considering spatial effects [8].

In this context, modeling diarrhea cases in East Java Province using the Spatial Autoregressive regression method can provide a more comprehensive understanding of the contributing factors to diarrhea cases in 2018, taking into account their spatial effects. The results of this modeling can provide valuable insights to local governments and relevant stakeholders in designing more effective interventions to reduce diarrhea cases in this region [9].

By understanding the contributing factors to diarrhea cases in East Java Province through modeling using the Spatial Autoregressive regression, targeted prevention and control efforts can be implemented, considering the spatial patterns of diarrhea occurrences. This modeling can also provide new insights into the spatial relationship between risk factors and diarrhea cases, contributing to public health research and serving as a guide for better policy decision-making in the future.

2. Materials and Methods

Data and Research Variables

The data used in this study are secondary data obtained from the Health Service (Dinas Kesehatan), Central Agency of Statistics (BPS) and participant visit records of Social Security Agency (BPJS) to Primary Health Facilities (Fasilitas Kesehatan Tingkat Pertama/FKTP) in East Java Province in 2018. The dependent variable in this study is the number of individuals exposed to diarrheal disease (Y). The independent variables include the number of water supply providers (X_1), the number of healthcare facilities (X_2), population density (X_3), the percentage of households with access to safe drinking water (X_4), the percentage of households with access to proper sanitation (X_5) and the number of households experiencing water contamination (X_6).

Linear Regression Model

Regression model is an equation that describes the relationship between a dependent variable and independent variables. If the parameters in the model have a linear relationship with the dependent variable, it is called a linear regression model. Furthermore, this model can be used to predict the value of the dependent variable when the values of the independent variables are given.

Let y_i be the observation of the dependent variable for the i-th observation, x_{it} be the value of the independent variable t for the i-th and ε_i be the error of the i-th observation. Therefore, the regression model can be written as follows [10].

$$\begin{aligned} y_1 &= \beta_1 + x_{21}\beta_2 + \cdots + x_{k1}\beta_k + \varepsilon_1 \\ y_2 &= \beta_1 + x_{22}\beta_2 + \cdots + x_{k2}\beta_k + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_1 + x_{2n}\beta_2 + \cdots + x_{kn}\beta_k + \varepsilon_n \end{aligned} \tag{1}$$

or it can be expressed in matrix form as follows.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where,

$\boldsymbol{Y}$: The vector of dependent variable with size $n \times 1$

$\boldsymbol{X}$: The matrix of k independent variables or regressor variables with size $n \times k$

$\boldsymbol{\beta}$: The vector of parameters with size $k \times 1$

$\boldsymbol{\varepsilon}$: The vector of errors with size $n \times 1$.

Spatial Autocorrelation

Spatial autocorrelation is one of the spatial influences expressed through weighting in the form of a matrix that depicts the proximity relationship between observations, or known as a spatial weighting matrix. [11].

Spatial Weighting Matrix

The general form of a spatial weighting matrix is as follows.

$$W_{ij} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix}$$

ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

INDONESIAN JOURNAL OF GEOGRAPHY

The underlying structure of neighborhood relationships is represented by weights of 0 and 1. The element W_{ij} will have a value of 0 if i and j do not have a neighboring relationship or if $i = j$, conversely, W_{ij} will have a value of 1 if i and j have a neighboring relationship or $i \neq j$ [12].

Moran's Index

Moran's I test is a statistical test used to assess the spatial autocorrelation and identify locations based on spatial autocorrelation values. The hypotheses used in Moran's I test are as follows.

$H_0: I = 0$ (There is no spatial autocorrelation).

$H_1: I \neq 0$ (There is spatial autocorrelation).

The statistical test used is as follows.

$$Z_{hit} = \frac{1 - E(I)}{\sqrt{Var(I)}} \quad (3)$$

where,

$$I = \left(\frac{n\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}w_{ij}(x_i - \bar{x})(x_j - \bar{x}}{W\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\right) \quad (4)$$

$$E(I) = -\frac{1}{n-1} \quad (5)$$

The value of index I is between-1 and 1. If $I > I_0$ then the data has positive autocorrelation and a clustering pattern, conversely if $I < I_0$ then the data has negative autocorrelation and a dispersing pattern [4].

Lagrange Multiplier (LM)

The Lagrange Multiplier (LM) test is used as a basis for selecting the appropriate spatial regression model for the data at hand. This test is also used to investigate spatial dependence, including lag dependence, using the Lagrange Multiplier Lag (LM_{lag}). The hypotheses used in LM_{lag} are as follows.

$H_0: \rho = 0$ (There is no spatial lag dependence).

$H_1: \rho \neq 0$ (There is spatial lag dependence).

The statistical test used is as follows.

$$LM_{lag} = \frac{\left(\frac{e'W_1y}{\sigma^2}\right)^2}{\frac{((W_1X\beta)'M(W_1X\beta) + T\sigma^2)}{\sigma^2}} \quad (6)$$

The decision criterion for drawing a conclusion is to reject the null hypothesis (H_0) if $LM_{lag} > \chi^2_{(\alpha,1)}$ or $p - value < \alpha$ [12].

Spatial Regression Model

A spatial model involving spatial influence is called a spatial regression model. The presence of spatial autocorrelation leads to the formation of spatial autoregressive and moving average parameters, resulting in the following form of spatial process [11].

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \quad (7)$$

$$\boldsymbol{u_t} = \lambda\boldsymbol{W}\boldsymbol{u_{t-1}} + \boldsymbol{\varepsilon} \quad (8)$$

where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$ there is no autocorrelation, and,

$\boldsymbol{y}$: The vector of dependent variable with size $n \times 1$

ρ : Spatial lag coefficient parameter

$\boldsymbol{W}$: Spatial weighting matrix with size $n \times n$

$\boldsymbol{X}$: The matrix of independent variables with size $n \times (p + 1)$

$\boldsymbol{\beta}$: Regression coefficient parameter with size $(p + 1) \times 1$

$\boldsymbol{u}$: The vector of errors with size $n \times 1$

λ : Spatial lag coefficient parameter on the error

$\boldsymbol{\varepsilon}$: The vector of errors with size $n \times 1$.

Spatial Autoregressive (SAR)

Spatial Autoregressive (SAR) is a spatial regression model that incorporates spatial influence on the response variable. The SAR model combines a linear regression model with spatial lag on the response variable using cross-sectional data [12]. The coefficients of spatial lag indicate the level of correlation between the influence of one region on the surrounding regions. The general form of the SAR model is as follows [13].

$$\boldsymbol{y} = \boldsymbol{\rho}\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (9)$$

$$\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma}^2\boldsymbol{I})$$

where,

$\boldsymbol{y}$: The vector of dependent variable with size $n \times 1$

ρ : Spatial lag coefficient parameter

$\boldsymbol{W}$: Spatial weighting matrix with size $n \times n$

$\boldsymbol{X}$: The matrix of independent variables with size $n \times (p + 1)$

$\boldsymbol{\beta}$: Regression coefficient parameter with size $(p + 1) \times 1$

$\boldsymbol{\varepsilon}$: The vector of errors with size $n \times 1$.

As for the parameter estimation method for the SAR model, it is as follows.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t(\boldsymbol{I} - \rho\boldsymbol{W})\boldsymbol{y} \quad (10)$$

ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

INDONESIAN JOURNAL OF GEOGRAPHY

Diarrhea

Diarrheal disease, which accounts for a significant number of fatalities among children under the age of five, is the second most common cause of death in this age group. It claims the lives of approximately 525.000 children annually. Diarrhea, characterized by prolonged episodes lasting multiple days, depletes the body of essential water and salts crucial for survival. Typically, diarrhea is an indication of an infection affecting the intestinal tract, caused by various types of bacteria, viruses, and parasites. Infections can spread through the consumption of contaminated food or water, as well as through person-to-person contact due to inadequate hygiene practices. The causes of diarrhea encompass the presence of infectious symptoms arising from diverse microorganisms, malnutrition, contamination of water sources with fecal matter or waste, and other contributing factors [14]. In East Java Province, the highest number of diarrhea cases occurs in Sidoarjo Regency and Surabaya City due to environmental sanitation, drinking water sources, clean and healthy living behaviors, nutritional status, and immunization [15].

The Methods of Analysis Data

The analysis method employed in this study is Spatial Autoregressive (SAR) analysis, with the following steps [4].

1. Conducting an initial exploration by displaying a sample of diarrhea cases' distribution in the form of a spatial map.
2. Estimating linear regression parameters using the following model.
$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{(\rho-1)} X_{(\rho-1)} + \varepsilon_i \quad (11)$$
The process of building a regression model begins with the selection of important variables for the model. Variable selection is done using the Stepwise method.
3. Determining the spatial weighting matrix (W). Based on the inter-location relationships, spatial weights for each pair of locations will be obtained.
4. Investigating the presence of spatial autocorrelation among districts to analyze it using spatial regression. The test used for this purpose is Moran's I test.
5. Investigating the presence of spatial dependence in the lag.
6. Estimating parameters using the Ordinary Least Squares (OLS) method.
7. Conducting significance testing of parameters to investigate the role of independent variables in the model using the Wald test.
8. Interpreting and concluding the obtained results.

3. Results and Discussion

Data Exploration

The data exploration stage in this research aims to showcase the number of occurrences depicted in the form of a thematic map. A thematic map is used to convey additional information. Thematic maps generally show that brighter colors in a region indicate lower values of the observed variable, while darker colors indicate higher values of that variable.

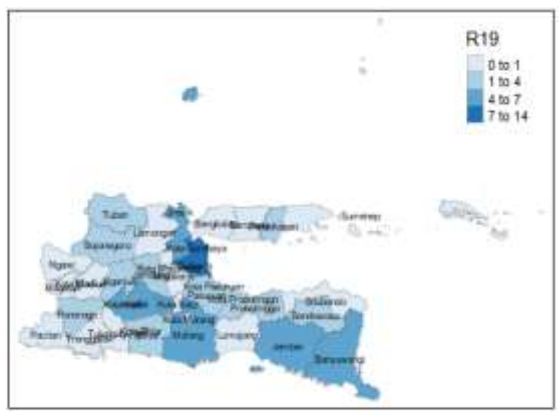

Figure 1. Map of Diarrhea Cases Distribution in East Java Province, 2018

In Figure 1, it can be observed that the cities/districts with the same color, namely Surabaya City and Sidoarjo Regency, are neighbors. This indicates that these cities/districts belong to the high category of diarrhea cases in East Java Province, with 7 to 14 cases. Furthermore, Mojokerto Regency is adjacent to Mojokerto City and Batu City. This suggests that these districts and cities fall into the low category of diarrhea cases in East Java Province, with 0-1 cases.

ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

Estimated Parameters of the Linear Regression Model

The linear regression model estimation was performed using R software version 4.3.0. The estimated linear regression model is as follows.

Table 1. Estimated Parameters of the Linear Regression Model

Parameters	Estimated Parameters	t_{value}	$p-value$
β_0	$-5{,}596$	$-2{,}52$	0,0170
β_1	$-0{,}00034$	$-0{,}23$	0,8158
β_2	0,127	7,027	$6{,}86 \times 10^{-8}*$
β_3	0,00041	2,441	$0{,}0206*$
β_4	0,0374	1,140	0,2632
β_5	$-0{,}0108$	$-0{,}70$	0,4863
β_6	0,0445	2,344	$0{,}0256*$

Based on Table 1, the estimated linear regression model obtained is as follows.

$$\begin{aligned}\hat{y}_i = &-5{,}596\\ &\pm 0{,}00034x_{1,i}\\ &+0{,}127x_{2,i}\\ &+0{,}00041x_{3,i}\\ &-0{,}0374x_{4,i}\\ &-0{,}0108x_{5,i}\\ &+0{,}0445x_{6,i}.\end{aligned}$$

Next, the process of selecting significant variables is performed using the Stepwise method. The predictor variables that are found to be non-significant (*p-value* $> \alpha = 0{,}05$ and not marked with $*$) are X_1, X_4 and X_5. Therefore, the model estimation is repeated by removing the non-significant predictor variables as follows.

Table 2. Estimated Parameters of the Linear Regression Model for Significant Variables

Parameters	Estimated Parameters	t_{value}	$p-value$
β_0	$-3{,}719$	$-6{,}397$	$2{,}64 \times 10^{-7}$
β_2	0,132	7,132	$3{,}86 \times 10^{-9}$
β_3	0,00046	3,996	0,000328
β_6	0,0385	2,302	0,0276

Based on Table 2, the estimated linear regression model is as follows.

$$\begin{aligned}\hat{y}_i = &-3{,}719\\ &+0{,}132x_{2,i}\\ &+0{,}00046x_{3,i}\\ &+0{,}0385x_{6,i}\end{aligned}$$

Next is to perform a spatial dependence test to investigate the spatial effects on the data, allowing for spatial regression analysis. Before conducting spatial regression analysis, a spatial weighting matrix needs to be constructed. In this study, the spatial weighting matrix is based on distance using the k-Nearest Neighbor (k-NN) method with $k = 1$.

Spatial Dependence Test

Spatial dependence can be tested using the Moran's I test and the Lagrange Multiplier test. Based on the Moran's I test $p-value = 1{,}57 \times 10^{-4}$ is obtained. Since the *p-value* smaller than $\alpha = 0{,}05$ we reject the null hypothesis (H_0), indicating the presence of spatial autocorrelation in the number of diarrhea case samples in East Java Province.

The value of I in the Moran's I test is 0,70303553 which is greater than $E(I) = -0{,}02702703$ ($I > I_0$). This indicates the presence of positive autocorrelation, meaning that nearby cities and districts have similar characteristics and exhibit clustering patterns. To observe this, you can refer to the Moran's I scatterplot shown below.

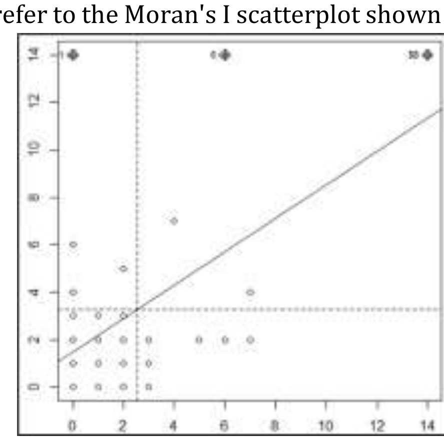

Figure 2. Moran's I Scatterplot of Diarrhea Cases Samples

Figure 2 shows the distribution of diarrhea sample counts in quadrant I (high-high) tends to be away from the cities and districts with low diarrhea sample counts or quadrant III (low-low).

Table 3. Lagrange Multiplier Test

Statistical Test	$p-value$
LM_{lag}	0,00856

Next, a Lagrange Multiplier test is conducted and based on Table 3, it indicates that LM_{lag} is significant ($p-value < 0{,}05$) therefore the alaysis

ISSN ....(online), ISSN .... (print)
Indonesian Journal of Geography Vol ..., No ... (......):
DOI: .... website :

can proceed with Spatial Autoregressive (SAR) analysis.

Spatial Autoregressive (SAR) Modelling

In the SAR modelling process, significance of the paraneters obtained is tested using the Wald test.

$H_0: \rho, \lambda, \beta_j = 0$ (Non-significant parameters).

$H_1: \rho, \lambda, \beta_j \neq 0$ (Significant Parameters).

with the test statistics used as follows.

$$W = \left[\frac{r}{se(r)}\right]$$

and the test criterion is to reject the null hypothesis (H_0) if the value $W > Z_{\alpha/2}$ or $p - value < \alpha$ [12]. The results of the Wald test are as follows.

Table 4. Wald Test fo the SAR Model

Parameters	Estimated Parameters	Wald	$p - value$
β_0	$-3{,}048$	$-5{,}6732$	$1{,}42 \times 10^{-8}$ *
β_2	$0{,}128$	$8{,}921$	$2{,}2 \times 10^{-16}$*
β_3	$0{,}00041$	$4{,}0354$	$5{,}452 \times 10^{-5}$ *
β_6	$0{,}0398$	$4{,}0354$	$0{,}00534$ *

From table 4. The obtained $p - value$ for the model are smaller than 0,05 indicating a decision to reject the null hypothesis (H_0). This means that variables X_2, X_3 and X_6 are significant, and the following model is obtained.

$$\begin{aligned}\hat{y}_i = &-3{,}048 \\ &+0{,}128X_{2,i} \\ &+0{,}00041X_{3,i} \\ &+0{,}0398X_{6,i}\sum_{j=1,i\neq j}^{38} W_{ij}u_j\end{aligned}$$

where,

u_j : Spatial residual of j-th regencies/cities.

W_{ij} : Element of the spatial weigthing matrix (k-NN distance)

Model Interpretation

Here are the interpretations of the previous SAR model.

1. The coefficient of X_2 indicates that if the number of drinking water providers in the regencies and cities of East Java Province increases, the number of samples exposed to diarrhea (Y) will increase by 0,128%.
2. The coefficient of X_3 indicates that if the population density in the regencies/cities of East Java Province increases, the number of samples exposed to diarrhea (Y) will increase by 0,00041%.
3. The coefficient of X_6 indicates that if the number of household water contamination sources in the regencies/cities of East Java Province increases, the number of samples exposed to diarrhea (Y) will increase by 0,0398%.

4. Conclusion

Based on the analysis results, the proximity between regencies/cities in East Java Province has an influence, whereby the geographical conditions in East Java Province affect the number of diarrhea cases. In the modeling conducted using Spatial Autoregressive (SAR), the obtained model is as follows: $\hat{y}_i = -3{,}048 + 0{,}128X_{2,i} + 0{,}00041X_{3,i} + 0{,}0398X_{6,i}\sum_{j=1,i\neq j}^{38} W_{ij}u_j$.

The significant factors that influence the number of samples exposed to diarrhea in East Java Province are the number of drinking water providers in the regencies/cities of East Java Province, population density in the regencies/cities of East Java Province, and the number of households experiencing water contamination in the regencies/cities of East Java Province. These factors have been found to have a significant impact on the occurrence of diarrhea cases in the region.

References

[1] Kemenkes RI, “Kementerian Kesehatan Republik Indonesia,” *Kementerian Kesehatan Republik Indonesia*, 2012. https://www.kemkes.go.id/article/view/13010200015/download-pusdatin-profil-kesehatan-indonesia-2012.html (accessed Jun. 16, 2023).

[2] BPS, “Statistik Indonesia 2019,” *Badan Pusat Statistik Indonesia*, 2019. https://www.bps.go.id/publication/2019/07/04/daac1ba18cae1e90706ee58a/statistik-indonesia-2019.html (accessed Jun. 16, 2023).

[3] J. Chin, *Manual Pemberantasan Penyakit enular*. Jakarta: Infomedika, 2009.

[4] I. G. A. D. Sulasih, M. Susilawati, and N. L. P. Suciptawati, “Pemodelan Kasus Diare di

Provinsi Bali dengan Metode Analisis Regresi Spasial," *E-Jurnal Matematika*, vol. 10, no. 2, pp. 95–102, May 2021.

[5] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Tahun 2018*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur, 2018.

[6] B. Gunawan, "Determinan Kejadian Diare pada Balita di Kabupaten Sidoarjo," *Jurnal Kesehatan Masyarakat*, vol. 6, no. 1, pp. 15–23, 2018.

[7] E. Yunitasari and R. Pratiwi, "Sanitasi, Perilaku Hidup Bersih dan Sehat (PHBS), dan Kejadian Diare pada Balita di Kabupaten Malang," *Jurnal Kesehatan Lingkungan Indonesia*, vol. 18, no. 2, pp. 65–72, 2019.

[8] H. Huang and J. LeSage, "Spatial Econometric Models of Interdependent Processes: Recent Advances and Future Directions," *J Reg Sci*, vol. 55, no. 3, pp. 317–339, 2015.

[9] S. F. Messner, L. Anselin, D. F. Hawkins, G. Deane, and S. E. Tolnay, "Spatial Analysis of Homicide Clearance Rates: The Importance of Considering Spatial Dependence," *J Quant Criminol*, vol. 33, no. 2, pp. 361–382, 2017.

[10] D. N. Gujarati and D. C. Porter, *Dasar-Dasar Ekonometrika*. Jakarta: Salemba Empat, 2010.

[11] L. Mubtadiah, "Estimasi Parameter Model Resgresi Spasial Error dengan Metode Maximum Likelihood Estimation," Universitas Islam Negeri (UIN) Maulana Malik Ibrahim, Malang, 2011.

[12] L. Anselin, *Spatial Econometrics: Method and Models*. Dordrecht: Kluwe Academic Publisher, 1988.

[13] J. LeSage and K. R. Pace, *Introduction to Spatial Econometrics*. London: Chapman and Hall/CRC, 2009.

[14] WHO, "Diarrhoeal disease," *World Health Organizatin*, May 02, 2017. https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease (accessed Jun. 15, 2023).

[15] BPS Provinsi Jawa Timur, "Kesehatan," *BPS Provinsi Jawa Timur*. https://jatim.bps.go.id/subject/30/kesehatan.html (accessed Jun. 15, 2023).