

GEOGRAPHICALLY WEIGHTED PANEL REGRESSION (GWPR) FOR COVID-19 CASE IN INDONESIA

Zakiyah Mar'ah¹*, Sifriyani²

¹Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Negeri Makassar
Daeng tata Street, Makassar, 90224, Indonesia

²Department of Statistics, Faculty of Mathematics and Natural Science, Mulawarman University
Barong Tongkok Street, Samarinda, 75123, Indonesia

Corresponding author's e-mail: *zakiyahm@unm.ac.id

ABSTRACT

Article History:

Received: 16th December 2022

Revised: 17th April 2023

Accepted: 20th April 2023

Keywords:

Covid-19;

Fixed Effect Model;

Fixed Gaussian Kernel;

Geographically Weighted

Panel Regression.

Coronavirus disease 2019 (COVID-19) is a newly emerging infectious disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which was declared a pandemic by the World Health Organization (WHO) on March 11th, 2020. The response to this ongoing pandemic requires extensive collaboration across the scientific community to contain its impact and limit further transmission. Modeling to see cause-and-effect relationships in an event usually uses the Multiple Linear Regression (Ordinary Least Square) method. But in the case of Covid-19, the spread of the virus occurred from one location to another, so there was an indication that there was a spatial effect on the incident. In this study, we did not only look at spatial perspective but also considered time series data, so the method used was Geographically Weighted Panel Regression (GWPR). This study modeled the number of positive cases of Covid-19 in 34 provinces in Indonesia that occurred from March 2020 to August 2021 and looked at what factors influenced the number of positive cases of Covid-19 in each province. GWPR was performed with the assumption of a Fixed Effect Model (FEM). The FEM assumption was used by considering that the conditions of each observation unit were different. Based on the results, the best GWPR model obtained was the GWPR model with a Fixed Gaussian Kernel. The predictor variables that influenced the number of positive cases of Covid-19 were different at each location and tend to cluster at certain locations.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

Z. Mar'ah and Sifriyani., "GEOGRAPHICALLY WEIGHTED PANEL REGRESSION (GWPR) FOR COVID-19 CASE IN INDONESIA," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 0879-0886, June, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id

Research Article • **Open Access**

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a newly emerging infectious disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which was declared a pandemic by the World Health Organization (WHO) on March 11th, 2020. The response to this ongoing pandemic requires extensive collaboration across the scientific community to contain its impact and limit further transmission [1]. Several scientific studies on Covid-19 that have been carried out include: Sarker et al. modeled Covid-19 in India using SARIQsq sensitivity analysis [2], Soukhovolsky et al. used ADL (Autoregressive Distributed Lag) mathematical modeling on the number of Covid-19 cases in several countries such as the United States, Russia, South Korea, and other countries [3], Sifriyani and Rosadi used SIR (Susceptible Infected Recovered) mathematical modeling to estimate the number of Covid-19 sufferers in Kalimantan and Samarinda [4], Mahdy used Geographically Weighted Regression to model the number of Covid-19 cases in West Java [5], Sifriyani et al. estimated the number of positive cases in Kalimantan Province using Geographically and Temporally Weighted Regression [6], and many other studies have been conducted to contain the impact and limit further transmission of the disease.

Based on previous studies, it is always suspected that population density is one of the factors causing the increase in the number of positive cases of Covid-19. Some news and discussion on social media related to Covid-19 in Jakarta have linked overcrowding with the spread of the virus. It was said that because of the high population density in Jakarta, the virus would spread more easily, and its penetration would be faster in areas with higher density [7]. Several other causative factors are thought to have influenced the number of positive cases of Covid-19, including the number of elderly people, the number of households with access to proper drinking water, the GRDP of business fields based on current prices, and the number of villages with health centers [6].

Modeling to see cause-and-effect relationships in an event usually uses Multiple Linear Regression (Ordinary Least Square) method. But in the case of Covid-19, the spread of the virus occurred from one location to another, so it was indicated that there was a spatial effect on the incident. Several studies conducted by Marhamah and Mindra Jaya [8], also Mahdy [5] have shown that modeling the number of positive cases of Covid-19 can be performed using the regression method with a point spatial approach, namely Geographically Weighted Regression.

In this study, we did not only look at spatial perspective but also considered time series data, so the method used was Geographically Weighted Panel Regression. This study modeled the number of positive cases of Covid-19 in 34 provinces in Indonesia that occurred from March 2020 to August 2021 and looked at what factors influenced the number of positive cases of Covid-19 in each province. Geographically Weighted Panel Regression (GWPR) was performed with the assumption of a Fixed Effect Model (FEM). The FEM assumption was used by considering that the conditions of each observation unit were different.

2. RESEARCH METHODS

2.1 Geographically Weighted Panel Regression (GWPR)

The main idea of GWPR is the same as GWR analysis. In GWPR, the time series of observations at a geographic location is assumed to be the implementation of a smooth spatiotemporal process. This process follows a distribution in which close observations (geographical location or time) are more relevant than distant observations. GWPR analysis aims to combine the overall location (cross section) and observations [9]. The GWPR method is a local regression with repeated data at location points for each spatial observation. In other words, GWPR focuses more on repeated spatial observations for each location [10].

GWPR is a development model that combines the GWR model with panel regression. In this study, it is assumed that the conditions of each unit of observation are different, so a panel regression with the FEM model is used, which is shown in Equation (1) [11].

$$\mathbf{y}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T_1 \quad (1)$$

One way to estimate the value of $\boldsymbol{\beta}$ in FEM model is to eliminate α_i with a within estimator [12], that is, calculating the average of Equation (1) against $t = 1, \dots, T$ to obtain the cross section equation as follows:

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta} + \alpha_i + \bar{\boldsymbol{\varepsilon}}_i \quad (2)$$

which,

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}, \quad \bar{\mathbf{x}}_i = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it}, \quad \boldsymbol{\varepsilon}_{it} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_{it}$$

then reduce **Equation (1)** with **Equation (2)**, until it becomes:

$$(y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \boldsymbol{\beta} + (\boldsymbol{\varepsilon}_{it} - \bar{\boldsymbol{\varepsilon}}_i) \quad (3)$$

or

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it} \boldsymbol{\beta} + \ddot{\boldsymbol{\varepsilon}}_{it}, \quad t = 1, 2, \dots, T_1 \quad (4)$$

Then perform Ordinary Least Square (OLS) method to obtain the estimation value of $\boldsymbol{\beta}$ by using regression equation \ddot{y}_{it} on $\ddot{\mathbf{x}}_{it}$ [13].

The GWPR model is obtained from a combination of the GWR model and panel regression. **Equation (5)** is a combination of the GWR equation and the FEM panel regression equation with the within estimator.

$$\ddot{y}_{it} = \beta_0(u_{it}, v_{it}) + \sum_{k=1}^p \beta_k(u_{it}, v_{it}) \ddot{x}_{itk} + \ddot{\boldsymbol{\varepsilon}}_i, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T_1 \quad (5)$$

where \ddot{y}_{it} is the demeaned average response at the i -th observation and t -th time. \ddot{x}_{itk} is the value of the k -th corrected average predictor variable at the i -th observation and t -th time. $\beta_0(u_{it}, v_{it})$ is the intercept of the equation formed at the i -th observation and t -th time. $\beta_k(u_{it}, v_{it})$ is the regression coefficient on the k -th corrected average predictor variable at the i -th observation and t -th time. k is the number of predictor variables. (u_{it}, v_{it}) is the coordinate point of the observation location at the i -th observation and t -th time. And $\ddot{\boldsymbol{\varepsilon}}_i$ is the random effect, which is assumed to be independent, identical and follows a normal distribution with zero mean and constant variance [14].

As with GWR, bandwidth can be obtained at each location to determine local sample locations [15]. Observations at the local sample location will be weighted with kernel weights. Then weighting is performed for all time periods. At local sample points, it is assumed that panel data can be aggregated into a single geographic space. Then the panel data estimation model can be applied to obtain predictor variable coefficients at certain locations [9].

2.2 Data and Research Variables

In this study, the data used were secondary data sourced from the Central Statistics Agency (BPS) and the Ministry of Health [16]. This study used balanced panel data consisting of time series data from March 2020 to August 2021 and cross section data covering 34 provinces in Indonesia, therefore the number of observations used in this study was 102 observation units. The data was divided into 3 periods where each period consists of 6 months as follows:

- a. Early 2020 is data from March 2020 to August 2020
- b. End of 2020 is data from September 2020 to February 2021
- c. Early 2021 is data from March 2021 to August 2021

Response (y) in this study is the number of positive cases of Covid-19 and the predictor variables used are:

- x_1 : Total population density
- x_2 : Percentage of elderly population
- x_3 : Percentage of households that have access to proper drinking water
- x_4 : GRDP of business sector based on current prices (billions of rupiah)
- x_5 : Number of villages/districts that have health centers

The selection of predictor variables in this study was based on previous research by Sifriyani et al. [6].

2.3 Stages of Data Analysis

Geographically Weighted Panel Regression (GWPR) was performed on the number of positive cases of Covid-19 in Indonesia from early 2020 to early 2021. The following are the stages in implementing the GWPR model.

1. Data exploration.
2. Testing the effect of the panel model using the Chow test and the Hausman test.
3. Testing the assumptions of the panel model.
4. Testing the spatial effect using the Breusch-Pagan Test.
5. Data transformation within estimator.
6. Modeling GWPR on panel data with Gaussian Kernel and Bisquare Kernel weights.
7. Interpretation of the results and conclusions.

3. RESULTS AND DISCUSSION

3.1. Data Exploration

The Covid-19 outbreak first emerged from the Wuhan area in China in December 2019 and in March 2020 it was confirmed that 2 Depok residents in Indonesia tested positive for Covid-19. The spread of this virus continues to expand in the archipelago. This epidemic is not only a national problem in a country but is already a global problem. **Figure 1** and **Figure 2** show the spread of positive cases of Covid-19 at the beginning and end of 2020 in 34 provinces in Indonesia.

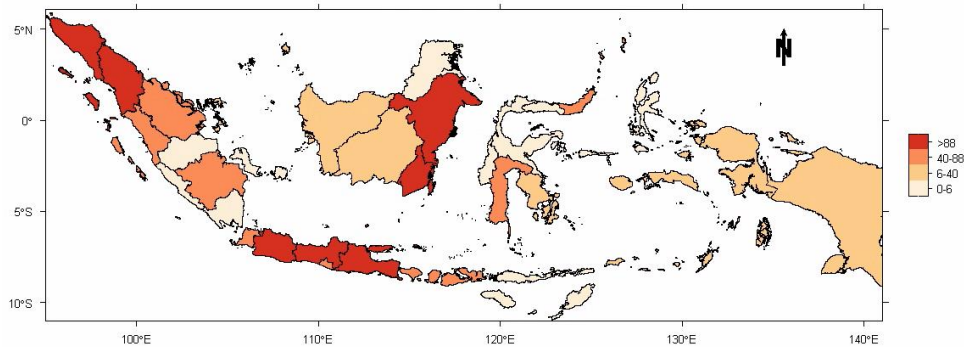


Figure 1. The spread of positive cases of Covid-19 at the beginning of 2020 in Indonesia

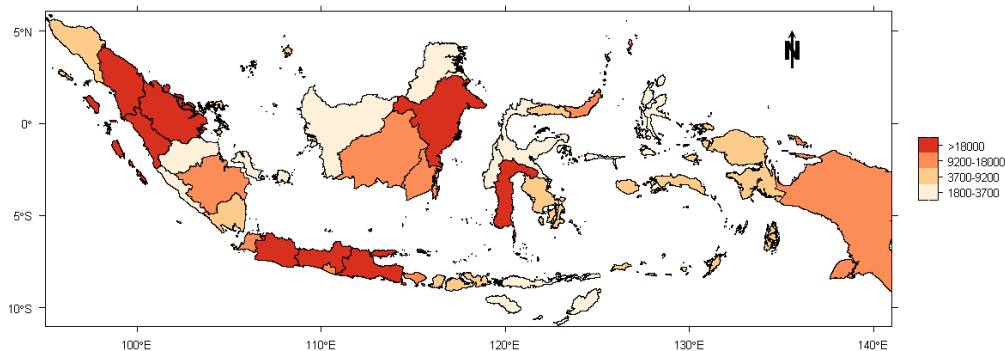


Figure 2. The spread of positive cases of Covid-19 at the end of 2020 in Indonesia

Figure 2 shows the drastic increase in the number of positive cases of Covid-19. The provinces with the highest number of Covid-19 cases from the beginning to the end of 2020 are located in DKI Jakarta, the provinces of West Java, Central Java, East Java, and East Kalimantan.

To contain the accelerated spread of Covid-19 in Indonesia, the government implemented Large-Scale Social Restrictions, and in January 2021 the Covid-19 vaccination program was carried out in Indonesia. **Figure 3** shows that there is still an increase in the number of positive cases, although the increase is not so drastic in several provinces, such as the Provinces of Maluku, Gorontalo, West Sulawesi, and Aceh, in fact, there is one province that has experienced a reduction in the number of positive cases of Covid-19, namely North Maluku Province. Meanwhile, provinces with high population density, such as South Sulawesi

Province and several provinces on Java Island, continue to experience a high increase in the number of positive cases of Covid-19.

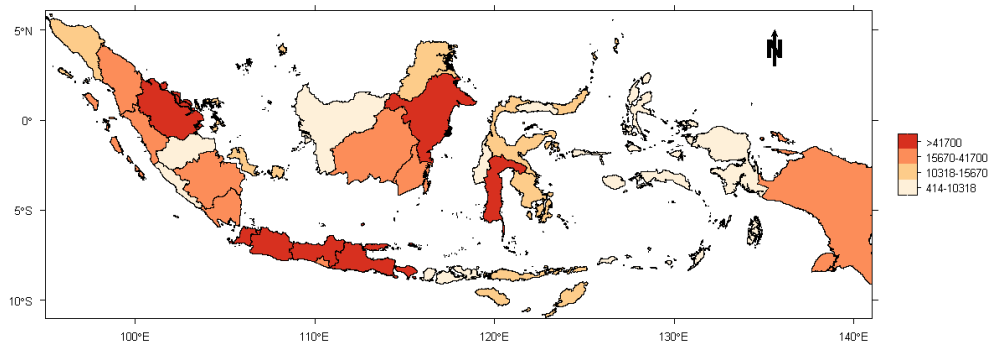


Figure 3. The spread of positive cases of Covid-19 at the beginning of 2021 in Indonesia

The descriptive of each predictor variable is shown in **Table 1** below:

Table 1. Data descriptive

		Number of positive cases (y)	Population density (x ₁)	Percentage of elderly population (x ₂)	Percentage of households with access to proper drinking water (x ₃)	GRDP of business field based on current prices (x ₄)	Number of villages/districts with community health centers (x ₅)
Min	Early 2020	0	9	3.34	61.22	36469	59
	End 2020	1896	9	5.19	57.6	39694	59
	Early 2021	414	9	3.63	62.47	41726	59
Median	Early 2020	40.5	102	7.51	86.12	182911	244
	End 2020	9142	103.5	11.99	86.61	187623	229.5
	Early 2021	15670	102.5	8.08	87.95	198185	236
Mean	Early 2020	97.29	734.59	7.755	83.52	440623	318.2
	End 2020	21386	742	12.65	84.23	471870	304
	Early 2021	48465	739.24	8.331	85.41	464285	308.2
Max	Early 2020	861	15764	13.97	96.38	2592607	1170
	End 2020	179660	15900	22.62	99.82	2816760	1071
	Early 2021	406205	15907	14.71	99.84	2772381	1094

Data source: Central Statistics Agency (BPS) and Ministry of Health

3.2. Panel Model Test

The panel model consists of 3 types of models, namely Fixed Effect Model, Random Effect Model, and Common Effect Model. The first test to be performed is Chow test which is used to compare the general effects model with the fixed effects model. The second test performed is the Hausman Test which is used to compare the random effects model with the fixed effects model. **Table 2** shows the results of the two panel model tests performed.

Table 2. Panel model testing

Testing	p-value	Selected model
Uji Chow	0.00156	FEM
Uji Hausman	9.363×10^{-6}	FEM

Table 2 shows the selected panel model is the Fixed Effect Model (FEM). **Table 3** shows the results of FEM panel regression model with significance level of 5%.

Table 3. Estimation result of FEM panel regression model

Variable	Estimate	SE	t-test	p-value	Explanation
Constant	-1.049×10^{-9}	1.609×10^{-2}	0.000	1.000000	Not significant
x_1	-8.87822	3.333539	-2.6633	0.009811	Significant
x_2	-0.02793	0.263869	-0.1058	0.916048	Not significant
x_3	0.551029	0.189231	2.9119	0.004962	Significant
x_4	-30.6465	6.710576	-4.5669	2.36E-05	Significant
x_5	5.848977	1.934046	3.0242	0.003604	Significant

3.3. Panel Model Assumptions Test

After the panel effect model was selected, the residual normality test of the model and the heterogeneity of variance test of the model were carried out. The normality test was carried out using the Shapiro-Wilk test and obtained a value of $W = 0.97522$ with a p-value = 0.05196. This indicates that the residual model is normally distributed. A heterogeneous test of variance of the model was carried out using the Breusch-Pagan test and obtained $BP = 19.722$ with p-value = 0.001409. This indicates that there is diversity in the data. Because the object is a province (location/spatial), it is concluded that there is spatial diversity in the panel data so that the data can be modeled using Geographically Weighted Panel Regression (GWPR).

3.4. Geographically Weighted Panel Regression Model

We transformed the data using within estimator because the selected panel model was FEM. Then, we performed GWPR using four kinds of bandwidth, namely, Fixed Gaussian Kernel, Adaptive Gaussian Kernel, Fixed Bisquare Kernel, and Adaptive Bisquare Kernel. **Table 4** shows the comparison of the GWPR models.

Table 4. Comparison of the models

Model	AIC	R ²
Panel Regression	-73.37139	0.598
GWPR with Fixed Gaussian Kernel	-314.721	0.982
GWPR with Adaptive Gaussian Kernel	-104.245	0.718
GWPR with Fixed Bisquare Kernel	-187.771	0.912
GWPR with Adaptive Bisquare Kernel	-196.284	0.914

Based on the AIC and R² values of the four GWPR models, the best GWPR model was the GWPR with Fixed Gaussian Kernel with the smallest AIC value of -314.721 and the largest R² value of 0.982. **Figure 5** shows that the predictor variables affecting the number of positive cases of Covid-19 are different at each location and tend to accumulate at certain locations.

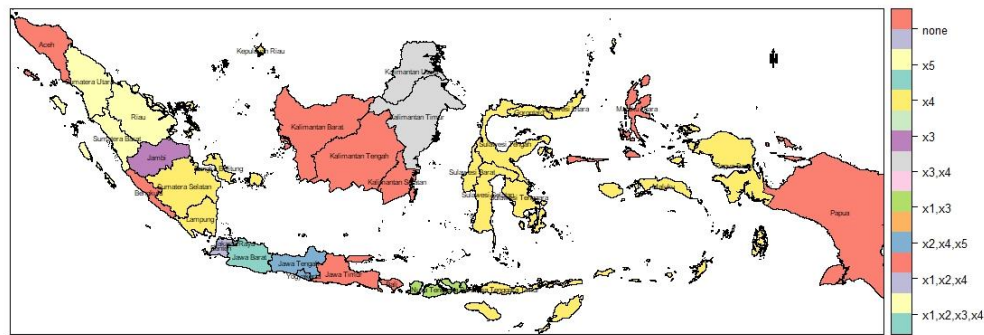


Figure 4. Map of distribution of predictor variables affecting Covid-19 in Indonesia

Figure 5 shows that the number of positive cases of Covid-19 in the provinces with the highest population density, namely West Java and DKI Jakarta, was affected by predictor variables x_1 (population density), x_2 (percentage of elderly population), x_3 (percentage of households having access to proper drinking water), and x_4 (GDP of the business sector at current prices). On the island of Sulawesi, the predictor variable that affected the number of positive cases of Covid-19 was only variable x_4 (GDP of business field at current prices), while in the provinces of Papua, Aceh, West Kalimantan, Central Kalimantan and South Kalimantan, there was no predictor variable that affected the number positive cases of Covid-19. The GWPR model based on significant predictor variables in several provinces is shown as follows:

$$\hat{y}_{\text{jawabarat}} = -2.38 \times 10^{-8} + 101.45x_1 + 3.43x_2 - 0.71x_3 - 327.20x_4$$

$$\hat{y}_{\text{DKI Jakarta}} = 6.24 \times 10^{-9} + 102.18x_1 + 3.40x_2 - 0.72x_3 - 335.64x_4$$

$$\hat{y}_{\text{Diyogyakarta}} = -1.20 \times 10^{-8} + 3.09x_2 - 101.58x_4 + 37.12x_5$$

4. CONCLUSIONS

Geographically Weighted Panel Regression can be used to model spatial data that takes time series into account. Based on the results of the panel model test on the number of positive cases of Covid-19 for 2020-2021 in 34 provinces in Indonesia, it was obtained that the selected panel model was Fixed Effect Model and based on the comparison of GWPR model with several weights, the best GWPR model was GWPR with Fixed Gaussian Kernel. The predictor variables that affected the number of positive cases of Covid-19 were different at each location and tend to accumulate at certain locations, hence the GWPR equation for each location was also different.

REFERENCES

- [1] M. T. Meehan *et al.*, "Modelling insights into the COVID-19 pandemic," *Paediatr. Respir. Rev.*, vol. 35, pp. 64–69, 2020.
- [2] K. Sarkar, S. Khajanchi, and J. J. Nieto, "Modeling and forecasting the COVID-19 pandemic in India," *Chaos, Solitons and Fractals*, vol. 139, p. 110049, 2020.
- [3] V. Soukhovolsky, A. Kovalev, A. Pitt, and B. Kessel, "A new modelling of the COVID 19 pandemic," *Chaos, Solitons and Fractals*, vol. 139, p. 110039, 2020.
- [4] S. Sifriyani and D. Rosadi, "Susceptible Infected Recovered (Sir) Model for Estimating Covid-19 Reproduction Number in East Kalimantan and Samarinda," *Media Stat.*, vol. 13, no. 2, pp. 170–181, 2020.
- [5] I. F. Mahdy, "Pemodelan Jumlah Kasus Covid-19 di Jawa Barat Menggunakan Geographically Weighted Regression," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 138–145, 2021.
- [6] S. Sifriyani, M. Rasjid, D. Rosadi, S. Anwar, R. D. Wahyuni, and S. Jalaluddin, "Spatial-Temporal Epidemiology of COVID-19 Using a Geographically and Temporally Weighted Regression Model," *Symmetry (Basel)*, vol. 14, no. 4, 2022.
- [7] N. Okthariza, "Kepadatan Penduduk dan Kasus Positif COVID-19 di Jakarta," *CSIS Comment.*, no. April, pp. 1–5, 2020.
- [8] E. Marhamah and I. G. N. Mindra Jaya, "Modeling positive covid-19 cases in Bandung City by means geographically weighted regression," *Commun. Math. Biol. Neurosci.*, vol. 2020, pp. 1–10, 2020.
- [9] D. Yu, "Exploring spatiotemporally varying regressed relationships: The geographically weighted panel regression analysis," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 38, pp. 134–139, 2010.
- [10] F. Bruna and D. Yu, "Geographically Weighted Panel Regression and Development Accounting for European Regions," *Int. Conf. Reg. Sci.*, pp. 1–20, 2016.
- [11] G. W. *Econometric Analysis*. New Jersey: Prentice-Hall Inc, 2000.
- [12] W. J. M., *Econometric Analysis of Cross Section and Panel Data*. London: MIT Press, 2002.
- [13] G. D. N. and P. D. C., *Basic Econometrics*. New York: McGraw-Hill/Irwin, 2009.

- [14] N. S. Rahayu, "Geographically Weighted Panel Regression for Modelling the Percentage of Poor People in Jawa Tengah Province," Thesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- [15] Z. Mar'ah, A. Djuraidah, and A. H. Wigena, "Semi-parametric Geographically Weighted Regression Modelling using Linear Model of Coregionalization," *Int. J. Sci. Basic Appl. Res.*, vol. 4531, no. July, pp. 178–186, 2017.
- [16] K. K. Ministry of Health, "Peta Sebaran dan Kasus COVID-19 di Indonesia," 2021. <https://in-422feksiemerging.kemkes.go.id/dashboard/covid-19/> (accessed Aug. 10, 2021).