Proceedings of The International Conference on Natural Sciences, Mathematics, Applications, Research, and Technology (ICON-SMART)
e-ISSN: 2963-3915

Vol. 2, December 2022
pp. 23-29

Model of Human Development Index in West Nusa Tenggara Province Using Geographically Weighted Ridge Regression Method

Miftha Sukma Adi Prajanati [1, a)], Lisa Harsyiah [1, b)], and Nurul Fitriyani [1, c)]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Mataram, Mataram, Indonesia.*

[a)] mifthasukma3@gmail.com
[b)] lisa_harsyiah@unram.ac.id
[c)] Corresponding author: nurul.fitriyani@unram.ac.id

Abstract. The Human Development Index (HDI) is a comparative measurement of life expectancy, education, and living standards for all countries that aim to classify a country as a developed, developing, or underdeveloped country. The Regency/City Human Development Index (HDI) value in West Nusa Tenggara Province has increased in the last five years. It is necessary to analyze the factors that cause the increase in the value of HDI. This study aims to determine the HDI model with six variables that affect HDI. HDI data is included in spatial information obtained from several regions. This study used ten observation areas: West Lombok Regency, Mataram City, Central Lombok Regency, North Lombok Regency, East Lombok Regency, Sumbawa Regency, West Sumbawa Regency, Dompu Regency, Bima Regency, and Bima City. The Geographically Weighted Ridge Regression (GWRR) method was used in modeling the HDI value. This method can be used for data that have multicollinearity problems. The striking difference between the GWRR method and the multiple linear regression method is the presence of weights. Therefore, the Weighted Least Square (WLS) method was the parameter estimation method used, with the weighted fixed Gaussian. As a result of the differences in the observation area, the models produced in this study consist of ten different models. The modeling was done by considering the Mean Square Error (MSE) values and Akaike's Information Criterion (AIC). The results of the analysis carried out show that the variables that have multicollinearity problems are the variables X_1, X_4, X_5, and X_6. The GWRR model obtained gives the MSE value of the GWRR method of 0.158861 and the AIC value of the GWRR method of -87.36997, with variables that significantly affect the value of the HDI data, are the expenditure variable per capita and the average length of schooling.

Keywords: Human Development Index; Geographically Weighted Ridge Regression; Weighted Least Square.

INTRODUCTION

The Human Development Index (HDI) is a comparative measurement of life expectancy, education, and standard of living for all countries that aim to classify a country as belonging to the category of developed, developing, or underdeveloped countries. According to [1], three dimensions are indicators of development performance in HDI: knowledge, health, and a decent life. An index will give a category to a region in the form of higher and lower categories. However, the provision of categories with HDI is only macro-scale, which means that it cannot reflect the actual human situation.

Modeling is one way that can be used to identify factors that significantly affect HDI data. Later, the HDI model can help solve problems that result in low HDI scores since the HDI significantly affects the poverty level of an area [2], [3], [4]. One of the statistical methods that can be used in modeling is the regression method. Regression is one of the statistical methods commonly used to find the relationship between independent and dependent variables. For spatial data, one way that can be used for modeling is the Geographically Weighted Regression (GWR) method. The GWR method can be used in modeling spatial data because each parameter is calculated based on the observation location, so it has different parameters for each observation location [5].

Spatial data analysis consists of 3 groups, namely visualization (providing information related to the results of the spatial study), exploration (processing spatial data using statistical methods), and modeling (predicting spatial patterns using spatial or non-spatial data methods which aim to find out the existence of interaction processes) [6]. Ridge regression is a method that can operate on data that has multicollinearity problems. The multicollinearity test shows the presence or absence of a high correlation between dependent variables in a multiple linear regression model. A data is said to have multicollinearity if the value of the Variance Inflation Factor (VIF) > 10 [7]. Ridge regression is a development of the least squares method that minimizes the number of squares of errors by adding a bias setting. In spatial data with multicollinearity problems, the technique that can be used to overcome it is a mixed method between the GWR method and the ridge method known as the Geographically Weighted Ridge Regression (GWRR) method [8]. This study aims to determine the HDI model of regencies/cities in West Nusa Tenggara Province and determine the factors that cause the HDI value.

METHODOLOGY

The data used is data obtained from the Central Bureau of Statistics West Nusa Tenggara Province, with a total of 10 data obtained from 10 regencies/cities in West Nusa Tenggara Province. This study used the variables of the Human Development Index (HDI) value in each regency/city in West Nusa Tenggara Province (y_1), old school expectations (x_1), per capita expenditure (x_2), the average length of schooling (x_3), life expectancy (x_4), population (x_5), and number of poverty (x_6). The analysis steps used in HDI models using the GWRR method and determining significant variables in the HDI are as follows.

1. Data input. In this study, the data used was data on the HDI of regencies/cities in West Nusa Tenggara Province in 2020.
2. Normal distribution test of data using the Kolmogorov-Smirnov test.

   $$D_n = sup|\, S_n(y) - F_Y(y)|$$

   $S_n(y)$ is the empirical distribution function of the sample for all y and $F_Y(y)$ is the empirical distribution of a sample. If the normality assumption is not fulfilled, the data transformation is carried out first before the next step is carried out [9].
3. Descriptive statistics on the variables.
4. Spatial aspect test using the Moran's I assumptions.

   $$Z_i = \frac{I - E(I)}{\sqrt{Var\ (I)}}$$

   with,

   $$I = \frac{(X-\bar{X})^T W (X-\bar{X})}{(X-\bar{X})^T (X-\bar{X})}$$

   $$E(I) = I_0 = -\frac{1}{n-1}$$

   $$Var\ (I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2-1) S_0^2}$$

   Spatial heterogeneity test using the Breusch-Pagan test [5].

   $$BP = \frac{0.5 \times SSR}{\left(\frac{SSE}{n}\right)^2}$$

   SSR represents the sum of regression squares, and SSE represents the sum of squares of errors.
5. Multicollinearity test in the GWR model [7].

   $$VIF_k = \frac{1}{1 - {R_k}^2}$$

   ${R_k}^2$ is the coefficient of determination between independent variables of x_k and other independent variables.

6. Geographically Weighted Regression (GWR) model determination.

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ki} + \varepsilon_i$$

$\beta_0(u_i, v_i)$ represents intercept value of GWR models, $\beta_k(u_i, v_i)$ represent coefficient regression of the predictor the k-th variable at the i-th observation site, and (u_i, v_i) is coordinates of geographical location (latitude, longitude) at the i-th observation location [10].

7. Geographically Weighted Ridge Regression (GWRR) model determination.
8. Model goodness criteria determination using the Mean Square Error (MSE) and Akaike's Information Criterion (AIC) values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

and,

$$AIC = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left\{\frac{n + tr(L)}{n - 2 - tr(L)}\right\}$$

n represents the amount of data, y_i show the value of the dependent variable, and $\hat{y}_i$ represent the predicted results of y [5].

RESULT AND DISCUSSION

The spatial aspect tests are carried out, which include spatial heterogeneity and dependency tests. In the heterogeneity test using the Brunch-Pagan test, the statistical results of the BP test were obtained $(58.52) > \mathcal{X}^2_{0.05;6}$ (12.592) and in the spatial dependence test using Moran's I test, the results of the Z test statistics were obtained $(1.6103) > Z_{table}$ (0.4463). Therefore, there are cases of spatial heterogeneity and spatial dependence, which means that each observation area has different characteristics, and HDI data in the NTB province can be processed using the GWRR method.

Geographically Weighted Regression (GWR)

The first step in modeling using the GWR method is to determine the optimum bandwidth value using the CV criterion. The method used in determining bandwidth is the Fixed Gaussian method, and a bandwidth value of 1.318584 is obtained. This bandwidth value is used to determine the weighting value to be used in determining the estimated parameter. The estimation parameter used is the Weighted Least Square (WLS) method. The following are the GWR models obtained for the districts/ cities in the NTB Province.

1. Mataram City

$$\hat{y} = 11.6971 + 0.9649x_1 + 0.001x_2 + 1.2770x_3 + 0.3611x_4 - 1.1657e^{-07}x_5 + 0.0009x_6$$

2. West Sumbawa Regency

$$\hat{y} = 11.3952 + 0.9596x_1 + 0.001x_2 + 1.2800x_3 + 0.3667x_4 - 1.0395e^{-07}x_5 + 0.0008x_6$$

3. Dompu Regency

$$\hat{y} = 10.9705 + 0.9538x_1 + 0.001x_2 + 1.2813x_3 + 0.3742x_4 - 1.0065e^{-07}x_5 + 0.0009x_6$$

4. East Lombok Regency

$$\hat{y} = 10.7346 + 0.9523x_1 + 0.001x_2 + 1.2811x_3 + 0.3782x_4 - 9.7308e^{-08}x_5 + 0.001x_6$$

5. Central Lombok Regency

$$\hat{y} = 9.8094 + 0.9310x_1 + 0.001x_2 + 1.2836x_3 + 0.3949x_4 - 1.4124e^{-07}x_5 + 0.0017x_6$$

6. West Lombok Regency

$$\hat{y} = 10.4777 + 0.9461x_1 + 0.001x_2 + 1.2829x_3 + 0.3829x_4 - 1.0190e^{-07}x_5 + 0.0011x_6$$

7. Sumbawa Regency

$$\hat{y} = 11.1737 + 0.9509x_1 + 0.001x_2 + 1.2831x_3 + 0.3710x_4 - 1.1322e^{-07}x_5 + 0.001x_6$$

8. North Lombok Regency

$$\hat{y} = 11.2936 + 0.9637x_1 + 0.001x_2 + 1.2769x_3 + 0.3670x_4 - 9.9901e^{-08}x_5 + 0.0008x_6$$

9. Bima City

$$\hat{y} = 11.4493 + 0.9625x_1 + 0.001x_2 + 1.2788x_3 + 0.3655x_4 - 1.0119e^{-07}x_5 + 0.0008x_6$$

10. Bima Regency

$$\hat{y} = 11.3228 + 0.9429x_1 + 0.001x_2 + 1.2834x_3 + 0.3857x_4 - 1.0641e^{-07}x_5 + 0.0012x_6$$

From the model, further testing is carried out to determine the variables that significantly affect the model using the partial test. Here are the partial test results obtained, for example, in Mataram City.

TABLE 1. Parameter Test of GWR Method in Mataram City

Variable	Estimate	Error standard	Statistic *t*
Intercept	1.074	7.336	1.463
x_1	0.9467	0.3419	2.769*
x_2	0.009590	0.0001973	8.936*
x_3	1.279	0.1562	8.168
x_4	0.3783	0.1440	2.628*
x_5	-0.0000001418	0.000001383	-0.103
x_6	0.0014115624	0.01126	0.126

*significant

Therefore, it is concluded that the variables that significantly affect the HDI value in Mataram City based on the GWR models are x_1 (old school expectations), x_2 (per capita expenditures), x_3 (average length of schooling), and x_4 (life expectancy). Moreover, the following explanations show the variables significantly affecting the HDI in other districts/ cities.

1. Variable x_1

   Significant at Sumbawa Regency, West Sumbawa Regency, Bima City, Central Lombok Regency, East Lombok Regency, North Lombok Regency, West Lombok Regency, Bima Regency, and Bima City.

2. Variable x_2

   Significant at Sumbawa Regency, West Sumbawa Regency, Central Lombok Regency, East Lombok Regency, West Lombok Regency, Bima Regency, and Bima City.

3. Variable x_3

   Significant at Sumbawa Regency, Bima City, Central Lombok Regency, North Lombok Regency, West Lombok Regency, and Bima Regency.

4. Variable x_4

   Significant at Sumbawa Regency, West Sumbawa Regency, Bima City, West Lombok Regency, Bima Regency, and Bima City.

Before modeling using the GWRR method is carried out, multicollinearity testing is carried out on the data. The presence of cases of multicollinearity in the data is characterized by the value of the VIF > 10. The case of multicollinearity signifies that there is a relationship between dependent variables. Here are the results of the multicollinearity test conducted.

TABLE 2. Table of Variance Inflation Factor (VIF) values

Variable	VIF value
x_1	15.166*
x_2	7.728
x_3	9.672
x_4	13.244*
x_5	52.141*
x_6	62.809*

*significant

Table 2 explains the results of the multicollinearity test in the HDI data in the NTB province. The results show that four variables have multicollinearity cases. The variables are x_1 (old school expectations), x_4 (life expectancy), x_5 (population), and x_6 (the amount of poverty).

Geographically Weighted Ridge Regression

Similar to the GWR method, the GWRR method requires weighting in determining parameter estimates. The following are the GWRR models obtained for the districts/ cities in the NTB Province.

1. Mataram City

$$\hat{y} = 0.213934x_1 + 0.320902x_2 + 0.310682x_3 + 0.180463x_4 - 0.05127x_5 - 0.11328x_6$$

2. West Sumbawa Regency

$$\hat{y} = 0.178973x_1 + 0.368002x_2 + 0.326733x_3 + 0.165838x_4 - 0.05278x_5 - 0.08953x_6$$

3. Dompu Regency

$$\hat{y} = 0.163854x_1 + 0.408838x_2 + 0.317444x_3 + 0.163043x_4 - 0.04719x_5 - 0.07289x_6$$

4. East Lombok Regency

$$\hat{y} = 0.21457x_1 + 0.320011x_2 + 0.310011x_3 + 0.180829x_4 - 0.0513x_5 - 0.11385x_6$$

5. Central Lombok Regency

$$\hat{y} = 0.205287x_1 + 0.332203x_2 + 0.314548x_3 + 0.176646x_4 - 0.05201x_5 - 0.10812x_6$$

6. West Lombok Regency

$$\hat{y} = 0.194554x_1 + 0.346566x_2 + 0.320741x_3 + 0.171887x_4 - 0.0523x_5 - 0.10015x_6$$

7. Sumbawa Regency

$$\hat{y} = 0.164177x_1 + 0.392786x_2 + 0.331374x_3 + 0.161261x_4 - 0.05131x_5 - 0.07655x_6$$

8. North Lombok Regency

$$\hat{y} = 0.206824x_1 + 0.33043x_2 + 0.314724x_3 + 0.17677x_4 - 0.05142x_5 - 0.10833x_6$$

9. Bima City

$$\hat{y} = 0.175413x_1 + 0.407437x_2 + 0.287445x_3 + 0.173769x_4 - 0.04623x_5 - 0.07814x_6$$

10. Bima Regency

$$\hat{y} = 0.172003x_1 + 0.405602x_2 + 0.295034x_3 + 0.173282x_4 - 0.04455x_5 - 0.07631x_6$$

The deletion of β_0 value in the above models is caused by the scaling and centering processes that have been carried out.

Moreover, in simultaneous tests of the model, results obtained were statistic F by 210.3 and F table of 8.79. For partial test results are described as follows for an example in Mataram City.

TABLE 3. Influential Parameter in Mataram City

Parameter	Estimate	Error standard	Statistics t
x_1	0.16862	0.10395	1.622
x_2	0.17894	0.02404	7.444*
x_3	0.19021	0.03882	4.900*
x_4	0.26497	0.15798	1.677
x_5	-0.03500	0.02434	-1.438
x_6	0.03421	0.02803	1.221

*significant

Table 3 explains the variables that significantly affect the HDI model in Mataram City using the GWRR method. Two variables significantly affect the HDI, namely x_2 (per capita expenditure) and x_3 (average length of school). Moreover, the following explanations show the variables significantly affecting the HDI in each district/city.

1. Variable x_1

   Significant at East Lombok Regency, North Lombok Regency, West Lombok Regency, and Bima City.

2. Variable x_2

   Significant at Sumbawa Regency, West Sumbawa Regency, and East Lombok Regency.

3. Variable x_3

   Significant at Sumbawa Regency, Central Lombok Regency, West Lombok Regency, and Bima Regency.

4. Variable x_4

   Significant at West Sumbawa Regency and Bima City.

Furthermore, the methods used in selecting the best model in this research are the MSE and AIC methods. The criterion for the goodness of a model is if it has the smallest MSE and AIC value. In the end, the GWR and GWRR models were compared with the following results.

TABLE 4. Comparison Results of GWR Models 2ith GWRR Models

Model	MSE	AIC
GWR	3.42796×10^{12}	2.566153
GWRR	0.158861	-87.36997

Table 4 explains that the MSE and AIC values in models using the GWRR method are smaller than in models using the GWR method. So that the conclusion was reached that the model obtained using the GWRR method is better than the model obtained using the GWRR method.

CONCLUSION

From the results of the research and analysis that has been carried out, it can be concluded that there are ten different models in ten research observation areas in each regency/city in NTB province, with GWWR giving better results than the GWR method. In addition, the variables used have significant effects that differ between districts/ cities in the NTB Province.

ACKNOWLEDGMENTS

This work was supported by the Statistics Areas of Expertise, Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Mataram.

REFERENCES

1. Central Bureau of Statistics (BPS), (2021), Indeks Pembangunan Manusia (Human Development Index) (https://ntb.bps.go.id/), *accessed on* April 4th, 2021.
2. S. Novianto and H. Sudarsono, Jurnal Ekonomi Pembangunan **16(1)**, 1-12 (2018).
3. N. Hasanah, S. Bahri, and N. Fitriyani, "The Effect of Human Development Index on Poverty Model in Indonesia using Penalized Basis Spline Nonparametric Regression", in *International Conference on Science, Technology, Engineering and Industrial Revolution*, IOP Conference Series: Materials Science and Engineering 1115, (IOP Publishing Ltd, West Java, Indonesia, 2020), 012055.
4. Sapurah, I.G.E. Gunartha, and N. Fitriyani, Eigen Mathematics Journal **4(2)**, 60-67 (2021).
5. R.N. Hidayah, S.P. Wulandari, and S. Pramonom, Jurnal Sains dan Seni ITS **3(1)**, 146-151 (2014).
6. T. Wuryandari, A. Hoyyi, D.S. Kusumawardani, and D. Rahmawati, Media Statistika **7(1)**, 1-10 (2014).
7. D. Veronica, H. Yasin, and T. Widiharih, Jurnal Gaussian **5(3)**, 383-393 (2016).
8. T. Hastie, R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009), pp. 43-99.
9. J.D. Gibbons, and S. Chakraborti, *Nonparametric Statistical Inference* (Marcel Dekker Inc, New York, 2003), pp. 103-150.
10. A.S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (John Wiley & Sons, England, 2003), pp. 27-64.