International Journal of Social Science and Human Research

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 07 Issue 07 July 2024

DOI: 10.47191/ijsshr/v7-i07-50, Impact factor- 7.876

Page No: 5040- 5046

Spatial Geographically Weighted Regression (GWR) Model on Toddler Stunting in Java Island

Fita Airudelia[1], Ni Luh Putu Suciptawati[2], I Made Eka Dwipayana[3]

[1,2,3] Departement of Mathematics, Udayana University, Bali Indonesia,

ABSTRACT: Stunting is a condition where a child fails to grow properly in height for their age caused by long term chronic malnutrition, repeated infection and insufficient of psychosocial stimulation. Children with stunting are prone to poor of cognitive development and intelligent, metabolic disease, and lack of immune systems to prevent chronic diseases such as diabetes and cancer. Java is one of the most populous island in Indonesia with high prevalent of stunting in 2021. The high population density results in less access to the right to a healthy life, triggering the problem of stunting. Therefore, it is important to conduct research to determine the factors that influence stunting toddlers in districts/cities on the island of Java. The different percentage of stunted toddlers in each district/city on the island of Java shows that there are differences in the characteristics of each region due to the factors behind it, causing spatial heterogeneity. Therefore, the appropriate method used in this study is Geographically Weighted Regression (GWR). GWR is one of the developments of multiple linear regression models that take into account the location of the region. Spatial influence is considered in the construction of GWR model. The results showed that there were six groups of districts/cities based on predictor variables that had a significant effect on stunting toddlers, then exclusive breastfeeding and proper sanitation became the dominant predictor variables that had a significant effect on stunting toddlers in all districts/cities in Java Island.

KEYWORDS: Stunting, Geographically Weighted Regression, Java Island, Spatial Heterogeneity, Spatial Influence

I. INTRODUCTION

The proper growth and development of infants and toddlers are closely related to the quality of nutrition intake. Undernutrition and malnutrition occur due to inadequate intake of nutrients needed by the body. The content of foods such as carbohydrates, proteins, vitamins, and fats are examples of nutrients that play an important role in the growth process in toddlers [1]. The growth of toddlers will be disrupted if they get poor food quality [2]. A condition where a person is malnourished due to past nutritional status problems (chronic nutritional problems) is called stunting [3]. Malnutrition is an effect that occurs due to nutritional problems in the short or long term [4]. The growth and development of toddlers who experience stunting will be different from children in general. The impact of stunting is more visible in adulthood, such as reduced immunity (easily infected with diseases), decreased intellect, and decreased productivity which causes poor economic conditions [5]. The height-for-age index (TB/U) is the basis for measuring stunting in toddlers [6]. Stunted toddlers are rarely recognised due to negligence in monitoring growth such as measuring the height or length of toddlers [7].

The Ministry of Health of the Republic of Indonesia explained that the stunting rate in Indonesia is still ranked second in Southeast Asia. According to the World Health Organization (WHO), the global target for stunting prevalence is below 20% [8]. While the Indonesian government targets a reduction in the number of stunting cases in Indonesia to reach 14% by 2024 [9]. Based on the Indonesian Nutrition Status Study Report 2021, West Java and Banten (24.5%) were declared as provinces with the highest stunting rates in Java Island while DKI Jakarta (16.8%) had the lowest stunting cases. Seeing these conditions, cases of stunting in Java Island show that there are differences in each region due to the factors behind it, so the appropriate method used to analyse these cases is Spatial Regression Analysis [9]. Geographically Weighted Regression (GWR) is part of spatial regression analysis with weights based on geographic location and has the assumption of spatial diversity [10].

Previous research on stunting cases is [5] with the research title Determinants of stunting among under-five children in Ethiopia: a multilevel mixed effects analysis of 2016 Ethiopian demographic and health survey data. The results of research using different methods, namely A multilevel logistic regression model, show that factors at the individual and community levels determine stunting toddlers in Ethiopia. Next is Modelling of Risk Factors of Childhood Stunting Cases in Malang Regency using Geographically Weighted Regression (GWR) [11]. The difference with this study lies in the predictor variables used.

II. RESEARCH METHOD

The data used in this study are secondary data obtained from the Indonesian Nutrition Status Study (SSGI) 2021, the 2021 Health Profile for each province in Java Island, and the 2021-2022 District/City Stunting Special Index Report (IKPS). The response (dependent) variable in this study is the percentage of stunted toddlers and the variables that are thought to have an effect (predictor variables) include the percentage of vitamin A administration (X_1), the percentage of complete basic immunisation (X_2), the percentage of exclusive breastfeeding (X_3), the percentage of LBW (Low Birth Weight Babies) (X_4), the percentage of proper sanitation (X_5), the percentage of complementary foods (X_6), the percentage of proper drinking water (X_7), and the percentage of delivery assistance by health workers in health facilities (X_8). The coverage area in this study includes the Java Island region with a total of 85 districts and 34 cities.

In this study, GWR4 software and R software were used for data processing. The data analysis steps are as follows:

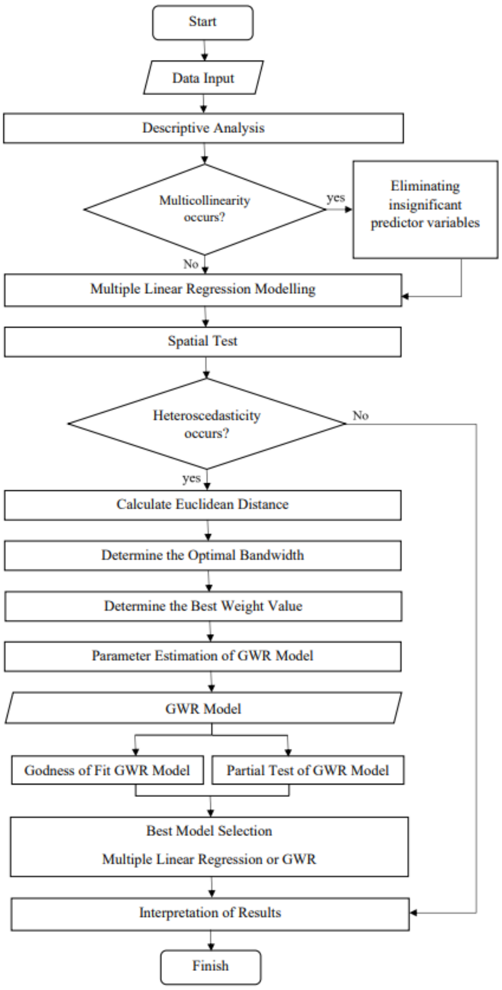

Figure 1. Flowchart of Data Analysis Steps

III. RESULTS AND DISCUSSION

A. Descriptive Analysis

Descriptive analysis was conducted to describe the condition of stunting toddlers in Java Island in 2021. The data presented in the table includes the highest and lowest percentages of stunting toddlers and suspected influencing factors.

Table 1. Descriptive Analysis

Variables	Minimum	Maximum	Mean
Y	6.90 (Mojokerto City)	38.90 (Bangkalan District)	21.2983
X_1	56.4 (Bangkalan District)	114.5 (Kep. Seribu District)	93.3467
X_2	24.3 (Sampang District)	100	76.3168
X_3	42.1 (Blitar District)	100	70.9688
X_4	0.3 (Jaksel, Jaktim, Jakbar City)	66.4 (Tuban District)	5.4233
X_5	39.4 (Bangkalan District)	98.8 (Tangsel City)	81.4588
X_6	50.5 (Surabaya City)	100	85.8647

X_7	73.3 (Lebak District)	100	94.4597
X_8	60.2 (Garut District)	100	94.5017

Table 1 shows that the highest percentage of stunting toddlers in Java in 2021 was 38.9% while the lowest percentage of stunting toddlers was 6.9%. Cases of stunting toddlers were highest in Bangkalan District and lowest in Mojokerto City. The high percentage of stunting toddlers in Bangkalan District is related to the low provision of vitamin A and the lack of access to proper sanitation. This is consistent with previous research which states that vitamin A supplements help the growth process in toddlers such as bone formation so as to reduce the incidence of stunting [12] and families who have problems with access to environmental sanitation such as inadequate latrines have a higher tendency to experience stunting [13].

B. Multicollinearity Test

Multicollinearity testing is important to determine whether the predictor variables have a high correlation. Multicollinearity test can be done using the VIF (Variance Inflation Factor) value criterion. If the VIF value is ≥ 5, it indicates multicollinearity between the predictor variables [14]. VIF value according to [15]:

$$\mathrm{VIF} = \frac{1}{1 - \mathrm{R}^2}$$

with R^2 is the coefficient of determination between predictor variables.

Table 2. VIF Values

Predictor Variables (X)	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	$\mathbf{X_5}$	$\mathbf{X_6}$	$\mathbf{X_7}$	$\mathbf{X_8}$
VIF	1.2251	2.1334	1.1468	1.0297	1.5346	1.1142	1.6301	2.4399

Table 2 shows the VIF values of the predictor variables are < 5, which indicates the absence of multicollinearity among the predictor variables.

C. Multiple Linear Regression Modelling

Multiple linear regression models are used when there are more than one predictor variables. The following is a multiple linear regression model [16].

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

with,
y_i : the response variable at observation-i (i=1,2,...,n)
β_0 : intercept
β_j : regression coefficient of the-j predictor variable
x_{ij} : value of the-j predictor variable at the-i observation
ε_i : error with IIDN assumption (identical, independent, normal) ~$N(0, \sigma^2)$
n : the number of observations.

The estimated multiple linear regression model for stunting toddlers in Java Island with a coefficient of determination of 49.25% is as follows.

$$\hat{Y} = 34.58321 + 0.05169X_1 + 0.09546X_2 + 0.09853X_3 + 0.04921X_4 + 0.22105X_5 + 0.04297X_6 + 0.09595X_7 + 0.05768X_8$$

D. Heteroscedasticity Test

The situation when one predictor variable responds differently at different locations in a research area is called spatial heterogeneity [17]. The difference in data characteristics allows the data to have a variety of variances. This diverse variance condition is called heteroscedasticity. The Breusch-Pagan test statistic is used to test for heteroscedasticity [18].
The test hypothesis is:

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_n^2 = \sigma^2$ (equal variance, homoskedasticity)
$H_1:$ at least one $\sigma_1^2 \neq \sigma^2$ (diverse variances, heteroscedasticity)

Test Statistics [18]:

$BP = \left(\frac{1}{2}\right) f^T Z(Z^T Z)^{-1} Z^T f$ with, the vector element f is $f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right)$ where,

f : is an n x 1 matrix

e_i^2 : the square of error for observation-i

σ^2 : the residual variance e_i

e_i : $y_i - \hat{y}_i$ is the least square residual for observation-i

ε_i : error with IIDN assumption (identical, independent, normal) $\sim N(0, \sigma^2)$

Z : is a matrix of size (n×(p+1)) containing the normalised vector for each observation.

Reject H_0 if $BP > X_P^2$ or if p-value < α with p being the number of predictors, which means that heteroscedasticity occurs in the model. The following are the test results using Breusch-pagan.

Table 3. Breusch-pagan Test

Breusch-pagan	P-Value	Description
17.244	0.02767	Reject H_0

Based on Table 3, it is shown that the P-value = 0.02767 < α = 0.05 so it is concluded that the data has a diverse variance which means it contains spatial heterogeneity. This results in the assumption of homogeneous error variance in multiple linear regression models not being met so that modelling using the Geographically Weighted Regression (GWR) method is required [19].

E. Creating a Geographically Weighted Regression (GWR) Model

One of the developments of multiple linear regression models is the Geographically Weighted Regression (GWR) model. It's just that the regression parameters of one location are different from the regression parameters of other locations because they consider spatial effects The following is the formulation of the GWR model [20].

$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k (u_i, v_i) x_{ik} + \varepsilon_i$; $i = 1, 2, \ldots, n$

with,

y_i : the observation value of the response variable for the-i location

x_{ik} : the observation value of the-k predictor variable at the-i observation location $k = 1, 2, \ldots, p$

$\beta_0(u_i, v_i)$: the intercept of GWR regression model

$\beta_K(u_i, v_i)$: the regression coefficient of the-k predictor variable at the-i observation location

ε_i : the error of the-i observation location (violating the assumption of homogeneous error).

In creating a GWR model, the first step is to calculate the euclidiean distance in each district/city-i to district/city-j in the coordinates (u_i, v_i) and (u_j, v_j) in Java Island. Then determine the optimum bandwidth value which will be used to calculate the weight matrix for each district/city. In determining the weighting value, this study uses the Adaptive Bisquare, Adaptive Gaussian, Fixed Bisquare, and Fixed Gaussian kernel weighting functions. After obtaining each model, the next step is to select the best GWR model by comparing the coefficient of determination (R^2) and Akaike Information Criterion (AIC).

Table 4. Comparison of GWR Model Weighting Function Results

Weighting Function	R^2	AIC
Adaptive Bisquare	**56.17%**	**678.3197**
Adaptive Gaussian	52.30%	684.7111
Fixed Bisquare	53.21%	683.2077
Fixed Gaussian	52.68%	684.0518

Table 4 shows that the Adaptive Bisquare weighting function produces the smallest AIC value followed by the largest coefficient of determination. Therefore, the GWR model with the Adaptive Bisquare weighting function is the best in modelling stunting toddlers in Java in 2021.

F. Godness of Fit GWR Model

This test aims to see if there is a significant difference between multiple linear regression and GWR. The test hypothesis is:
$H_0: \beta_k(u_i, v_i) = \beta_k, k = 1,2, \dots, p, i = 1,2, \dots, n$ (has no significant difference)
H_1: there is at least one $\beta_k(u_i, v_i) \neq \beta_k$ (has a significant difference)

Test Statistics [21]:

$$F_{count} = \frac{SSE\,(H_0)/df_1}{SSE\,(H_1)/df_2}$$

with,

$SSE\,(H_0)$	: $y^T(I-H)y$
$SSE\,(H_1)$	: $y^T(I-L)^T(I-L)y$
H	: $X(X^TX)^{-1}X^T$
df_1	: $n-p-1$ is the least square residual for observation-i
df_2	: $(n-2tr(S)+tr(S^TS))$
I	: identity matrix of size $n \times n$
L	: projection of GWR model value y into $\hat{y}$ at location (u_i, v_i)

Reject H_0 if $F_{count} > F_{table(\alpha; df_1, df_2)}$. Based on the results of the calculation of the best GWR model with Adaptive Bisquare weighting function, the value of F-Count = 2.04 > F-Table = 1.38 is obtained, so it is concluded that there is a significant difference between multiple linear regression models and GWR models.

G. Partial Test of GWR Model

This test aims to identify which parameters have a significant influence on stunting toddlers in Java Island.
The test hypothesis is:
$H_0: \beta_k(u_i, v_i) = 0$ (parameter $\beta_k(u_i, v_i)$ is not significant to the GWR model)
$H_1: \beta_k(u_i, v_i) \neq 0; i = 1,2, \dots, n;\ k = 1,2, \dots, p$ (at least one parameter $\beta_k(u_i, v_i)$ is significant to the GWR model)
Test Statistics [20]:

$$T_{count} = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{C_{kk}}}$$

with,

C	: $(X^TW(u_i, v_i)X)^{-1}X^TW(u_i, v_i)$
$\hat{\sigma}$	: $\sqrt{\frac{SSE\,(H_1)}{\delta_1}}$
δ_1	: $tr([(I-L)^T(I-L)]^i, i = 1,2$
df	: $\frac{\delta_1^2}{\delta_2}$
C_{kk}	: the-k diagonal element of the matrix CC^T

Reject H_0 if $|T_{count}| \geq t_{\frac{\alpha}{2}, df}$. The groups of districts/cities based on variables that have a significant effect on each district/city in Java Island with the Adaptive Bisquare weighting function are divided into six groups. The results show that geographical distance affects the grouping of districts/cities which are mostly in the same province, such as East Java Province which is in one group. The following is the parameter significance test for Tulungagung District which is included in the first group.

Table 5. Parameter Significance Test for Tulungagung District

Variables	$\hat{\beta}$	$SE\hat{\beta}$	t_{table}	t_{count}	Description
X_1	0.062483	0.058257	1.658697	1.07253	Not Significant
X_2	-0.14124	0.033815	1.658697	4.17679	**Significant**
X_3	0.093544	0.037663	1.658697	2.483719	**Significant**
X_4	0.010121	0.065686	1.658697	0.154083	Not Significant
X_5	-0.243755	0.03877	1.658697	6.28714	**Significant**
X_6	-0.03777	0.041659	1.658697	0.9066	Not Significant
X_7	0.113227	0.086891	1.658697	1.303082	Not Significant
X_8	0.046633	0.083748	1.658697	0.55682	Not Significant

Table 5 shows that the variables that significantly affect stunting toddlers in Tulungagung District are complete basic immunisation (X_2), exclusive breastfeeding (X_3), and proper sanitation (X_5).

H. Best Model Selection (Multiple Linear Regression Model or GWR Model)

In determining the best model between the Multiple Linear Regression model and the GWR model with the best weighting, namely Adaptive Bisquare for stunting toddlers in Java Island, it can be seen from the coefficient of determination and AIC value of each model [22]. The following is a comparison table of the coefficient of determination and AIC values.

Table 6. Coefficient of Determination and AIC Values

Model	R^2	AIC
Multiple Linear Regression	49.25%	701.2861
GWR (Adaptive Bisquare)	**56.17%**	**678.3197**

Based on Table 6, it can be seen that the coefficient of determination in the GWR (Adaptive Bisquare) model is greater than the Multiple Linear Regression model. Likewise, the AIC value shows the error rate of the GWR model is lower than the Multiple Linear Regression model. Thus the GWR model with the Adaptive Bisquare weighting function is the best model that can be used in modelling stunting toddlers in Java Island compared to the Multiple Linear Regression model.

I. Interpretation of Results

Based on the results of the best model in modelling stunting toddlers in Java Island, namely the GWR model with the Adaptive Bisquare weighting function, the next step is to interpret the GWR model. The following is an example of GWR model estimation in Tulungagung District which is in the first group.

$$\hat{Y}_{110} = 27.28116 - 0.14124X_2 + 0.093544X_3 - 0.243755X_5$$

The interpretation of the GWR model in Tulungagung District is that if complete basic immunisation (X_2) in Tulungagung District increases by one unit, the number of stunting toddlers will decrease by 0.14124 events with other variables considered constant. Then if exclusive breastfeeding (X_3) in Tulungagung District increases by one unit, the number of stunting toddlers will increase by 0.093544 events with other variables held constant. According to [23] it is often found that the relationship between exclusive breastfeeding and stunting is inconsistent because it is influenced by the quantity and quality of breast milk. In addition, it is necessary to ensure that infants six months and older get additional nutrition from complementary foods to complete a balanced nutritional intake. Furthermore, if proper sanitation (X_5) in Tulungagung District increases by one unit, the number of stunting toddlers will decrease by 0.243755 events with other variables held constant.

IV. CONCLUSIONS

Based on the analysis and discussion of the GWR model of stunting toddlers in Java Island, each district/city in Java Island has a different model, but there are similarities in predictor variables that have a significant effect on most of the adjacent districts/cities, such as districts/cities in East Java Province where the variables of complete basic immunization (X_2), exclusive breastfeeding (X_3), and proper sanitation (X_5) have a significant effect on stunting toddlers.

Furthermore, groups of districts/cities based on predictor variables that significantly affect stunting children under five in the GWR model with Adaptive Bisquare weights are divided into six groups. Exclusive breastfeeding and access to proper sanitation are the dominant variables that have a significant effect on stunting in all districts/cities in Java.

V. ACKNOWLEDGMENTS

Deepest gratitude to the Department of Mathematics, Udayana University for the memorable support in completing this research.

REFERENCES

1) Prakhasita, R. C. (2018). Hubungan Pola Pemberian Makan Dengan Kejadian Stunting pada Balita Usia 12-59 Bulan di Wilayah Kerja Puskesmas Tambak Wedi Surabaya. Skripsi, 1–119.
2) Menon, P., Headey, D., Avula, R., and Nguyen, P. H. (2018). Understanding the geographical burden of stunting in India: A regression-decomposition analysis of district-level data from 2015–16. Matern. Child Nutrition, 14(4), 1–10. doi: 10.1111/mcn.12620.
3) Khan, S., Zaheer, S., and Safdar, N. F. (2019). Determinants of stunting, underweight and wasting among children < 5 years of age: Evidence from 2012-2013 Pakistan demographic and health survey. BMC Public Health, 19(1), 1–15. doi: 10.1186/s12889-019-6688-2.
4) Bomela, N. J. (2009). Social, economic, health and environmental determinants of child nutritional status in three central asian republics. Public Health Nutrition, 12(10), 1871–1877. doi: 10.1017/S1368980009004790.
5) Fantay Gebru, K., Mekonnen Haileselassie, W., Haftom Temesgen, A., Oumer Seid, A., and Afework Mulugeta, B. (2019). Determinants of stunting among under-five children in Ethiopia: A multilevel mixed-effects analysis of 2016 Ethiopian

demographic and health survey data. BMC Pediatr, 19(1), 1–13. doi: 10.1186/s12887-019-1545-0.

6) Seboka, B. T., et al. (2022). Spatial trends and projections of chronic malnutrition among children under 5 years of age in Ethiopia from 2011 to 2019: a geographically weighted regression analysis. J. Heal. Popul. Nutr, 41(1), 1–17. doi: 10.1186/s41043-022-00309-7.
7) Sutarto., Mayasari, D., and Indriyani, R. (2018). Stunting, Faktor Risiko dan Pencegahannya. Foss. Behav. Compend, 5, 243–243. doi: 10.1201/9781439810590-c34.
8) de Onis, M., et al. (2018). Prevalence thresholds for wasting, overweight and stunting in children under 5 years. Public Health Nutr, 22(1), 175–179. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/52FB155B69DC75990CEFEE0C13A65A65/S1368980018002434a.pdf/prevalence-thresholds-for-wasting-overweight-and-stunting-in-children-under-5-years.pdf
9) Ahda, S. (2023). Pemodelan Kasus Stunting di Indonesia dengan Menggunakan Analisis Regresi Spasial, C, 1–23. http://scholar.unand.ac.id/215841/
10) Wu, X., and Zhang, J. (2021). Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). Environ. Sci. Pollut. Res, 28(32), 43732–43746. doi: 10.1007/s11356-021-13653-8.
11) Anismuslim, M., Pramoedyo, H., Andarini, S., and Sudarto. (2023). Modeling of Risk Factors of Childhood Stunting Cases in Malang Regency using Geographically Weighted Regression (GWR). Open Public Health J, 16(1), 1–11. doi: 10.2174/18749445-v16-e230420-2022-165.
12) Guantanamo, M. W. D. S. (2023). Hubungan Kelengkapan Pemberian Vitamin a Terhadap Stunting Pada Balita Di Puskesmas Lebdosari.
13) Indrastuty, D., and Pujiyanto. (2019). Determinan Sosial Ekonomi Rumah Tangga dari Balita Stunting di Indonesia: Analisis Data Indonesia Family Life Survey (IFLS) 2014. J. Ekon. Kesehatan. Indonesia, 3(2), 68–75. doi: 10.7454/eki.v3i2.3004.
14) O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673–690. doi: 10.1007/s11135-006-9018-6.
15) Hocking, R. R. (1997). Methods and Applications of Linear Models. Technometrics, 39(3), 332. doi: 10.2307/1271138.
16) Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). Applied Linear Statistical Models, 29(2). doi: 10.1080/00224065.1997.11979760.
17) Amin, R. A. (2021). Performa Bandwidth Cross Validation dan Akaike Information Criterion Corrected pada Model Geographically Weighted Regression (Studi Kasus : Jumlah Penduduk Kabupaten / Kota di Provinsi Sulawesi Selatan Tahun 2018).
18) Anselin, L. (2021). Spatial Models in Econometric Research. Oxford Res. Encycl. Econ. Financ. doi: 10.1093/acrefore/9780190625979.013.643.
19) Istifaiyah, I. (2018). Pemodelan Gizi Buruk Pada Balita di Provinsi Jawa Timur Menggunakan Geographically weighted Regression (GWR).
20) Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). Geographicallu Weighted Regression (the analysis of spatially varying relationships). https://www.academia.edu/33626785/Geographically_Weighted_Regression_The_Analysis_of_Spatially_Varying_Relationships_Wiley_2002
21) Fadhilah, N. (2015). Geographically Weighted Regression dan Spatial Pattern Analysis untuk Pemodelan Kejadian Penyakit Malaria.
22) Monalisa. K. A. (2022). Estimasi Angka Kematian Bayi di Indonesia Dengan GWR dan MGWR.
23) Hikmahrachim, H. G., Rohsiswatmo, R., and Ronoatmodjo, S. (2020). Efek ASI Eksklusif terhadap Stunting pada Anak Usia 6-59 bulan di Kabupaten Bogor tahun 2019. J. Epidemiologi. Kesehatan. Indonesia, 3(2), 77–82. doi: 10.7454/epidkes.v3i2.3425