Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 14(1) 2021: 89-97
http://ejournal.undip.ac.id/index.php/media_statistika

SPATIAL AUTOREGRESSIVE (SAR) MODEL WITH ENSEMBLE LEARNING-MULTIPLICATIVE NOISE WITH LOGNORMAL DISTRIBUTION (CASE ON POVERTY DATA IN EAST JAVA)

Dewi Retno Sari Saputro, Sulistiyaningsih, Purnami Widyaningsih
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Sebelas Maret

e-mail: dewiretnoss@staff.uns.ac.id

DOI: 10.14710/medstat.14.1.89-97

Article Info:
Received: 14 May 2020
Accepted: 22 June 2021
Available Online: 30 June 2021

Keywords:
Additive noise; *Ensemble*, *parameter estimation*; *SAR Model*

Abstract: The regression model that can be used to model spatial data is Spatial Autoregressive (SAR) model. The level of accuracy of the estimated parameters of the SAR model can be improved, especially to provide better results and can reduce the error rate by resampling method. Resampling is done by adding noise (noise) to the data using Ensemble Learning (EL) with multiplicative noise. The research objective is to estimate the parameters of the SAR model using EL with multiplicative noise. In this research was also applied a spatial regression model of the ensemble non-hybrid multiplicative noise which has a lognormal distribution of cases on poverty data in East Java in 2016. The results showed that the estimated value of the non-hybrid spatial ensemble spatial regression model with multiplicative noise with a lognormal distribution was obtained from the average parameter estimation of 10 Spatial Error Model (SEM) resulting from resampling. The multiplicative noise used is generated from lognormal distributions with an average of one and a standard deviation of 0.433. The Root Mean Squared Error (RMSE) value generated by the non-hybrid spatial ensemble regression model with multiplicative noise with a lognormal distribution is 22.99.

1. INTRODUCTION

Poverty is one of the major economic problems, especially in developing countries. Indonesia Central Bureau of Statistics (BPS) stated that poverty is measured based on the ability to meet fundamental needs, such as education, employment, per capita outcomes, health, and housing facility. As a developing country, Indonesia has a considerable poverty stage. In fact, BPS noted that through March 2015, the number of impoverished inhabitants reached 28.59 million or 11.22% of the total residents.

Java island is the largest contributor of the impoverished inhabitants population in Indonesia, where East Java ranks first (BPS of East Java Province)(Badan Pusat Statistik, 2004). Until March 2015, the population of impoverished inhabitants in this region reached 4.79 million or 12.34% of the total residents. Therefore, one of the attempts to cope with poverty is identifying the significant causative factors. Technically, the regression model can be used to figure out these factors.

The poverty level of an area is due to the influence of surrounding regions. Therefore, the spatial condition or closeness of the regions highly affects the poverty level. This corresponds with the Tobler Law, that everything is related, but near things are more related (Anselin, 2014; Cressie, 2015). Also, the existence of spatial effect indicates there is dependency on a region's poverty to others. Furthermore, the model which substitutes the spatial effect is known as Spatial Regression Model (Cressie, 2015). The models used are Spatial Autoregressive (SAR) and Spatial Error Model (SEM). In addition, the SAR is based on the spatial dependency existence on SEM temporary *lag* in its error.

The advancements in technology facilitate many amendments and analytic development. This does not occur with new algorithm, and is not easy or affordable on the computation side. Also, the development occurred on the final model paradigm used in carrying out prediction. In the computation era, a huge modeling barrier is considered, which is the idea that an applied algorithm is to use only one of many that are available. Therefore, this circumstance brings new ideas to predict not only based on one single model (which is considered as the best model), but also by combining the prediction results of several models (Saifudin, A. dan Wahono, 2016). This paradigm is known as Ensemble Learning (EL). According to Baba et al. (2015); Mevik et al. (2005); Vrrag & Nyitrai (2014), ensemble can raise the model accuracy. Hence, the principle is to raise the parameter estimation accuracy of the regression model. It also raises the prediction ability of several standard models (Zhu, 2012). Furthermore, ensemble learning leads to solution with more accurate prediction than single models found in several scientific journals. In fact, the ability of this approach is not only stated in many scientific papers but also in solved applicative cases as seen in data science competition on www.kaggle.com. This competition is freely opened to data science and mining users to offer predictive solutions according to the cases delivered by many international scale enterprises.

Based on current development, ensemble approach in predictive modeling is an exact choice for those who work on results with satisfying prediction. Besides the parameter estimation of regression spatial model, accuracy can be raised with resampling, a process of adding noise, which is an irregular disturbance on a data (Wu Z, 2005).

A single dataset provides several prediction models when different techniques or similar algorithm are used. Therefore, each model gives prediction, which could be different from one to another. Meanwhile, the EL approach combines different types of predictions into a final one. As a prediction model, the accuracy stage is an important factor to give an exact prediction. Furthermore, EL is capable of giving better classification and raising the prediction ability. Therefore, this study aims to estimate and apply the parameter model into Poverty in East Java data with Ensemble Non-hybrid spatial regression, Multiplicative Noise, and lognormal distribution.

2. LITERATURE REVIEW

Kim et al. (2003) conducted a study using ensemble technique and the result showed better accuracy on Support Vector Machine (SVM) compared to the non-ensemble techniques. Furthermore Canuto et al. (2005) applied the ensemble technique into spectrum data of Near-Infrared (NIR), hence the Partial Least Squares Regression (PLSR) became more robust than many types of added noise. Canuto et al. (2005) also applied the Ensemble Non-hybrid technique on two different databases, in which A is an image collection containing 7 outdoor images, and B is a promoter gene sequence, which are equally distributed over two classes (promoter and non-promoter).

Djuraidah (2012) used the SAR model to analyze factors that affect the poverty in East Java and declared that SAR is better than the regression model. Several applications with other poverty data use different approach models such as geographically weighted regression model in spatial data with multi collinearity problems and MARS (Fadliana et al., 2020); (Lembang et al., 2019). Even Saputro et al.(2019) used a nonparametric regression model approach with Fourier series approximation and penalized least squares (PLS) Fourier series approximation and penalized least squares (PLS). As well, Sinta et al. (2014) applied the Ensemble K-Nearest Neighbor (KNN) to determine a prediction model of rice price in Indonesia and resulted a high accuracy prediction. To solve the fundamental issue of supervised learning such as dimension decreasement. variance-biased and noise, are used the EL concept to design the optimised ensemble global model (Anwar H, Qamar U, 2014).

Rohmawati et al. (2015) applied the ensemble hybrid technique with additive noise on dengue fever (DBD) data in Central Java, and the intake used standard deviation calculation of dependent variable. Furthermore, Behera et al. (2016) explained that EL is a process with several models, obtained strategically and combined to certain computational intelligence problem solving. According to Lu et al. (2019) and Vrrag & Nyitrai (2014), ensemble learning is widely applied in areas such as web rank algorithm, classification and clustering, time series, and regression issues .

Spatial Autoregressive (SAR) is one of the models used based on spatial lag effects of the area approach. Accuracy stages of the SAR estimation parameter can be raised, especially to provide better results and decrease the error rate by the resampling method. Hence, the difference between previous literature and this study is focusing the Ensemble Non-hybrid Multiplicative Noise with lognormal distribution.

3. MATERIALS AND METHOD

3.1. Research Data

Secondary data of BPS and Health Office, East Java province was used. This includes the percentage of impoverished inhabitants (y) in 38 districts/cities in 2016, households using quality water source (X_1), health facility (X_2), household with the most spacious ground (X_3), inhabitants using good sanitary facilities (X_4), the total number of inhabitants (X_5), and the total number of hospitals (X_6).

3.2. Research Methods

The methods are explained as follows

1. Data exploratory using the poverty data in East Java.
2. Estimate the parameter of multiple linear regression using the Least Square Method.
3. Examine the parameter significance of multiple linear regression model.
4. Input the thematic maps of East Java, analyzing and interpreting the multiple linear regression model.
5. Determine the weighted spatial $\boldsymbol{W}$ matrix of the alluded $Queen$ matrix.
6. Spatial dependency test with I_M.
7. Examine is $|I_M| > Z_\alpha$?, if it is yes then continue to step 8 and if it is no then to the step 12.
8. Examine is $LM_\rho > \chi^2\alpha{:}\,1$? if it is yes, continue to step 9 and if it is no then to the step 10.
9. Estimate the SAR model parameter.

10. Examine is $LM_\lambda > \chi^2\alpha: 1$? if it is yes, continue to step 11 and if it is no then to the step 13.
11. Estimate the SEM parameter.
12. Initialize the multiple regression model and continue to step 25.
13. Obtaining the results that do not fulfil the SAR and SEM model continue to step 25.
14. Based on step 9 and step 11 then continue to step 15.
15. Determine and choose model with larger R^2
16. Initialize $r = 0$; $\sigma^2 = sd\ (ln\ y)$; $k =$ the amount of $resampling$.
17. Calculate $r = r + 1$.
18. Resurrect $noise\ z \sim \Lambda(1, \sigma^2)$.
19. Add $noise$ to y.
20. Determine the parameter estimation of spatial regression model on data that is noise added.
21. Examine is r=k ? if it is no, turn back into Step 17, if it is yes continue to step 22.
22. Calculate k spatial regression model.
23. Determine the mean of parameter estimation of k spatial model using ensemble technique
24. Analyze and interpret the ensemble spatial regression model.
25. Research method is completed.

Data structure that is shown in Table 1.

Table 1. Dependent and Independent Variable Data Structure

No.	y	X_1	X_2	X_3	X_4	X_5	X_6
1	y_{11}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
2	y_{21}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	y_{381}	x_{381}	x_{382}	x_{383}	x_{384}	x_{385}	x_{386}

Before the SAR model is constructed, due to EL that is described in the introduction, resampling for noise (ε_z) with $\varepsilon_z \sim LogN(0, \sigma^2)$is needed and its repeated 10 times so then data structure for dependent variable that are added with noise (y_k) are shown in Table 2.

Table 2. Dependent variable with Additive Noise Data Structure

No.	y	Noise	y_1
1	y_1	ε_1	y_{11}
2	y_2	ε_1	y_{21}
⋮	⋮	⋮	⋮
10	y_{10}	ε_1	y_{110}

⋮

No.	y	Noise	y_1
1	y_1	ε_q	y_{q1}
2	y_2	ε_q	y_{q2}
⋮	⋮	⋮	⋮
10	y_{10}	ε_q	y_{q10}

Using dependent variable $y_1, y_2, \dots y_{10}$, that can be constructed to 10 single SAR models is written in Table 3.

Table 3. Q Structures SAR Model

No.	SAR Model
1	$\hat{y}_1 = \rho_1 W y_1 + X\hat{\beta}_1$
2	$\hat{y}_2 = \rho_2 W y_2 + X\hat{\beta}_2$
$\vdots$	$\vdots$
10	$\hat{y}_{10} = \rho_{10} W y_{10} + X\hat{\beta}_{10}$

where **W** is a spatial weighted matrix explained as follows.

Lesage (1999) stated that alluded matrix is determined based on the information or the proximity of one area to another. In this case, the weighted matrix spatial data structure with area approach is the Queen alluded matrix which is described as

$$\boldsymbol{C} = (c_{11}\ c_{12}\ \ldots\ c_{1n}\ c_{21}\ c_{22}\ \ldots\ c_{2n}\ \vdots\ \vdots\ \ddots\ \vdots\ c_{n1}\ c_{n2}\ \ldots\ c_{nn})$$

C matrix is an angle side alluded matrix that define $c_{ij} = 1, i, j = 1,2, \ldots, n$, where n is the amount of observation for an area, in which common side or vertex are converged with concerned area and $c_{ij} = 0$ for another areas, where c_{ij} is the $\boldsymbol{C}$ element that states the weighted spatial size between the i area and j area.

Weighted spatial matrix is also stated as a matrix that describes the interaction strength between locations. This can be obtained through standardized $\boldsymbol{C}$ matrix. In the C matrix, value 1 shows the area which is neighboring to each other. Therefore, to notice how considerable the effect of each neighbor is towards an area, it can be calculated using the value of certain area with total values neighbor. Standardizing $\boldsymbol{C}$, each element is matched with (3).

$$w_{ij} = \frac{c_{ij}}{c_i} \tag{1}$$

Standardized alluded *Queen* matrix $\boldsymbol{C}$ is written as

$$\boldsymbol{W} = (w_{11}\ w_{12}\ \ldots\ w_{1n}\ w_{21}\ w_{22}\ \ldots\ w_{2n}\ \vdots\ \vdots\ \ddots\ \vdots\ w_{n1}\ w_{n2}\ \ldots\ w_{nn})$$

SAR using EL is a model obtained based on the mean parameter estimation results of 10 regression models

$$\hat{y} = \frac{1}{q}\sum_{k=1}^{10} \hat{y}_k \tag{2}$$

Where $\hat{y}$ is the mean of 10 model SAR estimation, q is the amount of resampling, and $\hat{y}_k$ is k-*th* SAR model. Therefore, SAR model with EL is generally written as

$$\boldsymbol{y} = \boldsymbol{\rho W y_k} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 I}) \tag{3}$$

4. RESULTS AND DISCUSSION

The results and discussion is in line with the aim explained in the introduction. SAR model with EL and additive noise was constructed in this study, the model parameter was estimated and applied to poverty data in East Java with Ensemble Non-hybrid, Multiplicative Noise, and lognormal distribution. Hence, the result and discussion are consistent.

4.1. SAR Model

General spatial regression model (Lesage 1999) is written as

$$\boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{u} \tag{4}$$
$$u = \lambda W + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I)$$

where y is dependent variable vector in size $n \times 1, \rho$ is spatial *lag* coefficient parameter, W is weighted matrix $n \times n$, X is independent variable matrix $n \times (p+1)$, β is parameter of regression coefficient $(p+1) \times 1$, and ε is a vector *error* sized $n \times 1$. Therefore, with every elements normally distributed, the mean is zero and the variance $\sigma^2 I$ along I is an identity matrix $n \times n$, u is an error vector $n \times 1$, λ is spatial error of coefficient parameter and p is the amount of regression coefficient parameters. Also, (4) is the SAR model when $\rho \neq 0$ and $\lambda = 0$. This model shows there is spatial effect on the dependent variable (Anselin, 1988), hence, the generally single SAR model is written as

$$\boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \text{ where } \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 I}) \tag{5}$$

Moran index is used to check when there is any significant spatial dependency on the given dataset (Griffith, 2012).

4.2. Ensemble Learning

Resampling process with Ensemble Learning (EL) proceeds by adding noise on data. Meanwhile, multiplicative noise is a type often recognized, and the result from dependent variable data on model (4) when noise is added is written as follows

$$\boldsymbol{y_k} = \boldsymbol{y} + \boldsymbol{\varepsilon_z} \tag{5}$$

Where y_k , $k = 1, 2, \ldots, q$ is the dependent variable vector when noise was added, with sized $n \times 1$ and $\varepsilon_z \sim logN(0, \sigma^2 I)$ obtained from the generating results in q repetitive.

According to De Bock & Coussement (2010), there are two types of EL, namely hybrid and non-hybrid ensemble techniques. Hybrid uses many spatial regression models and combines the estimation results of each into one final spatial model. Meanwhile, non-hybrid uses one spatial regression model which repeatedly uses results from different models and combine their estimated outcomes into one. In addition, non-hybrid ensemble is popularly used than the hybrid technique.

4.3. SAR Model Parameter Estimation with EL

The SAR single parameter on model (4) is estimated using maximum likelihood method. The likelihood of SAR model functions with dependent variable y

$$L(\boldsymbol{\rho}, \boldsymbol{\beta}, \boldsymbol{\sigma^2}; \boldsymbol{y}) = \left|\frac{\boldsymbol{I} - \boldsymbol{\rho W}}{(2\pi\sigma^2)^{n/2}}\right| exp\left(-\frac{(\boldsymbol{y} - \boldsymbol{\rho W y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{\rho W y} - \boldsymbol{X\beta})}{2\sigma^2}\right) \tag{6}$$

where $|\boldsymbol{I} - \boldsymbol{\rho W}| = \left|\frac{\partial(\boldsymbol{y} - \boldsymbol{\rho W y} - \boldsymbol{X\beta})}{\partial \boldsymbol{y}}\right|$

In line with maximum likelihood parameter estimation method, each estimation for σ^2 and β are obtained

$$\widehat{\boldsymbol{\sigma}}^{\boldsymbol{2}} = \frac{(\boldsymbol{y} - \boldsymbol{\rho W y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{\rho W y} - \boldsymbol{X\beta})}{n} \text{ and } \widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{I} - \boldsymbol{\rho W})\boldsymbol{y}$$

4.4. Application (Case on data's Poverty in East Java)

4.4.1 Multiple Linear Regression Model. The correlation between poverty and its causative factors was determined using multiple linear regression model. The multiple linear regression model for the poverty in East Java is written as

$$\hat{y}_{mkt} = 16.956 - 0.045\, X_1 - 0.046\, X_2 + 0.155\, X_3 - 0.018\, X_4 \tag{7}$$

$$+ 0.261\, X_5 \; - \; 0.186\, X_6$$

Model (7) fulfilled the normality, non-multicollinearity, and homoscedasticity assumptions. Based on model (7) parameter significance test, it was shown that the independent variable significantly affected X_3. This showed there is another impact which should be included in the regression model. The impact is known to be the spatial effect.

4.4.2. Spatial Dependency and Spatial Model Regression. Spatial dependency testing of model was carried out using Moran index (I_M). Based on the calculation, I_M = 0.397$ was obtained. There is a positive spatial dependency, hence, the contiguous areas have similar error values which tend to grouping. Furthermore, LM_ρ and LM_λ values were calculated to examine the existence of spatial dependency on lag and error. Based on the examination, the statistic test for each LM lag and LM $error$ are $LM_\rho = 9.698 > \chi^2_{(0.05;1)} = 3.84$ and $LM_\lambda = 9.517 > \chi^2_{(0.05;1)} = 3.84$. This means there is spatial dependency on lag and $error$. Therefore, the SAR and SEM model is constructed and written as

$$\hat{y}_{sar} = 0.390Wy + 3.796 + 0.165\, X_1 + 0.339\, X_5 - 0.236\, X_6 \quad (7)$$

$$\hat{y}_{sem} = 16.300 - 0.102\, X_1 + 0.168\, X_3 + 0.299\, X_5 - 0.198\, X_6 + u;$$

$$u = 0.563W \quad (8)$$

Based on the test results on model (8) and (9) errors, both fulfilled the normality and homoscedasticity assumptions. The R^2 values for each model are 67.16% and 73.94%.

Non-hybrid ensemble spatial regression uses one type of model. Meanwhile, two types (SAR and SEM) were obtained in this study. Therefore, ensemble technique was applied to spatial regression model with larger R^2, i.e. SEM. Technically, calculating with EL starts from generating the multiplicative noise as k = 10 repeatedly. This is because it is the ideal repetition count with the minimum error after several experiments. Also, the multiplicative noise generated a random variable and lognormal distribution of $\mu = 1$ and $\sigma = 0.433$, and the resampling process proceeded by adding the multiplicative noise on dependent variables. Hence, 10 SEMs were obtained from the resampling process as shown in Table 4.

Table 4. 10 SEMs Resampling Result

k	*SEM*
1	$\hat{y}_1 = 41.387 - 0.316X_1 + 0.410X_3 + 0.628X_5 - 0.051X_6 + u;\ u = 0.074Wu$
2	$\hat{y}_2 = 60.019 - 0.413X_1 - 0.176X_3 + 1.289X_5 - 0.633X_6 + u;\ u = 0.474Wu$
3	$\hat{y}_3 = 31.054 - 0.116X_1 + 0.701X_3 + 0.840X_5 - 0.622X_6 + u;\ u = 0.395Wu$
4	$\hat{y}_4 = 26.755 - 0.060X_1 + 0.637X_3 + 0.450X_5 - 0.377X_6 + u;\ u = 0.154Wu$
5	$\hat{y}_5 = 73.062 - 0.645X_1 + 0.887X_3 - 0.197X_5 + 0.196X_6 + u;\ u = -0.009Wu$
6	$\hat{y}_6 = 45.886 - 0.357X_1 + 0.829X_3 + 0.405X_5 + 0.047X_6 + u;\ u = -0.228Wu$
7	$\hat{y}_7 = 59.195 - 0.458X_1 + 0.725X_3 + 0.604X_5 - 0.413X_6 + u;\ u = -0.270Wu$
8	$\hat{y}_8 = 44.026 - 0.238X_1 + 0.316X_3 + 1.356X_5 - 0.961X_6 + u;\ u = 0.118Wu$
9	$\hat{y}_9 = 78.215 - 0.655X_1 + 0.957X_3 + 0.196X_5 - 0.411X_6 + u;\ u = 0.481Wu$
10	$\hat{y}_{10} = 34.954 - 0.106X_1 + 0.309X_3 + 1.031X_5 - 0.518X_6 + u;\ u = -0.122Wu$

The parameter estimation of non-hybrid ensemble spatial regression model was obtained through the mean of 10 models collected from the resampling process. Therefore, non-hybrid ensemble spatial regression model with multiplicative noise is written as

$$\hat{y} = 49.455 - 0.336X_1 + 0.559X_3 + 0.660X_5 - 0.374\, X_6 + u; \quad (10)$$

$$u = 0.092Wu$$

Based on the model (10), it can be interpreted that the increment of households that use quality sources of water is 10% and presumed that it will decrease the poverty percentage by 3 36%. Meanwhile, the increment of the household with the most spacious ground is 10% and presumed that it will increase the poverty percentage by 5.59%. The increment of inhabitants is 100 and presumed that it will increase the impoverished by 0.660%. In addition, the increment in number of hospitals is 10 units which decreases the percentage of impoverished inhabitants by 3.74%.

The regression model accuracy was calculated using RMSE value. This value on model (10) is 22.99, and the highest is presumed because there is an over-fitting that causes the prediction value to be higher than the actual.

5. CONCLUSION

Based on the results and discussion, it can be concluded that non-hybrid SAR with EL using a single type of spatial regression model was obtained from the mean of the combined SAR model estimation. Furthermore, single SAR model with EL from each of the data produced by resampling include the dependent variable with noise additive type. Also, SAR model parameter estimation was determined by the maximum likelihood method. While in practice, ensemble non-hybrid spatial regression, multiplicative noise, with lognormal distribution towards poverty in East Java for 2016 is written as model (10).

ACKNOWLEDGEMENTS

The authors are grateful to the Research Institute and community service of UNS for their support through grant and an agreement to carry out study on non-tax state revenue funds (PNBP) for the fiscal year 2020 number: 452 / UN27.21 / PN / 2020. The authors are also grateful to two anonymous reviewers for their valuable comments to revise the paper.

REFERENCES

Anselin, L. (2014). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anwar H, Qamar U, M. Q. A. (2014). Global Optimization Ensemble Model for Classification Methods. *The Scientific World Journal*.

Baba, N.M., Makhtar, M., Fadzli, S.A., dan Awang, M. . (2015). Current Issues in Ensemble Methods and Its Applications. *Journal of Theoretical and Applied Information Technology*, *81*(2), 266–276.

Badan Pusat Statistik. (2004). *Data dan Informasi Kemiskinan*.

Behera, R.N., Roy, M., dan Dash, S. (2016). Ensemble based Hybrid Machine Learning Approach for Sentiment Classification-A Review. *International Journal of Computer Applications.*, *146*(6), 0975–8887.

Canuto, A. M. P., L. Oliveira, J.C.X. Junior, A. Santos, M. A. (2005). Performance and Diversity Evaluation in Hybrid and Non-hybrid Structures of Ensembles. *Proceedings of the Fifth International Conference on Hybrid Intellegent Systems*.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

De Bock, K. W., K. Coussement, and D. V. D. P. (2010). Ensemble Classification Based on Generalized Additive Models. *Computational and Data Analysis*, *54*, 1535–1546.

Djuraidah, A. dan A. H. W. (2012). Regresi Spasial untuk Menentukan Faktor-faktor Kemiskinan di Provinsi Jawa Timur. *Statistika*, *12*, 1–8.

Fadliana, A., Pramoedyo, H., & Fitriani, R. (2020). Implementation Of Locally Compensated Ridge-Geographically Weighted Regression Model In Spatial Data With Multicollinearity Problems (Case Study: Stunting among Children Aged under Five Years in East Nusa Tenggara Province). *Media Statistika*, *13*(2), 125–135. https://doi.org/10.14710/medstat.13.2.125-135

Griffith, D. A. (2012). *Advanced Spatial Statistics: Special Topics in The Exploration of Quantitative Spatial Data Series* (Vol. 12). Springer Science & Business Media.

Kim, H.C., Pang, S., Je, H.M., Kim, D., and Bang, S. . (2003). Constructing Support Vector Machine Ensemble. *Pattern Recognition*, *36*, 2757–2767.

Lembang, F. K., Patty, H. W. M., & Maitimu, F. (2019). Analisis Kemiskinan Di Kabupaten Maluku Tenggara Barat Menggunakan Pendekatan Multivariate Adaptive Regression Spline (MARS). *Media Statistika*, *12*(2), 188. https://doi.org/10.14710/medstat.12.2.188-199

LeSage. (1999). *The Theory and Practice of Spatial Econometrics*. University of Toledo.

Lu, X., Zhou, W., Ding, X, Shi, X., Luan, B. and Li, M. (2019). Ensemble Learning Regression for Estimating Unconfined Compressive Strength of Cemented Paste Backfill. *IEEE*, *7*, 72125–72133.

Mevik, B.H., Segtnan, V.H. dan Naes, T. (2005). Ensemble Methods and Partial Least Squares Regression. *Journal of Chemometrics*, *18*, 498–507.

Rohmawati, N., Wijayanto, H and Wigena, A. H. (2015). Ensemble Spatial Autoregressive Model on the Poverty Data in Java. *Applied Statistical Sciences*, *9*, 2103–2110.

Saifudin, A. dan Wahono, R. S. (2016). Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, *1*(1).

Saputro, D. R. S., Sukmayanti, A., & Widyaningsih, P. (2019). The Nonparametric Regression Model Using Fourier Series Approximation and Penalized Least Squares (PLS) (Case on Data Proverty in East Java). *Journal of Physics: Conference Series*, *1188*(1). https://doi.org/10.1088/1742-6596/1188/1/012019

Sinta, D., Wijayanto, H. and Sartono, B. (2014). Ensemble K-Nearest Neighbor Method to Predict Rice Price in Indonesia. *Applied Statistical Sciences*, *8*, 7993–8005.

Vrrag M, N. T. (2014). The Application of Ensemble Methods in Forecasting Bankruptcy. *Financial and Economic Review*, *13*(4), 178–193.

Wu Z, H. N. (2005). Ensemble Empirical Mode Decomposition: A Noise Assisted Data Analysis Method, Kernels and Ensemble. *Perspective on Statistical Learning*.

Zhu, M. (2012). Kernel and Ensemble: Perspective on Statistical Learning. In *The American Statistician*. (Vol. 62, Issue 2, pp. 97–109).