# Geographically Weighted Lasso Method in Modeling The Gross Regional Domestic Product of The Bali-Nusa Tenggara Region

**Hairunnisa[1], Mustika Hadijati[2*], Nurul Fitriyani[3]**

[1] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Mataram, Mataram, 83125, Indonesia,* [2,3] *Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Mataram, Mataram, 83125, Indonesia*
*Corresponding Author: E-mail address:* mustika.hadijati@unram.ac.id

## ARTICLE INFO

## Abstract

**Introduction:** Statistics Indonesia announced that economic growth in 2020 is still in the negative zone, and the group of provinces in the Bali-Nusa Tenggara region has the most negligible impact on economic growth. The value of Gross Regional Domestic Product (GRDP) measures Indonesia's economic growth. GRDP is the total added value all regional business units generate at a particular time. **Background Problems:** This research aims to apply and interpret the results of the Geographically Weighted Lasso (GWL) method for GRDP in the Bali-Nusa Tenggara region. **Novelty:** Preliminary analysis in this study shows that the data used has the effect of spatial heterogeneity, which is a requirement for  modelling using the GWR method. In addition, there is a multicollinearity problem between independent variables. Therefore, the GWL method is used to solve the problem of spatial heterogeneity and multicollinearity in  modelling the GRDP of the Bali-Nusa Tenggara Region. **Research Methods:** The GWL method further develops the Geographically Weighted Regression (GWR) approach by adding the Least Absolute Shrinkage and Selection Operator (LASSO) method. The GWL method simultaneously selects insignificant variables by reducing the value of the regression coefficient to zero using the LASSO method. The data used has the effect of spatial heterogeneity and multicollinearity, a prerequisite for modelling with the GWL method. **Results:** Based on the analysis conducted, there are 41 different GRDP models for each district/city in the Bali-Nusa Tenggara region. The resulting GWL model provides a coefficient of determination of 95.84 % so that the resulting model can be used and is considered valid.

## 1. Introduction

In 2020, the Indonesian economy decreased by 2.07 % compared to 2019. Statistics Indonesia indicated that the Indonesian economy in the third quartile of 2020 was -3.49 %, meaning that economic growth in the third quartile of 2020 was still in the negative zone. When viewed spatially, all provincial economies in all islands in Indonesia also experienced negative growth. The most profound economic contraction was in the group of Bali-Nusa Tenggara (Bali-Nusra) provinces, with economic growth of -

6.80%. Based on this, the contribution of the economic growth of the Bali-Nusra region to the Indonesian economy is only 2.92%. The primary condition for the continued economic development of an area is the condition of high economic growth. According to Statistics Indonesia, a region's overall economic growth rate can be measured using Gross Regional Domestic Product (GRDP) at constant prices. GRDP preparation can be done through the Production, Expenditure, and Income approaches [1].

In a previous study, the factors used were the value of GRDP through the Production and Human Resources (HR) Approach, which consisted of 13 independent variables [2]. Based on this description, this study used factors that can affect GRDP using independent variables that have been studied previously by adding other factors, namely through the Expenditure Approach that is adjusted to the factors that affect GRDP in the Bali-Nusra region. A regression analysis model is needed to modelling the relationship between the GRDP of an area and the factors that influence it. The regression analysis used in this study is spatial regression because the data used is data that contains information on regional or location (spatial) related to the latitude and longitude coordinates of an area.

Geographically Weighted Regression (GWR) is a non-stationary method that models spatially varying relationships with coefficients in the GWR method, namely the function of spatial location. The GWR method aims to explore spatial heterogeneity in data relationships [3]. Several studies on the application of the regression model to spatial data have been carried out by previous researchers, including modelling the poverty rate in Central Java using the GWR model [4], the human development index model in West Nusa Tenggara using the Geographically Weighted Ridge Regression (GWRR) Method [5], the poverty model in the West Nusa Tenggara using the Geographically Weighted Logistic Regression (GWLR) model [6], and modelling the number of infant mortality in the East Lombok using Geographically Weighted Poisson Regression [7].

The drawback of the GWR method is that it has not been able to overcome cases of multicollinearity. Multicollinearity indicates a situation where there is a strong correlation between the independent variables [8]. The impact that can be caused by multicollinearity is the variance of the regression coefficient to be large. The magnitude of the variance can cause several problems, including the standard error, and the resulting interval will be large or wide. If the standard error is too large, the estimate of β will likely be insignificant. Therefore, other methods are needed to overcome the multicollinearity case [9].

The method to overcome the case of multicollinearity in the spatial model is the GWL method. The GWL method is an evolution of the GWR method, adding the LASSO method to the modeling. The objective of the GWL method is to overcome the problems of spatial heterogeneity and multicollinearity present in the least squares method [10]. The GWL method overcame the problem of spatial heterogeneity and multicollinearity in spatial data with a case study of food insecurity in Tanah Laut Regency [11]. This research results show that the GWL method performs better than the GWR method. Another study in the case of GRDP in West Java, using the GWL method with weighting the Fixed Exponential Kernel kernel function, was able to overcome the problems of spatial heterogeneity and multicollinearity. The result is that the model obtained is feasible to use or can be said to be valid [2].

Preliminary analysis in this study shows that the data used has the effect of spatial heterogeneity, which is a requirement for modeling using the GWR method. In addition, there is a multicollinearity problem between independent variables. Therefore, the GWL method is used to solve the problem of spatial heterogeneity and multicollinearity in modeling the GRDP of the Bali-Nusa Tenggara Region.

## 2. Material and Methods

This research aims to determine the GRDP model for the Bali-Nusra region and determine the factors that significantly influence the GRDP in the Bali-Nusra region using the GWL method. The method used is spatial regression analysis. The steps in this research are divided into the following stages:

### 2.1. Preparations

The preparations carried out in this research include setting the research topic, collecting as much information as possible from the literature related to the research, and then collecting the data used in the study. In this study, p represents the number of independent variables, n indicates the number of observations, $Y_i$ is the value of the dependent variable on $i$-th observation, $\beta_0$ is the constant, $\beta_j$ represents the regression coefficient of independent variable $X_j$, $X_{ij}$ indicates the value of $j$-th

independent variable on $i$-th observation, and $\varepsilon_i$ represents the error value which is assumed to be identical, independent, and normally distributed with zero means and $\sigma^2$ variance.

## 2.2. *Performing multicollinearity checking*

Multicollinearity is a condition where there is a high correlation between independent variables. The multicollinearity test aims to identify cases of multicollinearity in the independent variables using the $VIF$ test criteria. If the $VIF$ value is greater than 10, then multicollinearity occurs, whereas if the $VIF$ value range from $1 \leq VIF \leq 10$, then multicollinearity does not occur [9]. The VIF value can be obtained using Formula (1):

$$VIF_j = \frac{1}{1-R_j^2}; j = 1,2,\dots,p \tag{1}$$

where $R_j^2$ is the coefficient of determination of a multiple regression model with the $j$-th variable $(X_j)$ as the model's dependent variable and the other independent variables as its explanatory variables [11].

## 2.3. *Performing spatial heterogeneity test*

The spatial heterogeneity test used the Breusch-Pagan test formula. Tests with Breusch-Pagan were carried out to identify the data's spatial heterogeneity, i.e., there was a diversity of observational data between locations. If there is spatial heterogeneity in the model, then the requirements for testing using the GWR method are met [12]. The Breusch-Pagan test formula is shown in Formula (2).

$$BP = \left(\frac{1}{2}\right) \boldsymbol{f}^T \boldsymbol{Z} (\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \boldsymbol{f} \tag{2}$$

with,

$$f_i = \left(\frac{e_i^2}{\sigma^2} - 1\right), \ Z = \frac{x_i - x_i}{\sigma_{x_i}}, i = 1,2,\dots,n.$$

## 2.4. *Performing GWR modeling*

a. Calculating the Euclidean distance between the $i$-th observation location and the $j$-th observation location, which is defined in Formula (3).

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{3}$$

with $u$ latitude and $v$ longitude coordinates [13].

b. Estimating the value of bandwidth (b) with a kernel function that minimizes the value of cross-validation (CV). The mathematical equation of the CV value, according to [14], is in Formula (4).

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)]^2 \tag{4}$$

where $\hat{y}_{\neq i}$ is the estimated value of $y_i$ at the $i$-th observation location with a certain bandwidth removed from the estimation process [15].

The kernel function used in this research is the fixed exponential kernel function which is defined in Formula (5):

$$w_{ij} = exp\left(\frac{-d_{ij}}{b}\right) \tag{5}$$

where $d_{ij}$ is the Euclidean distance, and b is the bandwidth [13].

c. Forming a weighting matrix for each observation location using the bandwidth value that has been obtained previously. The weighting matrix used to estimate the parameters at each observation location is expressed in a diagonal matrix, with its elements being the kernel

function of each point of observation location. The form of the weighting matrix is shown in Formula (6):

$$W(u_i, v_i) = \begin{bmatrix} w_{(u_i,v_i)}^{i,1} & \cdots & 0 \\ 0 & \ddots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & w_{(u_i,v_i)}^{i,n} \end{bmatrix} \tag{6}$$

with $w_{(u_i,v_i)}^{in}$ is the value of the kernel function for the data at the $n$-th point in the model test at the $i$-th observation location.

d. Estimating GWR parameters using the WLS method by adding a weighting element in the estimate. The form of parameter estimation of the GWR model for each observation location is shown in Formula (7):

$$\widehat{\beta}(u_i, v_i) = (X^T W X)^{-1} X^T W y \tag{7}$$

e. Detecting cases of local multicollinearity in the GWR model. It can be seen from the local VIF value greater than 10 to determine the presence of local multicollinearity in the GWR model.

## 2.5. *Performing GWL modeling.*

The GWL method is a concept from the LASSO method applied in GWR modeling to overcome the problem of spatial heterogeneity and local multicollinearity. The parameter estimation in the GWL model uses the WLS method by adding a Lagrange multiplier function ($\lambda$) to the estimate. The value of $\lambda$ is an additional component in the LASSO regression that controls the magnitude of the shrinkage of the regression parameter value by producing a solution where the number of parameters is zero. The form of parameter estimation of the GWL model for each observation location is shown in Formula (8):

$$\widehat{\beta}(u_i, v_i) = (X^T W X + \lambda I)^{-1} X^T W Y \tag{8}$$

with $\lambda$ referred to as a tuning parameter that is useful for controlling the amount of shrinkage in the value of the regression parameter [8].

## 2.6. *Testing the sustainability of the model.*

Testing the suitability of the model using the coefficient of determination ($R^2$) with the equation in Formula (9):

$$R^2 = 1 - \frac{SSE}{SST} \tag{9}$$

$SSE$ is the sum of squared errors obtained by the formula $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ and $SST$ is the sum of the squared total obtained by the formula $\sum_{i=1}^{n}\left(y_i - y_i\right)^2$. The value of the coefficient of determination ranges from $0 \leq R^2 \leq 1$. That is if the value of $R^2$ obtained is close to 1, it can be said that the independent variable has a powerful influence on the dependent variable; in other words, the model used is good at explaining the effect of the variable [16].

## 3. Results And Discussion

It can be seen by looking at the $VIF$ value to determine whether multicollinearity occurs. If the $VIF$ value is $> 10$, then multicollinearity occurs, whereas if the $VIF$ value is around $1 \leq VIF \leq 10$, there is no multicollinearity. The results of the multicollinearity test are presented in Table 1:

**Table 1.** Rice coefficient for various climatic conditions

| Variables | VIF Value |
|---|---|
| $X_1$ | 30.7318 |
| $X_2$ | 11.5388 |
| $X_3$ | 4.4546 |

| Variables | VIF Value |
|-----------|-----------|
| $X_4$ | 3.6861 |
| $X_5$ | 4.5036 |
| $X_6$ | 17.5488 |
| $X_7$ | 2.9251 |
| $X_8$ | 1.7150 |
| $X_9$ | 13.2390 |
| $X_{10}$ | 60.7279 |
| $X_{11}$ | 90.6596 |
| $X_{12}$ | 10.7505 |
| $X_{13}$ | 33.1903 |
| $X_{14}$ | 73.5313 |
| $X_{15}$ | 10.8764 |
| $X_{16}$ | 62.3569 |
| $X_{17}$ | 65.2403 |

Table 1 shows that the VIF values that are less than 10 are $X_3, X_4, X_5, X_7$ dan $X_8$, while the other independent variables have a VIF value of more than 10. Therefore, it can be said that there is multicollinearity between the independent variables, so it needs to be overcome by the GWL method.

## 3.1. Spatial Heterogenity

A spatial heterogeneity test was conducted to identify whether or not there was a spatial effect on the data. If there is spatial heterogeneity in the regression model, then the requirements for completing the test using the GWR method are met to continue the research. The spatial heterogeneity test was carried out using the Breusch-Pagan (BP) test with the following hypothesis:

$H_0 : \sigma_1^2 = \sigma^2$ (no spatial heterogenity)

$H_1$ : at least one $\sigma_i^2 \neq \sigma^2$ (there is spatial heterogenity)

## 3.2. Geographically Weighted Regression (GWR) Modeling

Based on the results of the tests carried out, the BP value = 43.73, and the Chi-Square table value $\chi^2_{(p,\alpha)}$ (with $p = 17$ and α=0.05) =27.59. Because $BP > \chi^2_{(17,0.05)}$ then $H_0$ is rejected, meaning that there is spatial heterogeneity in the regression model. Therefore, the heterogeneity test is met, so the research can be continued to model the GRDP data for the Bali-Nusra region using the GWR spatial model.

The GWR model obtained is 41 models from 41 regencies/ cities in the Bali-Nusra region, so the parameter estimates from the model are very large. Therefore, the following is an example of a GRDP model in the Bali-Nusra region using the GWR method, one of which is the GRDP model for West Lombok Regency with model in Formula (10):

$$\hat{y}_1 = -3325.2670 + 0.2765x_1 - 0.0055x_2 + 43.1414x_3 + 66.2337x_4 - 5.8356x_5 - 0.1076x_6 +$$
$$0.0020x_7 + 0.9999x_8 + 0.3629x_9 - 15.9082x_{10} + 0.9629x_{11} + 0.3102x_{12} + 1.2276x_{13} +$$
$$1.3702x_{14} - 4.4293x_{15} + 1.9280x_{16} - 0.3204x_{17} \tag{10}$$

## 3.3. Local Multicollinearity

The next step is to identify local multicollinearity. It can be seen from the local VIF value greater than 10 to determine the presence of local multicollinearity in the GWR model. A summary of the total local VIF values for each observation location is presented in Table 2.

**Table 2.** Summary of local VIF values for all observation sites.

| Independent Variable | VIF > 10 |
|---|---|
| $X_1$ | 41 |
| $X_2$ | 34 |
| $X_3$ | 0 |
| $X_4$ | 0 |
| $X_5$ | 0 |
| $X_6$ | 41 |
| $X_7$ | 0 |
| $X_8$ | 0 |
| $X_9$ | 41 |
| $X_{10}$ | 41 |
| $X_{11}$ | 41 |
| $X_{12}$ | 41 |
| $X_{13}$ | 41 |
| $X_{14}$ | 41 |
| $X_{15}$ | 25 |
| $X_{16}$ | 41 |
| $X_{17}$ | 41 |

Based on the results of the local multicollinearity test, 12 independent variables have a VIF value greater than 10, with ten independent variables in which all observation locations have a VIF value greater than 10 and 2 other independent variables, respectively 34 and 25 observation locations have a VIF value greater than 10. greater than 10. The VIF value generated from the GWR model analysis is greater than that generated in the OLS method due to adding a weighting matrix to the parameter estimation. In addition, the results of local multicollinearity (with the GWR method) are similar to the results of global multicollinearity (with the OLS method), namely 12 independent variables have a VIF value greater than ten, and only five independent variables have a VIF value of less than 10. The results of the VIF values indicate a local multicollinearity problem in the GWR model. Therefore, the GWL method was applied to solve this problem.

## 3.4. Geographically Weighted Lasso (GWL) Modeling

The GWL method is a concept from the LASSO method applied in GWR modeling to overcome the problem of spatial heterogeneity and local multicollinearity. The regression coefficient generated by the GWL method will be depreciated to zero using the LASSO approach. Therefore, a coefficient that is zero does not affect the model; in other words, the GWL method selects variables that are not significant. Based on the parameter estimation results, 41 different models were obtained for each observation location. In other words, there were only significant variables in the obtained model. As an example of the results of the GRDP model for the Bali-Nusra Region using the GWL method, the following models are given:

1. GWL model for West Lombok, as an example for West Nusa Tenggara Province:

$$\hat{Y}_1 = 0{,}0090X_6 + 0{,}2565X_{11} + 0{,}1678X_{14} \tag{11}$$

Based on the model obtained, shown in Formula (11), it can be seen that the variables that significantly influence the value of GRDP in West Lombok Regency are per capita expenditure ($X_6$), construction ($X_{11}$), and household consumption expenditure ($X_{14}$).

2. GWL model for Jembrana, as an example for Bali Province:

$$\hat{Y}_{11} = 0{,}1495X_6 + 0{,}0802X_{13} + 0{,}3910X_{14} + 0{,}1886X_{17} \tag{12}$$

Based on the model obtained, shown in Formula (12), it can be seen that the variables that significantly influence the value of GRDP in Jembrana are per capita expenditure ($X_6$), provision of accommodation and food and drink ($X_{13}$), household consumption expenditure ($X_{14}$), and gross fixed capital formation ($X_{17}$).

3. GWL model for Alor, as an example for East Nusa Tenggara Province:

$$\hat{Y}_{20} = -0,0085 + 0,0064X_4 + 0,0334X_6 + 0,0137X_7 + 0,2527X_8 + 0,0097X_9 + 0,0080X_{10} +$$
$$0,0256X_{11} + 0,2755X_{13} + 0,6010X_{14} + 0,0687X_{16} \qquad (13)$$

Based on the model obtained, shown in Formula (13), it can be seen that the variables that significantly influence the value of GRDP in Jembrana are old-school expectations $(X_4)$, expenditure per capita $(X_6)$, agriculture, forestry, and fisheries $(X_7)$, mining and quarrying $(X_8)$, processing industry $(X_9)$, water procurement, waste management, waste, and recycling $(X_{10})$, construction $(X_{11})$, provision of accommodation and food and drink $(X_{13})$, household consumption expenditure $(X_{14})$, and government consumption expenditure $(X_{16})$.

After all the tests were carried out, 41 different models were obtained using the GWL method for each observation location, with the significant variables for each district/ city given in Table 3.

**Table 3.** Significant variables for each district/city

| No | District/ City | Significant Variables |
|----|----------------|----------------------|
| 1  | West Lombok    | $X_6, X_{11}, X_{14}$ |
| 2  | Central Lombok | $X_6, X_8, X_{13}, X_{14}, X_{17}$ |
| 3  | East Lombok    | $X_6, X_8, X_{10}, X_{13}, X_{14}, X_{17}$ |
| 4  | Sumbawa        | $X_6, X_{11}, X_{13}, X_{14}, X_{17}$ |
| 5  | Dompu          | $X_6, X_8, X_{10}, X_{13}, X_{14}, X_{17}$ |
| 6  | Bima           | $X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}, X_{17}$ |
| 7  | West Sumbawa   | $X_{11}$ |
| 8  | North Lombok   | $X_2, X_3, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 9  | Mataram City   | $X_6, X_8, X_{13}, X_{14}, X_{17}$ |
| 10 | Bima City      | $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 11 | Jembrana       | $X_6, X_{13}, X_{14}, X_{17}$ |
| 12 | Tabanan        | $X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 13 | Badung         | $X_6, X_8, X_{10}, X_{13}, X_{14}, X_{17}$ |
| 14 | Gianyar        | $X_6, X_8, X_{13}, X_{14}, X_{17}$ |
| 15 | Klungkung      | $X_2, X_3, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{13}, X_{14}, X_{15}, X_{16}$ |
| 16 | Bangli         | $X_2, X_3, X_5, X_6, X_7, X_8, X_9, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}$ |
| 17 | Karangasem     | $X_6, X_{11}, X_{14}$ |
| 18 | Buleleng       | $X_2, X_3, X_5, X_6, X_7, X_8, X_9, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}$ |
| 19 | Denpasar City  | $X_3, X_6, X_7, X_8, X_{10}, X_{13}, X_{14}, X_{16}$ |
| 20 | Alor           | $X_4, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 21 | Belu           | $X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 22 | Ende           | $X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 23 | East Flores    | $X_6, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}$ |
| 24 | Kupang         | $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 25 | Lembata        | $X_1, X_2, X_4, X_6, X_7, X_8, X_9, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}$ |
| 26 | Malaka         | $X_6, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 27 | Manggarai      | $X_6, X_7, X_8, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 28 | West Manggarai | $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 29 | East Manggarai | $X_6, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 30 | Nagekeo        | $X_2, X_3, X_4, X_7, X_8, X_9, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |

| No | District/ City | Significant Variables |
|----|----------------|----------------------|
| 31 | Ngada | $X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}, X_{17}$ |
| 32 | Rote Ndao | $X_1, X_2, X_4, X_6, X_7, X_8, X_9, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 33 | Sabu Raijua | $X_1, X_2, X_4, X_6, X_7, X_8, X_9, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 34 | Sikka | $X_2, X_3, X_4, X_5, X_7, X_8, X_9, X_{11}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 35 | West Sumba | $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}$ |
| 36 | Southwest Sumba | $X_2, X_3, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 37 | Sumba Tengah | $X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 38 | East Sumba | $X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}, X_{17}$ |
| 39 | Southern Central Timor | $X_6, X_8, X_{10}, X_{11}, X_{13}, X_{14}$ |
| 40 | Northern Central Timor | $X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, X_{16}$ |
| 41 | Kupang City | $X_6, X_{10}, X_{11}, X_{14}$ |

## 3.5. Model Fit Test

The goodness of fit test is used to determine the stability of the regression model obtained in representing the observational data, and it can be seen from the value of the coefficient of determination ($R^2$) obtained. The $R^2$ value generated from the GWL model is 0.9584, meaning that the independent variable in the model obtained using the GWL method is able to explain the diversity of the GRDP value at 41 observation locations in the Bali-Nusra region of 95.84%, with 4.16% explained by other variables not included in the model. Based on the results obtained, it can be said that the model obtained is feasible to use or is considered valid because it has a high $R^2$ value.

## 4. Conclusions

Based on the analysis that has been carried out, the GRDP model for the Bali-Nusra region based on the GWL method was obtained from as many as 41 different models, which can be declared suitable for use and considered valid based on the value of the determination coefficient obtained. Based on the model obtained for each district/ city, it can be seen that the variables that significantly influence the value of GRDP are different from each other. Therefore, it is necessary to provide different policies related to Gross Regional Domestic Product (GRDP), depending on the characteristics of the location.

## Ethics approval

Not required.

## Competing interests

All the authors declare that there are no conflicts of interest.

## Funding

## Underlying data

Data obtained via the Statistics Indonesia website https://bps.go.id/.

## References

[1] BPS, "Statistics Indonesia, https://bps.go.id/ (Accessed in 2022)."

[2] M. Kasyfurrahman, "Application of the Geographically Weighted Lasso Method in the Case of the Gross Regional Domestic Product of West Java ," Dissertation, Indonesian University of Education, 2020.

[3] B. Lu, M. Charlton, P. Harris, and A. Fotheringham, "Geographically Weighted Regression with A Non-Euclidean Distance Metric: A Case Study using Hedonic House Price Data," *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 660–681, 2014.

[4] M. Agustina, R. Wasono, and Darsyah MY, "Pemodelan Geographically Weighted Regression (GWR) pada Tingkat Kemiskinan di Provinsi Jawa Tengah (translate: Geographically Weighted Regression (GWR) Modeling at the Poverty Level in Central Java Province)," *Statistika*, vol. 3, no. 2, pp. 67–74, 2015.

[5] M. Prajanati, L. Harsyiah, and N. Fitriyani, "Model of Human Development Index in West Nusa Tenggara Province using Geographically Weighted Ridge Regression Method (GWRR)," in *Proceeding of The 3rd International Conference on Natural Sciences, Mathematics, Applications, Research, and Technology*, Bali, Indonesia, 2022.

[6] I. Sumarni, N. Fitriyani, and Z. Baskara, "Modeling of Factors Affecting Poverty in West Nusa Tenggara Province in 2020 With Geographically Weighted Logistic Regression," in *Mathematics National Conference (Konferensi Nasional Matematika) XXI (not published)*, Mataram, Indonesia, 2022.

[7] B. Justitiaski, N. Fitriyani, and S. Bahri, "Modeling the Number of Infant Mortality in East Lombok using Geographically Weighted Poisson Regression," *Eigen Mathematics Journal*, vol. 5, no. 2, pp. 100–108, 2022.

[8] A. Ramadhan, H. Pramoedyo, and R. Fitriani, "Perbedaan Metode Geographically Weighted Lasso (GWL)-Lokal dan Geographically Weighted Lasso (GWL) Global dalam Mengatasi Kasus Multikolinieritas pada Model Geographically Weighted Regression (GWR) (translate: Differences in Geographically Weighted Lasso (GWL)-Local and Global Geographically Weighted Lasso (GWL) Methods in Overcoming Multicollinearity Cases in Geographically Weighted Regression (GWR) Models)," *Jurnal Mahasiswa Statistik*, vol. 1, no. 2, pp. 93–96, 2013.

[9] R. Kurniawan and B. Yuniarto, *Analisis Regresi Dasar dan Penerapannya dengan R (translate: Basic Regression Analysis and Its Application with R)*. Jakarta: PT. Karisma Putra Utama, 2016.

[10] A. Setiyorini, J. Suprijadi, and B. Handoko, "Implementations of Geographically Weighted Lasso in Spatial Data with Multicollinearity (Case Study: Poverty Modeling of Java Island)," in *AIP Conference Proceedings, 1827(1):020003*, AIP Publishing LLC, 2017.

[11] J. Kim, "Multicollinearity and Misleading Statistical Results," *Korean J Anesthesiol*, vol. 72, no. 6, pp. 558–569, 2019.

[12] D. Desriwendi, A. Hoyyi, and Wuryandari, "Pemodelan Geographically Weighted Logistic Regression (GWLR) dengan Fungsi Pembobot Fixed Gaussian Kernel dan Adaptive Gaussian Kernel; Studi Kasus: Laju Pertumbuhan Penduduk Provinsi Jawa Tengah (translate: Geographically Weighted Logistic Regression (GWLR) Modeling with Weighted Functions of Fixed Gaussian Kernel and Adaptive Gaussian Kernel; Case Study: Population Growth Rate of Central Java Province)," *Jurnal Gaussian*, vol. 4, no. 2, pp. 193–204, 2015.

[13] R. Pamungkas, H. Yasin, and R. Rahmawati, "Perbandingan Model GWR dengan Fixed dan Adaptive Bandwidth untuk Persentase Penduduk Miskin di Jawa Tengah (translate: Comparison of the GWR Model with Fixed and Adaptive Bandwidth for the Percentage of Poor People in Central Java)," *Jurnal Gaussian*, vol. 5, no. 3, pp. 535–544, 2016.

[14] A. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex: John Wiley & Sons, 2002.

[15] M. Nadya, W. Rahayu, and V. Santi, "Analisis Geographically Weighted Regression (GWR) pada Kasus Pneumonia Balita di Provinsi Jawa Barat (translate: Geographically Weighted Regression (GWR) Analysis in Toddler Pneumonia Cases in West Java Province)," *Jurnal Statistika dan Aplikasinya*, vol. 1, no. 1, pp. 23–32, 2017.

[16] A. Widarjono, *Ekonometrika Teori dan Aplikasinya (translate: Econometrics Theory and Its Applications)*. Yogjakarta: Ekonisia, 2005.