Application of the Geographically Weighted Regression Method to the Human Development Index and Visualization on the Tableau Dashboard

Devi Octaviani Hasibuan[1*], Rokhana Dwi Bekti[2], Edhy Sutanta[3], I Wayan Julianta Pradnyana[4]

[1,2]*Department of Statistics, IST AKPRIND Yogyakarta, Indonesia*
[3,4]*Department of Informatics, IST AKPRIND Yogyakarta, Indonesia*

Article Information

Received: 10 November 2022
Revised: 30 November 2022
Published: 22 Desember 2022

Keywords
Human Development Index; Geographically Weighted Regression; Tableau

****Correspondence Email:***
hasibuandevioctaviani@email.com

Abstract

Spatial data is information that contains the location or geographic information of a region on a map of objects on Earth. One of the methods in spatial analysis is Geographically Weighted Regression (GWR). GWR is the development of the Ordinary Least Square (OLS) theory into a weighted regression model that takes into account spatial effects, resulting in a parameter estimator that can only be used to predict each point or location where the data is observed and concluded. The application of the GWR method is expected to produce an accurate Human Development Index (HDI) model. This study applies the GWR method to the Human Development Index in each Regency or city in Central Java in 2021. With a global GWR determination coefficient value of 0.8577, it means that 85.77% of population density, percentage of the poor population, gross regional domestic product at constant prices, and adjusted school enrollment rates have a global influence on HDI. From the GWR model obtained, it is known that the value of the HDI variable will decrease if the value of the PPM variable increases by one unit in each Regency or city. The value of the IPM variable will increase if the value of the PDRB, KP, and APS variables increases by one unit in each Regency or city. Therefore, Regency or city governments in Central Java Province are expected to be able to overcome the problem of poverty and reduce the percentage of poor people so that there is an increase in the HDI.

1. Introduction

The spatial analysis developed over the last few decades, consists of two main research fields, namely spatial data analysis and spatial modeling. Spatial statistics is a field of interest in geography-based statistics. The existence of spatial effects is something that often occurs between one region and another or the geographical location of a place using data.

Spatial data is data that contains information on the location or geography of an area, so it does not only contain what is measured. One method that can be used in spatial analysis is Geographically Weighted Regression

(GWR). GWR is the development of the Ordinary Least Square (OLS) theory into a weighted regression model taking into account spatial effects, resulting in a parameter estimator that can only be used to predict each point or location where the data is observed and concluded. GWR analysis is a method used to process spatial data. The GWR model is a model that takes into account geographical factors as independent variables that affect the response variable. The GWR model will generate a local parameter estimator for each point or location where the data is observed.

One of the problems that can be solved using the GWR method is the Human Development Index (HDI). HDI is an indicator used to see development progress in the long term. The achievement of human development can be measured by paying attention to three essential aspects, namely longevity and healthy life, knowledge, and a decent standard of living. According to the United Nations Development Program (UNDP), HDI is defined as a process of enlarging the choice of people. HDI measures the achievement of development results from a region/region in these three basic aspects of development. In general, Indonesia's human development continues to progress from 2010 to 2020. According to the Central Bureau of Statistics for Central Java Province, human development in Central Java in 2020 will experience progress, as indicated by an increase in the HDI. Even though it was affected by the Covid-19 outbreak, Central Java HDI in 2020 was still able to grow positively by 0.14 points, from 71.73 points in 2019 to 71.87 points in 2020.

The GWR method is a statistical method that is usually used on data that has spatial effects to model the diversity of relationships in spatial dimension visualization. Compared to global regression, the GWR method is able to model relationships that are weighted by the spatial component, namely distance. Spatial effects that occur between regions can be divided into two types, namely spatial dependence and spatial heterogeneity (Susanti, Lestia, & Sukmawaty, 2016). The fundamental thing of the GWR method is the proximity between regions, which is indicated by the weighting matrix. The closer the distance between regions, the greater the weight value. For this reason, the GWR method will be more accurate in statistically analyzing the spatial relationships of several variables, because it can overcome the problem of spatial diversity.

The application of the GWR method was carried out by Alfisyahrina (Alfisyahrina, 2021) who analyzed the factors that influence the Reading Literacy Activity Index in Indonesia. In this study, it was found that the best modeling of several regression methods is the GWR model, because it has a greater goodness-of-fit model than the linear regression model, which is equal to 92.46%. There is a significant influence on the literacy index number factors in Indonesia. In the grouping of variables that have a significant effect on the literacy activity index, there are 11 groups where in group 1 there is only one variable that has a significant effect on the literacy activity index, namely the percentage of Latin literacy in Papua Province. Meanwhile, in group 11, where all independent variables have a significant effect on the literacy activity index are in Jambi Province, South Sumatra Province, and Lampung Province.

Another study that applied the GWR method was conducted on Pneumonia Cases in East Java Province (Nila & Gde, 2019). In this study, it can be concluded that there is an influence of spatial heterogeneity aspects on pneumonia cases in East Java in 2016, so it is necessary to carry out an analysis using the GWR method. The results of the GWR analysis showed that the sum of the squared errors for the GWR model was smaller than the sum of the squared errors for the multiple linear regression model. This means that the GWR model is more appropriate for describing pneumonia cases that occurred in East Java in 2016.

Research with similar methods was carried out to analyze social vulnerability and its impact on social problems using methods in the city of Semarang (Hida, Sukmono, & Firdaus, 2020). The results of the analysis using this method show that the GWR model shows a positive relationship between social problems (Y) and population density (X1), the number of unemployed (X4), and the average length of schooling (X5), while a negative relationship with the sex ratio (X2) and life dependency rate (X3). The GWR model shows a significant level only for the population density factor with a value of t_count=2.065 ≥ t ((0.025;452))=2.059, and is not significantly different from the global regression model. However, the GWR method provides a better model with a higher coefficient of determination R2 of 0.326 and a lower RSS of 15.733. The application of the GWR method in this research is that the HDI case data in Central Java shows a spatial effect.

Visualization of the data will also help facilitate interpretation. In this case, visualization is defined as a method for presenting data or problems in a graphic format or image form that is easy to understand. The existence of data visualization will make it easier for readers to understand the information quickly and effectively conveyed by researchers using various attractive interactive graphics or images.

This study applies the GWR method with the hope of being able to produce an appropriate HDI model for each Regency/city in Central Java. Data is also visualized on the Tableau dashboard to provide a more attractive and easy-to-understand illustration.The food and beverage industry were hit hard by the COVID-19 outbreak. This includes offline food businesses, such as restaurants, cafés, and other dining spaces, which have been completely shuttered in some areas, while internet meal delivery services are still available. In addition, as customers race to stock their pantry shelves, the packaged food and beverage industries are seeing an increase in demand for shelf-stable goods and beverages, particularly milk products.

1.1 Literature Review

GWR is a spatial method involving the geographical conditions of each region as one of the factors thought to influence the dependent variable. GWR develops by adding geographic points at each location for each parameter. This development is based on the concept of non-parametric regression, which is applied to the general regression model. The GWR model obtained will be used to predict the magnitude of the response variable with the resulting parameters where each parameter is obtained from the location of the object. The basic thing of the GWR method is the proximity between regions, which is indicated by a weighting matrix. The closer the distance between regions, the greater the weight value. The general equation for GWR is as in equation (1).

$$y_i = \beta_0 (u_i, v_i) + \Sigma \frac{p}{t} = {}_1\beta_j(u_i, v_i)x_{ij} + \varepsilon_i \quad (1)$$

i=1, 2, …, n where:

y_i : the value of the dependent variable on the i-th observation

x_i : the value of the j-th independent variable in the i-th observation

$\beta_0(u_i, v_i)$: constant on the i-th observation

$\beta_j(u_i, v_i)$: the value of the function of the independent variable xj in the i-th observation

p : number of independent variables

(u_i, v_i) : the coordinates of the i-th observation location

ε : random error

The GWR model with an adaptive Gaussian Kernel weight is more suitable for modeling cases of malnutrition in children under five in West Java than the OLR model and the GWR model with a fixed Gaussian Kernel weight. This can be seen from the sum of the squared residuals of the GWR model with adaptive Gaussian Kernel weights and the coefficient of determination of the GWR model with Gaussian Kernel adaptive weights (Maulani, Herrhyanto, & Suherman, 2016).

Based on the SSE, R^2, and AIC values, a good model to use is to use a weighted near neighborhood kernel, whereas if based on a significant value with α = 5% it is known that a good model to use is to use a Kernel Bi-Square weight using 1 independent variable, namely the IPM variable because there are 17 regencies/cities, so it can be concluded that the best model in this study is to use bi-square kernel weights using 1 independent variable, namely the HDI variable because in the GWR model using near neighborhood kernel weights, none of the locations is significant (Dao & Kartiko, July 2019).

2. Research Methods

2.1. Data

The data used in this research is secondary data obtained from the Central Bureau of Statistics (BPS) of Central Java Province. The data used are HDI, Population Density (KP), Percentage of Poor Population (PPM), School Participation Rate (APS), and Gross Regional Domestic Product (PDRB) based on constant prices in 2021. The amount of data in each variable is 3

2.2. Research Variable

The variables used in this study are shown in Table 1.

Table 1. Research Variable

Variable	***Type***	***Operational Definition***	***Data Scale***
IPM	Dependent	Human Development Index by Regency/City in Central Java Province	Ratio
PDRB	Independent	Gross Regional Domestic Product Per capita at Constant Prices by Regency/City in Central Java Province (Million Rupiah)	Ratio
KP		Population Density by Regency/City in Central Java Province (Person/km2)	Ratio
PPM		Percentage of Poor Population by Regency/City in Central Java Province (percent)	Ratio
APS		School Participation Rates aged 7–18 Years by Regency/City in Central Java Province (Percent)	Ratio

2.3. Data Analysis Method

The method used in this research is descriptive analysis and spatial analysis of GWR, with the following stages:

1. Selecting variables that are suspected of influencing HDI to be involved in model building.
2. Identify descriptive analysis and spatial patterns of HDI variables to see the characteristics of the research data used through thematic maps.
3. Perform OLS analysis by testing assumptions.
4. Perform GWR analysis with the following stages:
5. election of the best model from the OLS model and the GWR model is based on the R2 and AIC values.
6. Dashboard visualization with Tableau.

Syntax used in R as in appendix.

3. Result and Discussion

3.1. Description of Characteristics and Spatial Patterns of Research Variables

Table 2 shows the average, minimum, maximum, median value, and standard deviation of each variable. The average HDI in 35 Regencys/cities in Central Java is 72.85. Regions that have a low HDI are Brebes Regency (66.32). The area that has a high HDI is the City of Salatiga (83.60). The average PPM in 35 Regencys/cities in Central Java is 11.393%. The region that has the lowest PPM is Semarang City (4.56%). The region that has the highest PPM is Kebumen Regency (17.83%). The average KP in 35 Regencies/Cities in Central Java is 2,102 people/KM2. The area that has a low KP is Blora Regency (491 people/KM2). The area that has a high KP is Surakarta City (11,361 people/KM2). The average GRDP in 35 Regencies/Cities in Central Java is 28.8 million rupiahs. The region that has a low GRDP is Pemalang Regency (12.75 million rupiahs). The region that has a high GRDP is Semarang City (87.36 million rupiahs). The average APS in 35 Regencies/Cities in Central

Java is 89.68 million rupiahs. The region that has the lowest GRDP is Brebes Regency (83.68%). The region that has a high GRDP is Magelang City (95.51%).

Table 2. Variable Descriptive Statistical Analysis

Variable	***Mean***	***Minimum Value***	***Maximum Value***	***Median Value***	***Standard***
IPM	72.85	66.32	83.6	72.36	4.45
PDRB	28.8	12.75	87.36	21.5	18.88
KP	2.102	491	11.361	1.124	2.421.8
PPM	11.393	4.56	17.83	10.68	3.56
APS	89.68	83.68	95.51	90.13	3.14

3.2. Regression Model Analysis

Table 3 shows results of estimating model parameters using OLS regression analysis produce parameter values.

Table 3. OLS Model Parameter Estimation Value

Parameter Estimator	***Parameter Estimator***	***Test Statistics t***		***Test Statistics F***	
		P-Value	$\lvert t_{count} \rvert$	***P-Value***	$\lvert F_{count} \rvert$
β_0	24.1553	0.0364	2.191	$2.098x10^{-11}$	38.46
β_1	0.0745	0.0055	2.992		
β_2	0.0005	0.0050	3.028		
β_3	-0.2688	0.0359	-2.197		
β_4	0.5412	$7.73x10^{-5}$	4.574		

Based on Table 3, it is known that the linear regression model that is formed is: IPM=24.1553+0.0745PDRB+0.0005KP-0.2688PPM+0.5412APS

The model can be explained as follows:

- If the independent variable is 0, then the CPI will experience an increase of 24.1553.
- When PDRB increases by one unit, then CPI will experience an increase of 0.0745.
- When KP increases by one unit, then CPI will experience an increase of 0.0005.
- When PPM increases by one unit, then IPM will experience a decrease of 0.2688.
- When APS increases by one unit, then IPM will experience an increase of 0.5412

Then Table 4 displays the *Residual Normality*, Table 5 displays *heteroscedasticity*, and Table 6 displays *autocorrelation*.

Table 4. Residual Normality

Test	***D Statistic***	***P-Value***
Kolmogorov-Smirnov	0.13083	0.5435

It is known that is residual data normally distributed with a 95% confidence level and H0 is rejected if the P-value<α. So that it can be obtained p – *value* = 0.5435 > a = 0.05 or D = 0.13083 < *Kolomogrov Smirnov*$_{(0,05;035)}$ = 0.244, then H_0 is not rejected or the assumptions are fulfilled. This means that the residuals are normally distributed with a 95% confidence level.

Table 5. Heteroscedasticity

Test		*P-Value*
Glejser	2.19	0.09016

It is known that is that there is no heteroscedasticity of the data with a 95% confidence level and H0 is rejected if the P-value<α. So that it can be obtained p – *value* = 0.09016 > a = 0.05 or $|F_{count}|$= 2.19 < $F_{(0.05;30)}$ = 2.69,then H_0 is not rejected or the assumptions are fulfilled. This means that there is no heteroscedasticity with a confidence level of 95%.

Table 6. Autocorrelation

Test	*DW Statistic*	*P-Value*
Durbin-Watson	0.1144	1.6686

It is known that H_0 is not autocorrelation with 95% confidence level and H0 is rejected if the P-value<α. p– *value* = 0.1144 > a = 0.05 atau 0 < DW = 1,6686 <dL = 1,2221 so H_0 is not rejected or assumptions are met. This means that there is no autocorrelation with a 95% confidence level.

In detecting the presence of multicollinearity is to look at the value of VIF (Variance Inflation Factor). A VIF value that is less than 5 indicates that there is no multicollinearity in the data. The multicollinearity test results are shown in Table 7.

Table 7. Multicollinearity Test VIF Value

Variable	*VIF Value*	*Conclusion*
PDRB	2.0551	There is no multicollinearity
KP	1.5522	There is no multicollinearity
PPM	1.7588	There is no multicollinearity
APS	1.2849	There is no multicollinearity

In the F test, it is known that H_0:$\beta_i = 0$ or there is no effect of the independent variable on the dependent variable with a 95% confidence level and H0 is rejected if $|F_{count}|$>$F_{(0.05;29)}$ Obtained $|F_{count}|$ = 38.46 > $F_{(0.05;29)}$ = 2.55 so that H_0 is rejected or the assumptions are fulfilled. This means that there is at least one effect of the independent variable on the dependent variable with a 95% confidence level. In the partial test, it is known that independent variables (PDRB, KP, PPM, APS) significantly influence the dependent variable (IPM) partially with a 95% confidence level and is rejected if P-value < α or $|t_{count}|$> $t_{(0.05;29)}$= 2.04. Based on Table 7, it can be concluded that the PDRB, KP, PPM and APS variables obtained P-value <α, then was rejected, thus the GRDP, KP, PPM and APS variables had a significant effect on the dependent variable (IPM) partially with a confidence level of 95%.

3.3. Spatial Effect Testing

The spatial dependency test used in this study is using Moran's I test with the output as shown in Table 8.

Table 8. Moran's I Output

Test	*Moran I Statistic*	*P-Value*	*Expectation*
Moran's I	0.2192	0.02955	-0.0294

It is known that H0 is that there is no spatial autocorrelation in HDI data with a 95% confidence level. H0 is rejected if the P-value < α. Obtained P-value = 0.2955 < α = 0.05, then H0 is rejected meaning that there is a spatial autocorrelation in the HDI data with a 95% confidence level. The value of I = 0.2192 is greater than E(I) = -0.0294, meaning that there is a positive but significant autocorrelation. A positive coefficient value also indicates that if an area has a high HDI, it will result in other neighboring regions having a high HDI as well. Geographical diversity can be seen by testing with spatial heterogeneity using the Breusch-pagan method.

Table 9. Spatial Heterogeneity

Test	*BP Statistic*	*P-Value*
Breusch Pagan	11.42	0.02223

It is known that H0 is that there is no spatial heteroscedasticity of the data with a 95% confidence level and H0 is rejected if the P-value<α. So that it can be obtained p-value = 0.02223 < α = 0.05 or *BP* = 11.42 > $X^2_{(0.05;4)}$ = 9.4877, so that H_0 is rejected or the assumptions are fulfilled. This means that there is spatial heteroscedasticity with a confidence level of 95%. In this study, the Queen Contiguity weight was used to determine the relationship between neighbors or adjacent locations, which would indicate a higher spatial dependency relationship than those located far apart (Tobler's First Law). Distance weighting is obtained from the latitude and longitude coordinates of a point or area.

3.4. Geographically Weighted Regression (GWR) Model Analysis

Location points u_i and v_i in this study were determined based on latitude and longitude for each Regency/city in the Central of Java Province. Longitude and latitude function to map the characteristics of variables in each Regency/city. Bandwidth is used as the basis for determining the weight of each observation of the GWR model based on the geographic location of Regencys/cities in Central Java. Table 10 shows the bandwidth values obtained from the results of using the Rstudio software.

Table 10. Bisquare Adaptive Kernel Bandwidth Results

Regency/city	***Bandwidth***	***Regency/city***	***Bandwidth***
Cilacap Regency	2.668391	Pati Regency	2.276911
Banyumas Regency	2.384422	Kudus Regency	2.099793
Purbalingga Regency	2.125499	Jepara Regency	2.099294
Banjarnegara Regency	1.893520	Demak Regency	1.835339
Kebumen Regency	2.042719	Semarang Regency	1.600524
Purworejo Regency	1.757408	Temanggung Regency	1.410482
Wonosobo Regency	1.680459	Kendal Regency	1.344281
Magelang Regency	1.415218	Batang Regency	1.618511
Boyolali Regency	1.761721	Pekalongan Regency	1.862148
Klaten Regency	1.804688	Pemalang Regency	2.082659
Sukoharjo Regency	2.005730	Tegal Regency	2.316963
Wonogiri Regency	2.244366	Brebes Regency	2.549546
Karanganyar Regency	2.170173	Magelang City	1.425687
Sragen Regency	2.090683	Surakarta City	1.960327
Grobogan Regency	2.070617	Salatiga City	1.615208
Blora Regency	2.531327	Semarang City	1.571014
Rembang Regency	2.668428	Pekalongan City	1.787245
Pati Regency	2.276911	Tegal City	2.347354

The results of determining the optimum bandwidth value with CV criteria produce different bandwidth values for each location. The optimum bandwidth value is used to obtain weights for each Regency. The next step is to calculate the Euclidean distance d_{ij} between locations (u_i, v_i), with the bandwidth value of each location using the Rstudio software. Equation (2) is a spatial weighting matrix based on adaptive kernel bisquare weighting

$$W_i(u_i, v_i) = \left[1 - \left(\frac{d_{ij}}{b_{i(q)}}\right)^2\right]^2 \quad (2)$$

Based on equation (2), the results of calculating the Euclidean distance and weights using the Rstudio Software. The model for each Regency/city in Central Java Province is shown in Table 11.

Table 11. GWR Model Estimation

Regency/City	***Intercept***	b_1	b_2	b_3	b_4
Cilacap Regency	26.1927	0.08729	0.00045	-0.2237	0.5073
Banyumas Regency	26.2088	0.08994	0.00044	-0.2281	0.50777
Purbalingga Regency	26.3874	0.09286	0.00043	-0.2309	0.50607
Banjarnegara Regency	26.3181	0.0963	0.00042	-0.2379	0.50821
Kebumen Regency	26.0845	0.09517	0.00042	-0.2426	0.51189
Purworejo Regency	26.6697	0.10137	0.00039	-0.2579	0.5083
Wonosobo Regency	26.3525	0.10009	0.0004	-0.2485	0.51015
Magelang Regency	26.9143	0.10148	0.00038	-0.2765	0.51045
Boyolali Regency	26.5644	0.07347	0.00052	-0.33	0.52622
Klaten Regency	27.6057	0.07944	0.00049	-0.3261	0.51327
Sukoharjo Regency	27.583	0.07213	0.00053	-0.3393	0.51632
Wonogiri Regency	28.2281	0.07132	0.00053	-0.345	0.50999
Karanganyar Regency	27.4051	0.06738	0.00056	-0.3463	0.51985
Sragen Regency	26.7007	0.06588	0.00057	-0.3437	0.52747
Grobogan Regency	25.7791	0.06494	0.00057	-0.3362	0.53663
Blora Regency	26.1651	0.06046	0.0006	-0.3455	0.53411
Rembang Regency	25.5958	0.05935	0.00061	-0.3407	0.53976
Pati Regency	24.9464	0.06219	0.00059	-0.331	0.54526
Kudus Regency	24.7425	0.06416	0.00058	-0.3263	0.54664
Jepara Regency	24.0573	0.06448	0.00058	-0.3178	0.55284
Demak Regency	24.5906	0.0687	0.00055	-0.3177	0.54661
Semarang Regency	25.7199	0.07783	0.0005	-0.3133	0.53259
Temanggung Regency	26.1207	0.09939	0.0004	-0.261	0.5163
Kendal Regency	25.4714	0.09364	0.00044	-0.2635	0.52446
Batang Regency	26.7418	0.09905	0.00042	-0.2377	0.50336
Pekalongan Regency	26.9013	0.09626	0.00043	-0.2315	0.50003
Pemalang Regency	27.0442	0.09332	0.00044	-0.2264	0.4974
Tegal Regency	27.0858	0.09059	0.00045	-0.2224	0.49626
Brebes Regency	27.0142	0.08832	0.00046	-0.2195	0.49671
Magelang City	26.8165	0.10156	0.00038	-0.2731	0.5108
Surakarta City	27.1824	0.07072	0.00054	-0.3392	0.52095
Salatiga City	26.042	0.07801	0.0005	-0.3164	0.52948
Semarang City	24.5038	0.07739	0.00051	-0.2995	0.54386
Pekalongan City	27.4393	0.09728	0.00043	-0.2303	0.49343
Tegal City	27.5858	0.09051	0.00045	-0.2201	0.48991

For example, the QWR model equation will be formed for regencies/cities in Central Java Province, then the model is:

$$IPM = 26.193 + 0.087PDRB + 0.00045KP - 0.2237PPM + 0.5073APS$$

The GWR model equation obtained in Cilacap district can be explained as follows:

- If the independent variable is 0, then the CPI will experience an increase of 26,193.

- When PDRB increases by one unit, then CPI will experience an increase of 0.087.
- When KP increases by one unit, then CPI will experience an increase of 0.00045.
- When the PPM increases by one unit, the IPM will experience a decrease of 0.2237.
- When APS increases by one unit, then IPM will experience an increase of 0.5073.

The model parameter significance test was carried out to see which independent variables affected the HDI in each regency/city in the Central Java Province. With the critical area stating that H_0 is rejected if $|t_{count}| > t_{(0.025;30)} = 2.04227$. The mapping of the variables that significantly influence HDI is shown in Fig. 1.

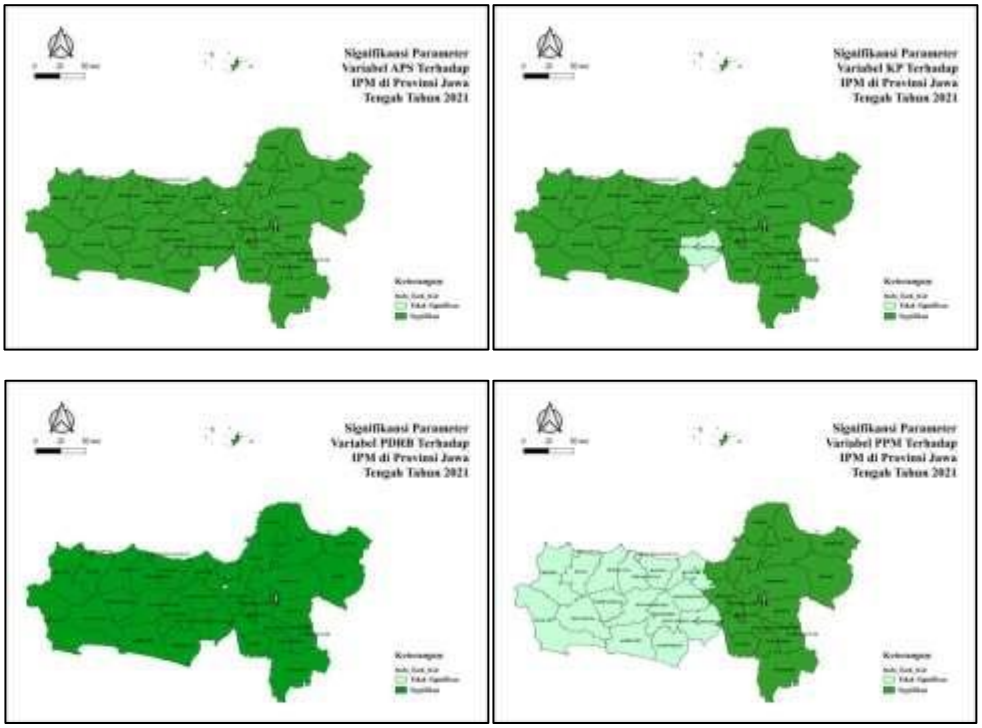

Fig. 1 Significance of Variable Parameters in Each Regency/City of Central Java Province in 2021

3.5. Comparison of Models

Selection of the best model is used to find out the best model that is good at estimating the chances of each model from the existing data. In this study, the two models (OLS and GWR) that have been obtained based on the and AIC. The results of the comparison of the two models are presented in Table 12. Based on Table 12, the results show that the GWR model has a larger compared to the OLS model, which is 0.8577 and the AIC is smaller than the OLS model, which is 142.2445. It is proven that the GWR model is good for use as a model in estimating model parameters.

Table 12. Comparison Result Value

Criteria	*OLS Model*	*GWR Model*
	0 8151	0 8577
AIC	169 3321	142 2445

3.6. Tableau Dashboard

The visualization that has been built is presented in Fig. 2 or can be accessed at bit.ly/3OHUWy9. This dashboard displays a scatter plot for each variable that affects the human development index of Central Java Province in 2021.

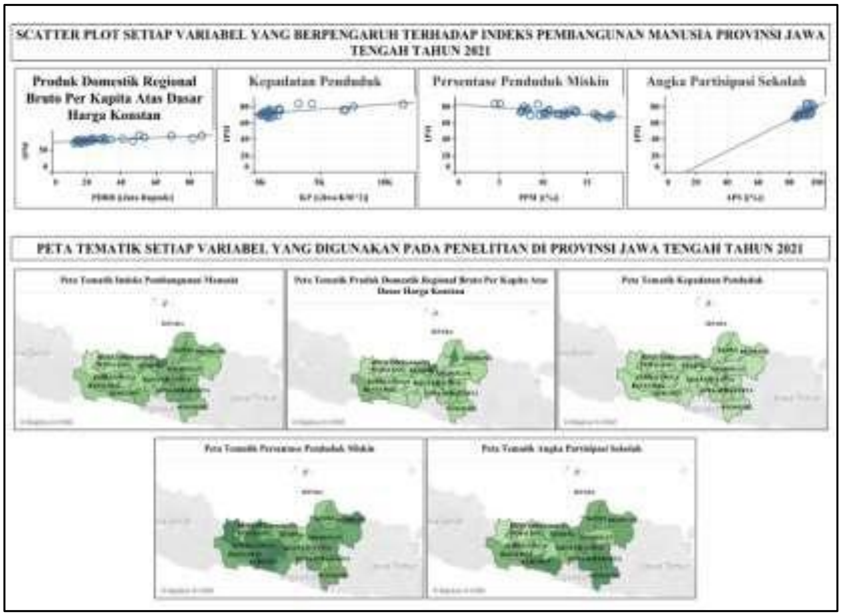

Fig. 2 Tableau Dashboard Visualization

4. Conclusions

Based on the results of the analysis using the GWR model with the adaptive kernel bisquare weighting function, information is obtained to answer the research problem, namely as follows:

a. Of the 35 regencies/cities in Central Java Province, it is known that each variable has a varying descriptive value in 2021. The spatial pattern of HDI achievements in Central Java Regencies/Cities in 2021 shows that there is a link between adjacent regions. This can be seen in the color differences in each district/city. In general, the darker the green color, the higher the HDI variable. Conversely, the faded green color on the map, the lower the HDI performance or tends to decrease.

b. Spatial effect testing in 35 regencies/cities in Central Java Province showed spatial dependencies and spatial heteroscedasticity occurred. The results of the study using the GWR model show that the variables GRDP, KP, PPM, APS significantly affect HDI, which means an increase in these four variables will increase HDI.

c. Testing the significance of the GWR model resulted in 35 GWR models for each district/city, where in general, the GRDP and APS variables affected the HDI in all districts/cities, the KP variable affected the HDI in 33 districts/cities and the PPM variable affected the HDI in 17 districts /city.

d. Selection of the best model seen in the value of the coefficient of determination and the AIC. The GWR model has a larger value than the OLS model, which is 0.8577 and the AIC is smaller than the OLS model, which is 142.2445, so it is proven that the GWR model is good to use as a model in estimating model parameters.

5. References

A, M. A., Nila, K. I., & Gde, S. I. (2019). Penerapan Metode Geographically Weighted Regression (GWR) pada Kasus Penyakit Pneumonia di Provinsi Jawa Timur. *E-Jurnal Matematika, 8(1)*, 27-34.

Alfisyahrina, D. H. (2021). Aplikasi Geographically Weighted Regression (GWR) Untuk Pemetaan Faktor yang Mempengaruhi Indeks Aktivitas Literasi Membaca di Indonesia. *Jurnal Riset dan Aplikasi Matematika, 5(2)*, 80-90.

Dao, A. F., & Kartiko. (Juli 2019). Analisis Geographically Weighted Regression Menggunakan Pembobot Adaptive Bisquare Dan Near Neighbourhood Kernel(Studi Kasus: Kemiskinan di Provinsi Sulawesi Tenggara tahun 2017). *JurnalStatistika Industri dan Komputasi*, Volume 4, No 2, 75 -93.

Hida, I., Sukmono, A., & Firdaus, H. (2020). Analisis Kerentanan Sosial Dan Pengaruhnya Terhadap Masalah Sosial Dengan Geographically Weighted Regression (GWR) (Studi Kasus:Semarang). *Jurnal Geodesi Undip, 9(1)*, 237-246.

Maulani, A., Herrhyanto, N., & Suherman, M. (2016). Aplikasi Model Geographically Weighted Regression (GWR) Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Kasus Gizi Buruk Anak Balita Di Jawa Barat. *EurekaMatika*, Vol.4, No.1.

Suryowati, K., Ranggo, M. O., Bekti, R. D., Sutanta, E., & Riswanto, E. (2021, July). Geographically Weighted Regression using Fixed and Adaptive Gaussian Kernel Weighting for Maternal Mortality Rate Analysis. In 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA) (pp. 115-120). IEEE.

Susanti, D., Lestia, A., & Sukmawaty, Y. (2016). Pemodelan Tingkat Kesejahteraan Peduduk Propinsi Kalimantan Selatan dengan Pendekatan Geographically weighted Regression (GWR). *Seminar Nasional MIPA2016.* Jatinangor: Peran Penelitian Ilmu Dasar dalam Menunjang Pembangunan Berkelanjutan.

Wagho, M. D., & Bekti, R. D. (2022). MODEL PERSAMAAN SIMULTAN UNTUK ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PERSENTASE KEMISKINAN DAN INDEKS PEMBANGUNAN MANUSIA DI PROVINSI JAWA TIMUR. Jurnal Statistika Industri dan Komputasi, 7(01), 60-70.

Wati, R. A., Suryowati, K., & Bekti, R. D. (2021, January). PEMODELAN ANGKA KEMATIAN IBU DENGAN PENDEKATAN GEOGRAPHICALLY WEIGHTED POISSON REGRESSION (GWPR) DI SULAWESI TENGGARA TAHUN 2016. In Prosiding Seminar Hasil Penelitian dan Pengabdian Masyarakat (Vol. 1, No. 1).

6. Appendix

Syntax in R

```
#Packages
library(spatial)
library(lattice)
library(spdep)
library(gstat)
library(maptools)
library(spgwr)
library(car)
library(ggplot2)
library(lmtest)
library(ape)
library(fields)
library(skedastic)
#Memanggil data
kabupaten<-readShapePoly("D:/DEVI/SEMESTER 7/KP DEVI/Data/Indo_Kab_Kot.shp")
plot(kabupaten)
```

```
datapeta<-data.frame(kabupaten)
#Statistika Deskriptif
summary(datapeta$IPM)
summary(datapeta$PDRB2)
summary(datapeta$KP)
summary(datapeta$PPM)
summary(datapeta$APS)

sd(datapeta$IPM)
sd(datapeta$KP)
sd(datapeta$PPM)
sd(datapeta$PDRB2)
sd(datapeta$APS)


#Analisis OLS
model=lm(IPM ~ PDRB2+KP+PPM+APS, data=datapeta)
summary(model)
#asumsi normalitas res=model$residuals ks.test(res,"pnorm",mean(res),sd(res),alternative = c("two.sided"))
#asumsi heteroskedastisitas
ujiglejser<-lm(abs(residuals(model))~datapeta$IPM+datapeta$PDRB2+datapeta$KP+
                                                   datapeta$PPM+datapeta$APS)
summary(ujiglejser)
#asumsi multikolinearitas
vif(model)
#asumsi autokorelasi dwtest(model)
#asumsi heterogenitas glejser(model)

#Titik ui dan vi c=coordinates(kabupaten) datapeta$long=c[,1] datapeta$lati=c[,2]
titik<-matrix(0,nrow=35, ncol=2)
titik
t.lokasi<-as.matrix(cbind(datapeta$KABKOT,datapeta$long,datapeta$lati))
t.lokasi

#menyusun pembobot/neighboring untuk nilai Moran's I
w<-poly2nb(kabupaten)
#untuk pembobot biner
ww1<-nb2mat(w, style="B",zero.policy = TRUE)
ww1
#untuk pembobot biner
ww2<-nb2mat(w, style="W",zero.policy=TRUE)
ww2
#rangkuman pembobot www<-nb2listw(w) www
Ketetanggaan<-data.frame(ww1)
Ketetanggaan

#Pengujian Efek Spasial
#Moran's I
moran<-moran.test(datapeta$IPM, alternative="two.sided", listw=www)
moran
#breusch Pagan
bptest(model,studentize = T,data=datapeta)

#Matriks Jarak Euclidean
```

```
#Jarak Euclidean
J.Euclidean<-as.matrix(dist(cbind(datapeta$long,datapeta$lati))) J.Euclidean<-data.frame(J.Euclidean)

#Penentuan Bandwidth Optimum GWR dengan Adaptive Kernel Bisquare
col.bw<-gwr.sel(datapeta$IPM                                                                  ~
datapeta$PDRB2+datapeta$KP+datapeta$PPM+datapeta$APS,coords=cbind(datapeta$long,datapeta$lati),
                    data=datapeta,adapt=TRUE,gweight = gwr.bisquare)

gwr<-gwr(datapeta$IPM   ~   datapeta$PDRB2+datapeta$KP+datapeta$PPM+datapeta$APS,data=datapeta,
    adapt=col.bw,coords=cbind(datapeta$long,datapeta$lati),
    gweight=gwr.bisquare,hatmatrix=TRUE,se.fit=TRUE)
#matriks pembobot untuk setiap kabupaten/kota
#Menyimpan pembobot bisquare bw<-gwr$bandwidth
col.bw2<-as.matrix(bw)
col.bw2

#Pembobot Matriks Fungsi Adaptive Kernel Bisquare
#matriks pembobot
matriks<-matrix(nrow=35,ncol=35)
matriks
for (i in 1:35) for (j in 1:35)
{
  matriks[i,1:35]<-
  (1-(J.Euclidean[i,j]/col.bw2[i,])^2)^2
}
matriks

#Estimasi summary(gwr) gwr$SDF data.frame(gwr$SDF) gwr$gweight print(gwr)

#Data Frame Model GWR
B0=gwr$SDF$`(Intercept)`
B1=gwr$SDF$`datapeta$PDRB2`
B2=gwr$SDF$`datapeta$KP`
B3=gwr$SDF$`datapeta$PPM`
B4=gwr$SDF$`datapeta$APS`
B1_se=gwr$SDF$`datapeta$PDRB2_se`
B2_se=gwr$SDF$`datapeta$KP_se`
B3_se=gwr$SDF$`datapeta$PPM_se`
B4_se=gwr$SDF$`datapeta$APS_se`
gwr_model=data.frame(B0,B1,B2,B3,B4,B1_se,B2_se,B3_se,B4_se)
gwr_model

#t_hitung t_PDRB=B1/B1_se t_KP=B2/B2_se t_PPM=B3/B3_se t_APS=B4/B4_se
t_hitung=data.frame(t_PDRB,t_KP,t_PPM,t_APS)
t_hitung
```