Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 2 (2024) 298 - 308 e-ISSN: 2580-0760

LR-GLASSO Method for Solving Multiple Explanatory Variables of the Village Development Index

M Yunus[1], Agus M Soleh[2], Asep Saefuddin[3], Erfiani[4]
[1,2,3,4]Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[1]muhammadyunus@apps.ipb.ac.id, [2]agusms@apps.ipb.ac.id, [3]asaefuddin@apps.ipb.ac.id, [4]erfiani@apps.ipb.ac.id

Abstract

Sustainable Development Goals (SDGs) are developments that maintain sustainable improvement in society's economic, social, and environmental welfare. Kemendes PDTT RI has issued the Village Development Index (VDI) to provide information and the status of village progress to support village development to improve the National SDGS. Modelling with multiple explanatory variables causes a high correlation between explanatory variables, multicollinearity, and coefficient estimation results, which have major variance and overfitting in prediction results. The modeling solution uses LASSO and GLASSO. Binary categorical response data uses binary logistic regression (LR), so LR-LASSO and LR-GLASSO are used. North Maluku Province has a VDI ranking that tends to fall in 2018-2022. Based on the mean and variance of the coefficient estimation results and misclassification errors, LR-GLASSO is better than LR-LASSO and LR. LR-GLASSO is recommended for analyzing VDI data because it has many explanatory variables, and the correlation between them is relatively high. The recommendation given by the Indonesian government, if it is to increase the status of VDI in Indonesia, especially in North Maluku province, is to increase the number of electricity users, food and beverage stalls, and other cooperatives. The Indonesian government also needs to pay attention to villages relatively far from the regent's office, between food and beverage stalls, and supporting health centres, because they still need to be developed compared to other villages, and more than 50% of villages are underdeveloped. If the Village SDGs are formulated through increasing VDI status, it will support the achievement of SDGs goals nationally.

Keywords: *logistic regression (LR); LR-GLASSO, LR-LASSO; multiple explanatory variables; village development index.*

How to Cite: M. Yunus, A. M. Soleh, A. Saefuddin, and Erfiani, "LR-GLASSO Method for Solving Multiple Explanatory Variables of the Village Development Index", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 8, no. 2, pp. 298 - 308, Apr. 2024.
DOI: https://doi.org/10.29207/resti.v8i2.5656

1. Introduction

Sustainable Development Goals (SDGs) are developments that maintain sustainable improvement in society's economic, social, and environmental welfare [1]. The Village SDGs formula can support the achievement of SDG goals nationally [2]. The village SDGs in Indonesia determine a more specific and focused village development path to help achieve the SDGs at the national and international levels [3]. Village SDGs can create villages without poverty and hunger, equitable economic growth, health care, the environment, education, and women-friendly, networked, and culturally responsive villages [4]. Village development is one of the critical targets of government policy in Indonesia to optimize its potential to the maximum [5]. Based on Law Number 6 of 2014 concerning Villages, regional governments from the district/city level to the village level need to understand the potential of each village to increase the possibility of villages in their area.

The Ministry of Villages, Development of Disadvantaged Regions, and Transmigration of the Republic of Indonesia (Kemendes PDTT RI) has issued the Village Development Index (VDI) to provide information and status of village progress to support village development. VDI is a Composite Index formed based on three indices: the Social Resilience Index, the Economic Resilience Index, and the Ecological/Environmental Resilience Index [6]. However, the three village development measurement tools have almost the same shortcomings, namely that there are indicators that village officials and village policies cannot follow up [7]. The strategy to increase the status of the Village Development Index is to reallocate village financial management based on the status of the village typology and its supporting

Received: 07-01-2024 | Accepted: 26-04-2024 | Published Online: 28-04-2024

composite index [8]. Technology, information and communication infrastructure influence the VDI, which is related to the geographical conditions of areas far from the city centre [9]. VDI has a value at the interval (0.1), which includes social, economic, and environmental aspects [10].

Linear modelling aims to see the influence of explanatory variables on response variables and predict response based on one or more explanatory variables. This model is often used for multiple explanatory variables to describe the conditions of the response variable. Modelling with multiple explanatory variables causes a high correlation between explanatory variables, multicollinearity, and coefficient estimation results, which have major variance and overfit of prediction results. Multicollinearity problems can be overcome by adding new data, applying principal component analysis, and using LASSO and Elastic-Net regression [11], [12]. One commonly applied solution is variable selection and shrinkage to estimate a model known as the Least Absolute Shrinkage and Selection Operator [13], [14]. Least Absolute Shrinkage and Selection Operator (LASSO) adds L_1 regularization to the loss function used in linear regression [13]. The advantage of the LASSO method is that it can produce coefficient estimates that have a minor variance and a simpler model, so it is easy to interpret [15], superior to the Stein-type shrinkage-based estimators [16] and produces the Mean Square Error (MSE) most minor [17]. The weakness of the LASSO method is that it still allows several essential variables to shrink towards zero or equal to zero which has a high correlation [18], [19]. The development of variable selection in estimating models is group-based selection, where variables are included in a group and then selected and shrunk to a group of significant variables in the model [20]. In the case of gene or categorical variables that already have a natural group structure and want to see group sparsity, a method is developed by the Group Least Absolute Shrinkage and Selection Operator [21].

Group Least Absolute Shrinkage and Selection Operator (GLASSO) is a selection and shrinkage method carried out at the group level [22], [23], [24]. GLASSO is better than the Ordinary Least Squares Method (OLS) and LASSO when there are groups β and in one group, $\beta_i > 1$ all or $\beta_i < 1$ all or $\beta_i = 0$ all [25]. GLASSO is better than backward elimination in interpretation and prediction accuracy [26]. This method provides sparsity in a set of groups. If a group is included in the model, then the coefficient in that group will be non-zero, and vice versa [27], [28], [29]. Modelling using the LASSO and GLASSO methods can be applied to numerical and categorical response data. Binary categorical response data uses binary logistic regression [30], [31], [32]. The research results show no difference between the logistic regression model and the geographically weighted logistic regression (GWLR) model based on the AIC value deviance ratio and degrees of freedom [33]. LASSO and GLASSO methods will be interesting to assess the level of village independence known as VDI. This method is an integrated assessment innovation solution to support national SDGs through the Village SDGs. This can be used for accurate policymaking at Kemendes PDTT RI.

Table 1 Village development index for North Maluku province

	Years				
	2018	2019	2020	2021	2022
Index	0.6065	0.5814	0.5811	0.5861	0.5924
Rank	25	27	30	31	31

Based on Table 1, it shows that North Maluku Province has a VDI ranking that tends to fall in 2018-2022 [34], [35], [36], [37], [38]. Thus, this research will use VDI data from North Maluku Provinces as empirical data. The government prioritizes Indonesia's development from the periphery by strengthening regions and villages within the framework of a unitary state. Statistics Indonesia collects data on Village Potential (Podes). The results of Podes data collection can be used as material for regional analysis regarding economic social potential and regional facilities/infrastructure. Apart from that, it can also be used in program evaluation and regional-based policy/strategy preparation. Podes data collection is an activity to collect various information, both regarding the potential of the villages and related to the vulnerabilities or challenges faced [39]. Thus, Podes data will be used as explanatory variables. Furthermore, this research analyses VDI Data in North Maluku Province using the Logistic Regression (LR) with LASSO or GLASSO. This method is in the future called LR-LASSO and LR-GLASSO.

2. Research Methods

2.1 Methods

Linear regression models are linear in parameters [40]. The OLS is used to estimate the linear regression coefficient (β) value [41]. The goal of OLS in Formula 1 is to minimize the sum of the squares errors [42].

$$\min_\beta \frac{1}{2}\|y - X\beta\|_2^2 \tag{1}$$

LASSO was introduced by [13], changing the penalty in ridge regression with L_1 regularization and became known after [43] discovered the Least Angle Regression (LARS) algorithm method. LASSO is a method that adds an L_1 regularization in Formula 2.

$$\min_\beta \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1 \tag{2}$$

GLASSO is a regression method that applies LASSO selection to explanatory variables divided into several groups. Regularization method that adds L_2 norm penalty in Formula 3 [21].

$$\min_\beta \frac{1}{2}\left\|y - \sum_{l=1}^{m} X^{(l)}\beta^{(l)}\right\|_2^2 + \lambda\sum_{l=1}^{m}\sqrt{p_l}\left\|\beta^{(l)}\right\|_2 \tag{3}$$

LASSO and GLASSO have been developed into other models, namely logistic regression [44], [45]. If the likelihood function $L(\beta)$, for the desired model, is logconcave, then for LASSO and GLASSO, minimize Formula 4 and 5.

$$\ell(\beta) + \lambda\|\beta\|_1 \quad (4)$$

$$\ell(\beta) + \lambda \sum_{l=1}^{m} \sqrt{p_l}\left\|\beta^{(l)}\right\|_2 \quad (5)$$

with $\ell(\beta) = (-1/n)\log\left(L(\beta)\right)$. A commonly used case is logistic regression. Logistic regression is a response variable (y) with n-binary response vectors, and X, n by p covariate matrix divided into m groups, $X^{(1)}, X^{(2)}, \dots, X^{(m)}$. In this case, LASSO and GLASSO require shapes in Formula 6 and 7.

$$\hat{\beta} = \operatorname{argmin}_{\beta} \frac{1}{n}\left[\left(\sum_{i=1}^{n} \log(1+\exp(x_i^T\beta)) + y_i x_i^T\beta\right)\right] + \lambda\|\beta\|_1 \quad (6)$$

$$\hat{\beta} = \operatorname{argmin}_{\beta} \frac{1}{n}\left[\left(\sum_{i=1}^{n} \log(1+\exp(x_i^T\beta)) + y_i x_i^T\beta\right)\right] + \lambda \sum_{l=1}^{m} \sqrt{p_l}\left\|\beta^{(l)}\right\|_2 \quad (7)$$

2.2 Research Data

The simulation data used in this research are 24 explanatory variables and one response variable. Explanatory variables $(X_1, X_2, X_3, \dots, X_{24})$ generated with an average of 1 each $(\mu_i = 1)$ and with correlations namely 0.1 and 0.9 $(\rho = \{0.1, 0.9\})$. The amount of data generated is 100 $(n = 100)$. Model parameter scenarios (β_i) specified, namely $\beta_i \leq -1, \beta_i = 0$ and $\beta_i \geq 1$, are presented in Table 2. The response variable (y) generated $f(X_1, X_2, X_3, \dots, X_{24})$. The link function used is logit in Formula 8.

$$\ln\left(\frac{\mu_i}{1-\mu_i}\right) = X_i\beta \text{ and } \mu_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \quad (8)$$

Table 2 Scenario coefficients (β_i) and groups on simulated data

Groups	β_i	Scenario
1	β_0	1
	β_1	2
	β_2	2
	β_3	2
	β_4	2
	β_5	2
	β_6	2
2	β_7	-2
	β_8	-2
	β_9	-2
	β_{10}	-2
	β_{11}	-2
	β_{12}	-2
3	β_{13}	1
	β_{14}	1
	β_{15}	1
	β_{16}	1
4	β_{17}	-1
	β_{18}	-1
	β_{19}	-1
	β_{20}	-1
5	β_{21}	0
	β_{22}	0
	β_{23}	0
	β_{24}	0

The Village Development Index (VDI) is a Composite Index formed based on three indices: the Social Resilience Index, the Economic Resilience Index, and the Ecological/Environmental Resilience Index [6]. In this context, social, economic, and ecological resilience work as dimensions that strengthen the process and achieve the goals of development and empowerment of village communities. VDI captures the development of village independence based on implementing the Village Law with the support of Village Funds and Village Assistants [3]. VDI directs the accuracy of policy interventions with the correlation of appropriate development interventions from the Government by Community participation, which correlates with the characteristics of the village area, namely typology and social capital [46].

Podes data collection was carried out three times over ten years as part of the ten-year cycle of census activities carried out by BPS. Podes are carried out by census in all districts/cities, sub-districts and the lowest government administrative areas at the village level. Podes data collection is carried out through direct interviews by trained officers with relevant sources in the census area, as well as searching for related documents [47]. Podes data processing is carried out at Regency/City BPS to speed up completion time and consider the ease of data validation because the data processing centre is close to the data source. Dissemination of the 2021 Podes data collection results is organized into several main types of publications, namely Indonesian Village Potential Statistics, Provincial Village Potential Statistics, and Indonesian Statistics. Information related to village/subdistrict potential includes employment, education, health, social culture, sports and entertainment, transportation, communication and information, economy, security, development and empowerment of village/subdistrict communities. Information related to vulnerabilities or challenges includes natural disasters, environmental pollution, social and health problems in the community, and security disturbances that occur in villages [39].

Empirical data are Village Potential (Podes) and Village Development Index (VDI) data in North Maluku Province. Podes data is secondary data collected by Statistics Indonesia (BPS) of North Maluku Province using a questionnaire [47]. VDI data is secondary data collected by the Directorate General of Village Community Development and Empowerment, Kemendes PDTT RI [36]. Podes and VDI data in North Maluku Province comprise 1,199 villages from 117 sub-districts in 10 regencies/cities. The VDI becomes the response variable, and 94 variables become explanatory variables, as presented in Table 3. The Village Classification data used is North Maluku Province in 2020. The data consists of 1,199 observations (Villages), but 141 Villages experiencing missing values were related to the response variable, so they were not included in the analysis. The remaining villages in this study are 1,058, and missing values will be predicted. Figure 1 shows that most of the Classification of Villages in North Maluku Province in 2020 are categorized as underdeveloped villages, namely 61% (647 Villages).

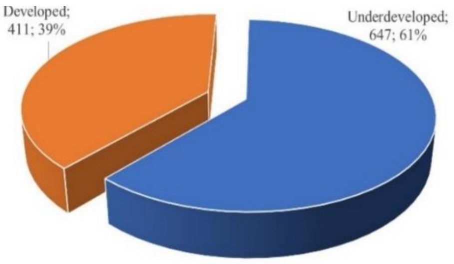


Figure 1. Distribution of village classification in 1,058 villages

Table 3 List of response variables and explanatory variables

	Response variable	Scale
1.	Village development index (VDI)	Nominal (Developed and Underdeveloped)
	Explanatory variables	Scale
1.	Number of families using electricity	Ratio
2.	Number of embungs in the village	Ratio
3.	Number of slum locations	Ratio
4.	Number of slum buildings	Ratio
5.	Number of slum families	Ratio
6.	Number of early childhood education posts	Ratio
⋮	⋮	⋮
92.	Number of village-owned enterprise business units	Ratio
93.	Number of village markets	Ratio
94.	Number of boat moorings	Ratio

2.3. Data Analysis Procedures

Logistic regression analysis of simulated using R software [48]. The simulation data analysis procedure to study the characteristics of LR-LASSO and LR-GLASSO is presented in Figure 2a. The steps are as follows: Explanatory variables $(X_1, X_2, X_3, \dots, X_{24})$ were generated with $(n = 100)$, $(\rho = \{0.1, 0.9\})$ and $\beta_i \leq -1, \beta_i = 0$ *and* $\beta_i \geq 1$; Generating response variable (y) with Formula 8; The model estimation uses the LR and LR-LASSO methods with optimal control parameters (λ) with Formula 6; Create groups of explanatory variables; The model estimation uses the LR-GLASSO method with optimal control parameters (λ) with Formula 7; Procedures a to e were repeated 300 times; The estimation results are presented in a boxplot on one graph; Calculate Mean $(\bar{x})$ and variance (s^2) of the coefficients (β_i) from 300 times; Misclassification Error results are presented in a histogram on one graph and table; Compare the best method with the criteria mean $(\bar{x})$ and variance (s^2) of the coefficients (β_i) and misclassification error; The best methods for applying empirical data are Village Potential (Podes) and Village Development Index (VDI) data.

The empirical data analysis procedure using LR-GLASSO on the VDI data is presented in Figure 2b. The steps are as follows: Data exploration as initial information and group formation; Looking at the correlation between explanatory variables and between explanatory variables and the response; Discards observation data missing value on each response variable; Create a group of explanatory variables based on the results of point 2; Carrying out an analysis using LR-GLASSO with optimal control parameters (λ) with Formula 7 via cross-validation (CV) process; Get the simplest model with the lowest misclassification rate; Analyzing 94 explanatory variables that have the most significant positive and negative influence on VDI; Provide recommendations to the Directorate General of Village Community Development and Empowerment, Ministry of Village PDTT RI and Statistics Indonesia (BPS) regarding variables that must be improved because they positively influence VDI; Provide recommendations to the Directorate General of Village Community Development and Empowerment, Ministry of Village PDTT RI and Statistics Indonesia (BPS) regarding variables that must be reduced because they negatively influence VDI; Make predictions on response data that has missing values; Get all village classifications in the North Maluku Province, Indonesia, 2020.

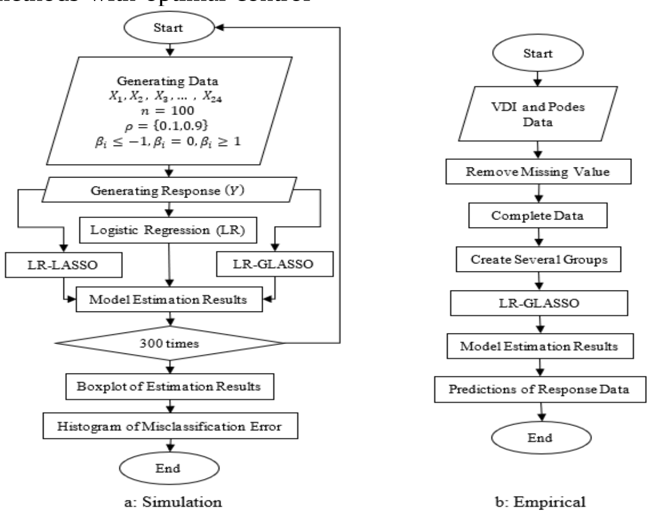


Figure 2 Data analysis procedures

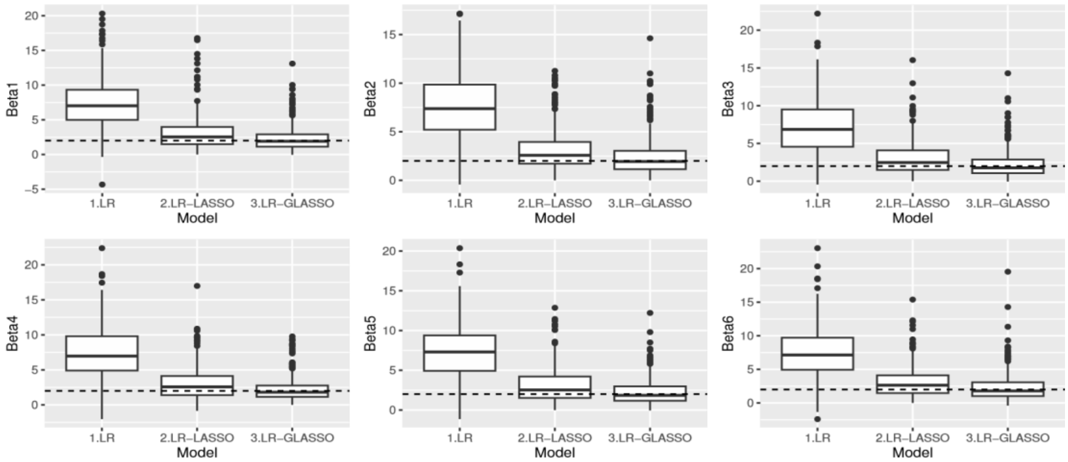


Figure 3 $\beta_1, \ldots, \beta_6 = (2,2,2,2,2,2)$ of 300 times simulation with $\rho = 0.9$

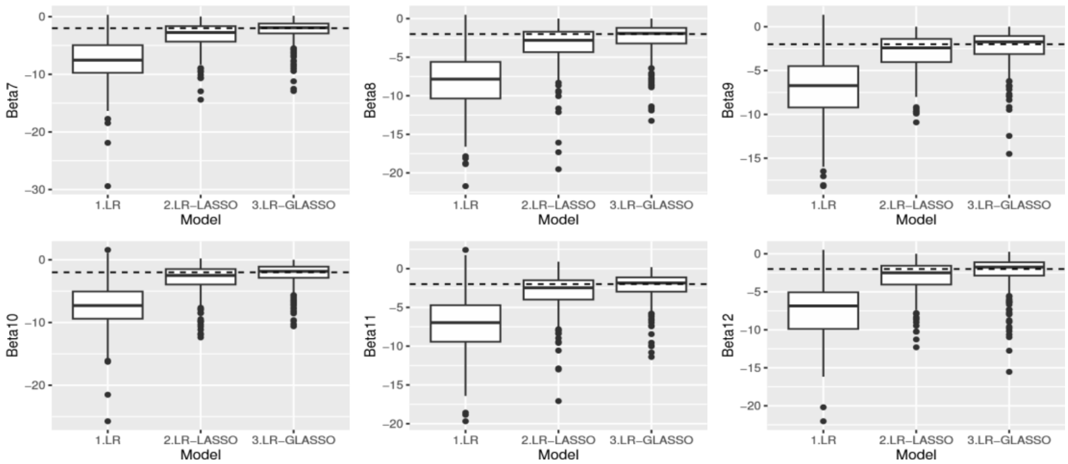


Figure 4 $\beta_7, \ldots, \beta_{12} = (-2,-2,-2,-2,-2,-2)$ of 300 times simulation with $\rho = 0.9$

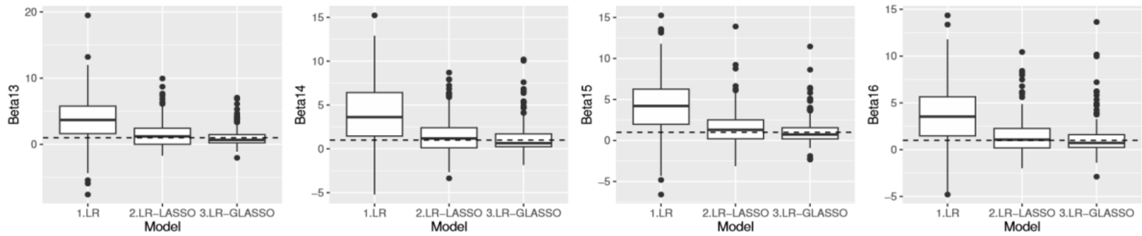


Figure 5 $\beta_{13}, \ldots, \beta_{16} = (1,1,1,1)$ of 300 times simulation with $\rho = 0.9$

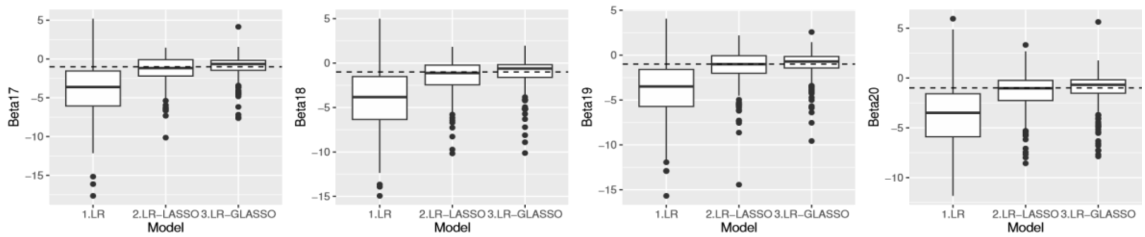


Figure 6 $\beta_{17}, \ldots, \beta_{20} = (-1,-1,-1,-1)$ of 300 times simulation with $\rho = 0.9$

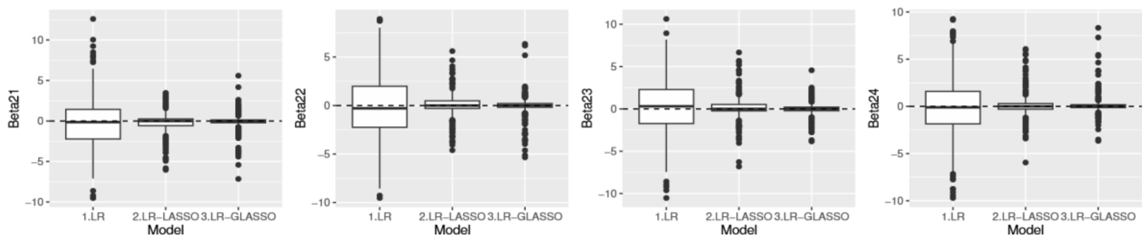


Figure 7 $\beta_{21}, \ldots, \beta_{24} = (0,0,0,0)$ of 300 times simulation with $\rho = 0.9$

Table 4 Mean ($\bar{x}$) and variance (s^2) of the coefficients (β_i) from 300 times simulation with $\rho = (0.9,0.1)$

			$\rho = 0.9$						$\rho = 0.1$		
β_i			LR	LR-LASSO	LR-GLASSO	β_i			LR	LR-LASSO	LR-GLASSO
β_1	2	$\bar{x}$	7.4572	3.1479	2.3469	β_1	2	$\bar{x}$	4.6504	1.9072	1.3892
		s^2	13.9875	6.7551	3.3280			s^2	3.0539	1.8684	1.4082
β_2	2	$\bar{x}$	7.6265	3.0829	2.4467	β_2	2	$\bar{x}$	4.9099	1.9911	1.4272
		s^2	11.4477	4.7734	4.1457			s^2	3.1274	1.7849	1.3962
β_3	2	$\bar{x}$	7.2075	3.0056	2.2984	β_3	2	$\bar{x}$	4.6421	1.8959	1.3529
		s^2	12.3380	5.3323	3.6304			s^2	2.9010	1.7455	1.1659
β_4	2	$\bar{x}$	7.4013	3.0882	2.3935	β_4	2	$\bar{x}$	4.7229	1.9227	1.3480
		s^2	13.3258	5.7236	3.6864			s^2	3.0949	1.8028	1.1590
β_5	2	$\bar{x}$	7.3532	3.0852	2.3276	β_5	2	$\bar{x}$	4.6723	1.9157	1.3504
		s^2	11.4896	5.1590	3.1395			s^2	3.0561	1.7816	1.0725
β_6	2	$\bar{x}$	7.4580	3.0929	2.4634	β_6	2	$\bar{x}$	4.5261	1.8595	1.3021
		s^2	13.8209	5.4943	4.9657			s^2	3.2445	1.7645	1.1075
β_7	-2	$\bar{x}$	-7.8018	-3.1948	-2.3984	β_7	-2	$\bar{x}$	-4.8635	-2.0198	-1.4107
		s^2	14.7086	5.4396	3.7244			s^2	3.1976	2.0149	1.3036
β_8	-2	$\bar{x}$	-8.0597	-3.3895	-2.5553	β_8	-2	$\bar{x}$	-4.6464	-1.8783	-1.3718
		s^2	13.1830	6.7383	4.3820			s^2	2.9029	1.5987	1.2687
β_9	-2	$\bar{x}$	-7.0418	-2.9458	-2.3638	β_9	-2	$\bar{x}$	-4.8944	-1.9563	-1.4319
		s^2	12.8622	4.6388	3.7766			s^2	3.1377	1.7094	1.3599
β_{10}	-2	$\bar{x}$	-7.3146	-3.0459	-2.3208	β_{10}	-2	$\bar{x}$	-4.6383	-1.8727	-1.3333
		s^2	13.0328	5.4467	3.3644			s^2	2.7380	1.6051	1.2088
β_{11}	-2	$\bar{x}$	-7.1099	-3.0129	-2.3140	β_{11}	-2	$\bar{x}$	-4.6774	-1.9568	-1.3728
		s^2	13.4351	5.7571	3.4973			s^2	2.5804	1.9927	1.1864
β_{12}	-2	$\bar{x}$	-7.4157	-3.0075	-2.3994	β_{12}	-2	$\bar{x}$	-4.7282	-1.9059	-1.3665
		s^2	13.2429	4.5933	4.6961			s^2	2.8322	1.6199	1.1891
β_{13}	1	$\bar{x}$	3.7794	1.4860	1.0209	β_{13}	1	$\bar{x}$	2.4177	0.8769	0.6165
		s^2	11.1370	2.6854	1.5441			s^2	1.9154	0.7075	0.4711
β_{14}	1	$\bar{x}$	3.9475	1.5140	1.1482	β_{14}	1	$\bar{x}$	2.3310	0.9112	0.6396
		s^2	12.4271	3.0763	2.3409			s^2	2.1481	0.8010	0.4693
β_{15}	1	$\bar{x}$	4.1644	1.5879	1.0883	β_{15}	1	$\bar{x}$	2.4369	0.8887	0.6305
		s^2	12.2250	3.3605	2.0882			s^2	2.3847	0.7308	0.5040
β_{16}	1	$\bar{x}$	3.7826	1.4922	1.1549	β_{16}	1	$\bar{x}$	2.2994	0.8507	0.6049
		s^2	10.4149	2.8450	2.5901			s^2	1.9616	0.6878	0.4783
β_{17}	-1	$\bar{x}$	-3.8023	-1.4621	-1.0211	β_{17}	-1	$\bar{x}$	-2.4168	-0.9004	-0.5971
		s^2	11.7424	2.6586	1.7938			s^2	2.8246	0.8800	0.4724
β_{18}	-1	$\bar{x}$	-4.0052	-1.5680	-1.1508	β_{18}	-1	$\bar{x}$	-2.4479	-0.9168	-0.6258
		s^2	12.2525	3.1805	2.4338			s^2	2.2918	0.7867	0.4583
β_{19}	-1	$\bar{x}$	-3.7593	-1.4282	-1.0404	β_{19}	-1	$\bar{x}$	-2.2655	-0.8503	-0.5694
		s^2	11.0997	3.2132	1.8736			s^2	2.1471	0.7661	0.4075
β_{20}	-1	$\bar{x}$	-3.6804	-1.4840	-1.0823	β_{20}	-1	$\bar{x}$	-2.3797	-0.9187	-0.6314
		s^2	10.6154	2.9672	2.3000			s^2	2.5799	0.9751	0.4262
β_{21}	0	$\bar{x}$	-0.1950	-0.1801	-0.0921	β_{21}	0	$\bar{x}$	-0.0294	-0.0174	-0.0193
		s^2	10.6706	1.7788	1.1071			s^2	2.1624	0.4547	0.1860
β_{22}	0	$\bar{x}$	-0.1089	-0.0152	-0.0644	β_{22}	0	$\bar{x}$	-0.0323	-0.0458	-0.0188
		s^2	11.0772	1.8188	1.3194			s^2	2.0957	0.3891	0.1488
β_{23}	0	$\bar{x}$	0.1753	0.0909	-0.0061	β_{23}	0	$\bar{x}$	0.0565	0.0153	0.0383
		s^2	10.8917	2.0769	0.7567			s^2	2.1767	0.3825	0.1747
β_{24}	0	$\bar{x}$	-0.0988	0.0653	0.1426	β_{24}	0	$\bar{x}$	0.0535	0.0221	0.0218
		s^2	10.9369	1.9513	1.3416			s^2	2.0656	0.4078	0.1816

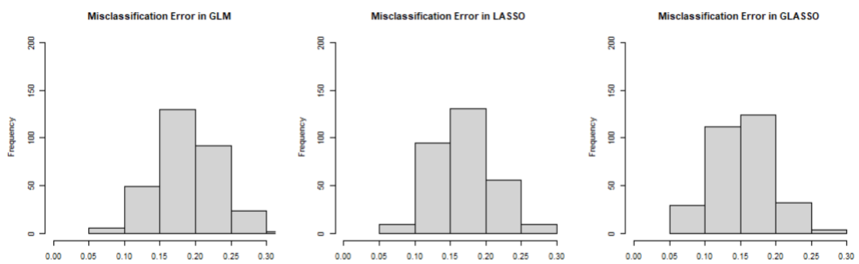


Figure 8 Misclassification error of 300 times simulation with $\rho = 0.9$

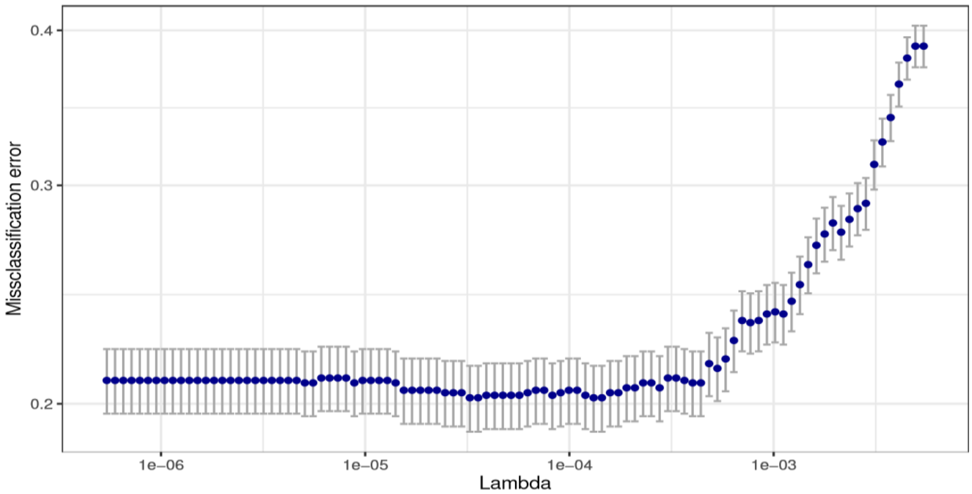
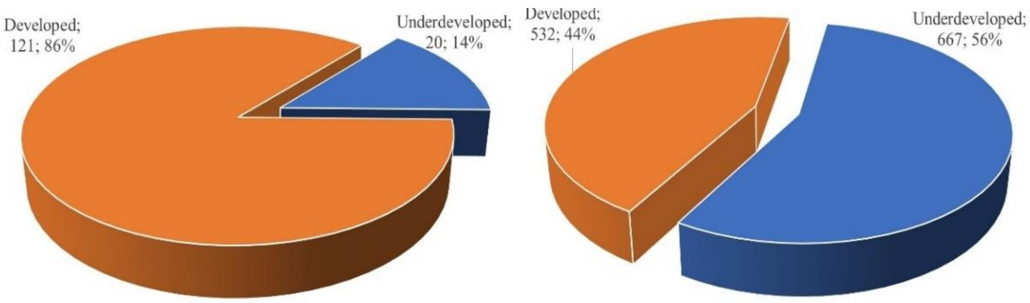


Figure 9 Misclassification error of the cross-validation (CV) process LR-GLASSO



a. Prediction results for 141 villages

b. Village classification in 1,199 villages

Figure 10 Village classification in North Maluku Province

3. Results and Discussions

3.1 Comparative simulation study of analytical methods

The results of simulation data analysis to study the characteristics of LR, LR-LASSO and LR-GLASSO, are presented in Figure (3 - 8), Appendix Figure (11 - 16) and Table (4 - 5). Based on Figures 3-7 and 11-17, it is better to use LR-GLASSO compared to LR and LR-LASSO because it has more precise coefficient estimation results and minor variance. The more precise LR-GLASSO coefficient estimation results show that all scenario coefficients are close to the actual values. The slight variance in the estimated results shows that the LR-GLASSO boxplot is smaller than LR and LR-LASSO. Based on the Boxplot results, they are sorted from the best to LR-GLASSO, LR-LASSO and LR.

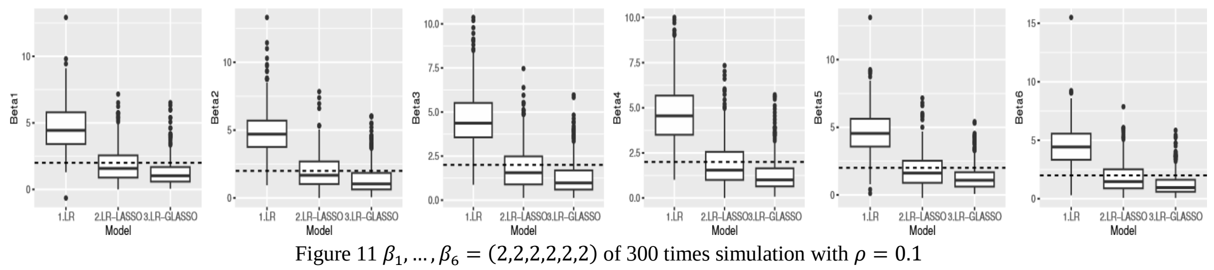


Figure 11 $\beta_1, \ldots, \beta_6 = (2,2,2,2,2,2)$ of 300 times simulation with $\rho = 0.1$

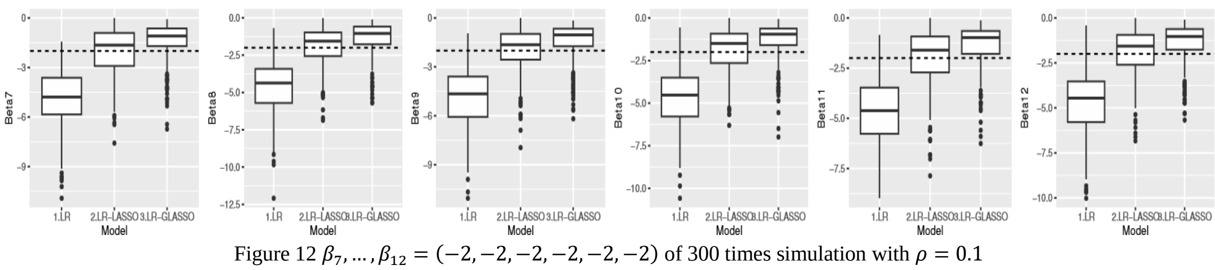


Figure 12 $\beta_7, \ldots, \beta_{12} = (-2,-2,-2,-2,-2,-2)$ of 300 times simulation with $\rho = 0.1$

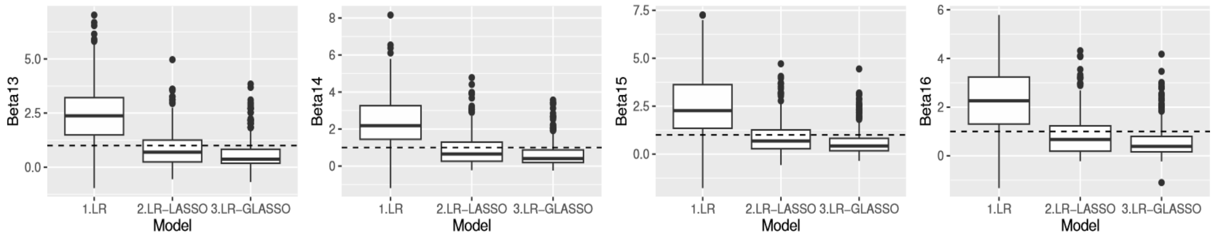


Figure 13 $\beta_{13}, \dots, \beta_{16} = (1,1,1,1)$ of 300 times simulation with $\rho = 0.1$

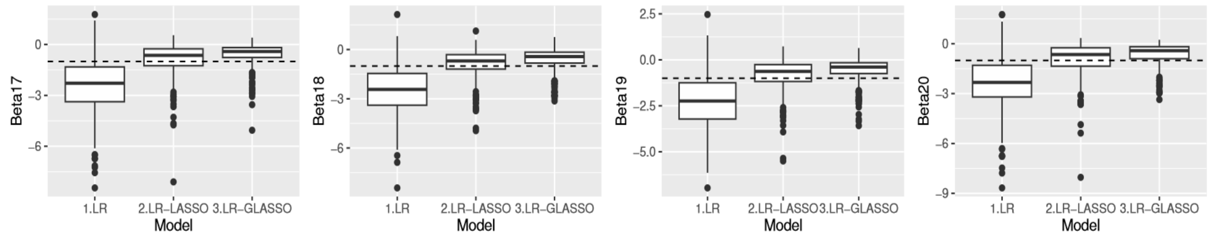


Figure 14 $\beta_{17}, \dots, \beta_{20} = (-1,-1,-1,-1)$ of 300 times simulation with $\rho = 0.1$

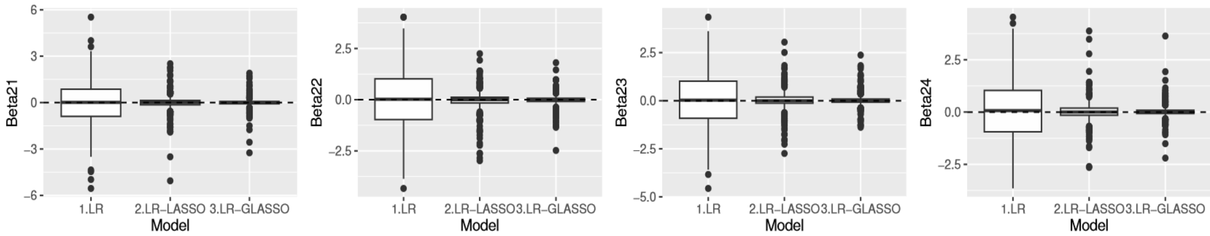


Figure 15 $\beta_{21}, \dots, \beta_{24} = (0,0,0,0)$ of 300 times simulation with $\rho = 0.1$

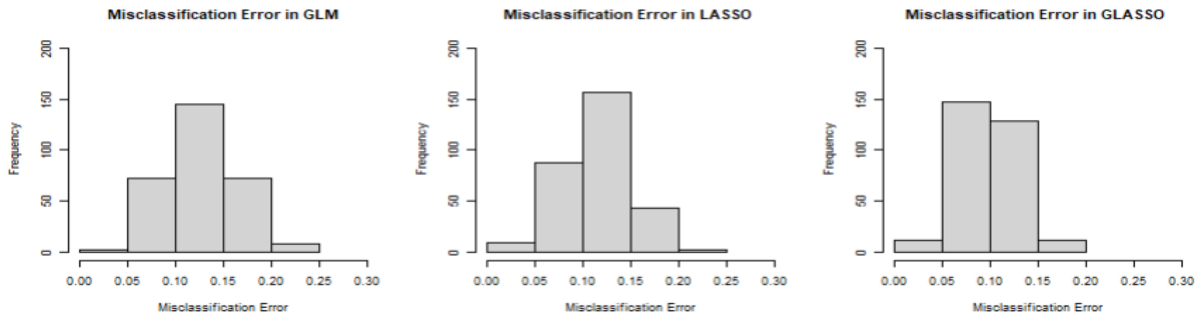


Figure 16 Misclassification error of 300 times simulation with $\rho = 0.1$

This is more detailed in Table 4, namely that based on the mean $(\bar{x})$ and variance (s^2) of the coefficient estimation results (β_i), LR-GLASSO is better than LR and LR-LASSO. In the $\rho = 0.9$ scenario, only the means of β_{22} and β_{24} are slightly better than LR-LASSO, but relative to all means and variances the coefficient estimation results show that LR-GLASSO is better and is closest to the actual value. In the $\rho = 0.1\ scenario$, if it is based on the mean of the coefficient estimation results, then LR-LASSO is better. However, if it is based on the variance of the coefficient estimation results, then LR-GLASSO is better. Based on the general correlation, the more significant the correlation between explanatory variables, the better the results and variance of LR-GLASSO coefficient estimates compared to LR and LR-LASSO.

Based on the misclassification error results in Figures (8 and 16) and Table 5, LR-GLASSO is better than LR and LR-LASSO. In the rho=0.9 scenario, the misclassification error value of LR-GLASSO (15.69%) is smaller than LR-LASSO (17.24%) and LR (19.44%). In the rho=0.1 scenario, the misclassification error value of LR-GLASSO (10.31%) is smaller than LR-LASSO (12.08%) and LR (13.15%). Thus, based on the mean and variance of the coefficient estimation results and misclassification errors, LR-GLASSO is better than LR-LASSO and LR.

Tabel 5 Misclassification error of 300 times

Correlation	Methods (%)		
	LR	LR-LASSO	LR-GLASSO
0.9	19.44	17.24	15.69
0.1	13.15	12.08	10.31

3.2 Empirical data analysis using LR-GLASSO

LR-GLASSO is recommended for analyzing VDI data because it has many explanatory variables, and the correlation between them is relatively high. LR-GLASSO is believed to produce better coefficient estimates and predictions than other methods. Based on the simulation study results, the best method used to analyze VDI data is LR-GLASSO analysis. LR-GLASSO analysis carrying out a cross-validation process of 100 lambdas is presented in Figure 9. The greater the lambda (penalty) given tends to produce higher misclassification. However, in more detail, it is presented in Table 6, which shows that lambda.min provides the lowest level of misclassification error, but the models and groups of variables produced are still too many, namely 94 explanatory variables. Lambda.1se is the largest value of lambda such that the error is within one standard error of the minimum. Lambda.1se provides a much simpler model and has relatively fewer variable groups, namely 62 explanatory variables. Therefore, the VDI modelling for North Maluku province in 2020 uses the lambda.1se value.

Table 6 Summary of LR-GLASSO analysis

	Lambda	CV measure	Non-zero variables	Active groups
Max.	0.0054	0.3885	0	0
Lambda.1se	0.0005	0.2136	62	9
Lambda.min	0.0001	0.2023	94	12
Min.	0.0000	0.2089	94	12

Analysis results of VDI response variables and 62 explanatory variables using LR-GLASSO. Seven variables with the greatest positive and negative influence compared to other variables are presented in Table 7.

Table 7 Variables that have the greatest positive/negative influence

	Variables have the biggest positive influence	
	Variables	β_i
1.	Number of electricity users	0.3447
2.	Number of food and beverage stalls	0.2455
3.	Number of other cooperatives	0.2063
4.	Number of grocery stalls	0.1928
5.	Several food and beverage industries	0.1161
6.	Distance to village maternity centre (polindes)	0.0974
7.	Semi-permanent building market	0.0897
	Variables have the biggest negative influence	
	Variables	β_i
1.	Distance from the regent's office	-0.2011
2.	Distance between food and beverage stalls	-0.1809
3.	Distance to supporting health centre	-0.1448
4.	Accommodation distance	-0.1222
5.	Number of polindes	-0.1145
6.	Restaurant distance	-0.1054
7.	Distance from hospital	-0.1038

This is following research conducted [49] that the value of the Economic and Social Security Index tends to be high. Several other core variables, such as village government human resources, village economic activities, and village development, have unique performance, which is by research [50]. The recommendation given by the Indonesian government through the Ministry of Villages PDTT RI and BPS, if it is to increase the status of VDI in Indonesia, especially in North Maluku province, is to increase the number of electricity users, number of food and beverage stalls, number of other cooperatives, number of grocery stalls, number of food and beverage industries, distance to village maternity centre (polindes), and semi-permanent building market. The Indonesian government also needs to pay attention to villages relatively far from the regent's office, between food and beverage stalls, supporting health centres, accommodation, restaurants and hospitals because they still need to be developed compared to other villages.

Predictions for one hundred and forty-one villages experiencing missing values related to the village classification response variable were carried out using the LR-GLASSO method. Figure 10a shows village classification prediction results from distribution in 141 villages. Many villages that experienced missing values were predicted to be villages with developed status, namely 86% (121 villages). There are 20 villages (14%) that are predicted to have underdeveloped status. Figure 10b shows the distribution of village classifications in 1,199 villages in North Maluku Province in 2020 (including prediction results), namely developed (532 villages/44%) and underdeveloped (667 villages/56%). The Indonesian government still needs to pay attention to villages in North Maluku province because more than 50% of villages are underdeveloped. If the Village SDGs are formulated through increasing VDI status, it will support the achievement of SDGs goals nationally.

4. Conclusions

LR-GLASSO is better than LR and LR-LASSO because it has more precise coefficient estimation results and minor variance. Based on the mean and variance coefficient estimation, results show that LR-GLASSO is better and closer to the actual value. Based on the general correlation, the more significant the correlation between explanatory variables, the better the results and variance of LR-GLASSO coefficient estimates compared to LR and LR-LASSO. Based on the misclassification error results, LR-GLASSO is better than LR and LR-LASSO. Thus, based on the mean and variance of the coefficient estimation results and misclassification errors, LR-GLASSO is better than LR-LASSO and LR. LR-GLASSO is recommended for analyzing VDI data because it has many explanatory variables, and the correlation between them is relatively high. LR-GLASSO is believed to produce better coefficient estimates and predictions than other methods. Analysis results of VDI response variables and 62 explanatory variables using LR-GLASSO with lambda.1se. Seven variables with the most significant positive and negative influence compared to other variables. The recommendation given by the Indonesian government through the Ministry of Villages PDTT RI and BPS, if it is to increase the status of VDI in Indonesia, especially in North Maluku province, is to increase the number of electricity users, number of food and beverage stalls, number of other cooperatives, number of grocery stalls, number of food and beverage industries, distance to village maternity centre (polindes), and semi-permanent building market. The Indonesian government also needs to pay attention to villages relatively far from the regent's office, between food and beverage stalls, supporting health centres, accommodation, restaurants and hospitals because they still need to be developed compared to other villages. The distribution of village classifications in 1,199 villages in North Maluku Province in 2020 (including prediction results), namely developed (532 villages/44%) and underdeveloped (667 villages/56%). The Indonesian government still needs to pay attention to villages in North Maluku province because more than 50% of villages are underdeveloped. If the Village SDGs are formulated through increasing VDI status, it will support the achievement of SDGs goals nationally. Future research will develop modelling using the Logistic Regression Sparse Group LASSO (LR-SGLASSO) method applied to VDI data. This method is to obtain a simpler model compared to LR-GLASSO.

References

[1] Kementerian PPN/Bappenas, "Pedoman Teknis Penyusunan Rencana Aksi Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs)," Jakarta, 2020.

[2] E. Kurniawan, Amidi, Gunawan, N. Susilowati, L. Paranti, and D. G. Santi, *Panduan UNNES GIAT Penguatan Generasi Milenial Mendukung SDGs Desa*. 2022.

[3] T. W. Rahmaddhani and N. Prasetyoningsih, "Achieving a Developing Village based on the Village Sustainable Development Goals in Tirtonirmolo Village, Bantul Regency," *Jurnal Penegakan Hukum dan Keadilan*, vol. 4, no. 1, pp. 11–29, Mar. 2023, doi: 10.18196/jphk.v4i1.16043.

[4] Permendesa Nomor 13, "Peraturan Menteri Desa, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia," Jakarta, 2020.

[5] N. A. Utami and A. W. Wijayanto, "Classification of Village Development Index at Regency/ Municipality Level Using Bayesian Network Approach With K-Means Discretization," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. Khusus, pp. 95–106, 2022.

[6] Muhtarom, N. Kusuma, and E. Purwanti, "Analisis Indeks Desa Membangun untuk Mengetahui Pola Perkembangan Pembangunan Desa Di Kecamatan Gadingrejo Kabupaten Pringsewu," *INOVASI PEMBANGUNAN – JURNAL KELITBANGAN*, vol. 6, no. 2, pp. 179–190, 2018, [Online]. Available: http://journalbalitbangdalampung.org

[7] H. S. B. Harmadi, U. Suchaini, and A. Adji, "Indonesia's Village Development Indicator: In Terms of Mismatch of Village Development Measurement Indicator," Jakarta, 2020. [Online]. Available: www.tnp2k.go.id

[8] E. P. Yudha, B. Juanda, L. M. Kolopaking, and R. A. Kinseng, "Rural development policy and strategy in the rural autonomy era. Case study of pandeglang regency-Indonesia," *Human Geographies - Journal of Studies and Research in Human Geography*, vol. 14, no. 1, pp. 125–147, May 2020, doi: 10.5719/hgeo.2020.141.8.

[9] A. R. T. Hidayat, Y. E. Prasetya, and D. Dinanti, "Village Development Index and ICT Infrastructure in Tourism Region," *J. Ind. Tour. Dev. Std*, vol. 7, no. 3, pp. 166–174, 2019, doi: 10.21776/ub.jitode.2019.007.03.05.

[10] A. N. Astika and N. S. Subawa, "Evaluasi Pembangunan Desa Berdasarkan Indeks Desa Membangun," *Jurnal Ilmiah Muqoddimah : Jurnal Ilmu Sosial, Politik dan Humaniora*, vol. 5, no. 2, 2021, [Online]. Available: http://jurnal.um-tapsel.ac.id/index.php/muqoddimah

[11] F. Abqorunnisa, Erfiani, and A. Djuraidah, "Performance of LASSO And Elastic-Net Methods on Non-Invasive Blood Glucose Measurement Calibration Modeling," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 1, pp. 0037–0042, Apr. 2023, doi: 10.30598/barekengvol17iss1pp0037-0042.

[12] P. Rintara, S. E. Ahmed, and S. Lisawadi, "Post Improved Estimation and Prediction in the Gamma Regression Model," *Thailand Statistician*, vol. 21, no. 3, pp. 580–606, 2023, [Online]. Available: http://statassoc.or.th

[13] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[14] R. N. Rachmawati, A. C. Sari, and Yohanes, "Lasso Regression for Daily Rainfall Modeling at Citeko Station, Bogor, Indonesia," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 383–390. doi: 10.1016/j.procs.2021.01.020.

[15] J. M. K. Aheto, H. O. Duah, P. Agbadi, and E. K. Nakua, "A predictive model, and predictors of under-five child malaria prevalence in Ghana: How do LASSO, Ridge and Elastic net regression approaches compare?," *Prev Med Rep*, vol. 23, Sep. 2021, doi: 10.1016/j.pmedr.2021.101475.

[16] O. Reangsephet, S. Lisawadi, and S. E. Ahmed, "Post Selection Estimation and Prediction in Poisson Regression Model," *Thailand Statistician*, vol. 18, no. 2, pp. 176–195, 2020, [Online]. Available: http://statassoc.or.th

[17] N. H. Pusponegoro, A. Kurnia, K. A. Notodiputro, A. M. Soleh, and E. T. Astuti, "Small Area Estimation of Sub-District's per Capita Expenditure through Area Effects Selection using LASSO Method," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 754–761. doi: 10.1016/j.procs.2021.01.064.

[18] A. M. Soleh and Aunuddin, "Lasso : Alternatif Seleksi Peubah Dan Penyusutan Koefisien Model Regresi Linier, Solusi," *FSK : Indonesian Journal of Statistics*, vol. 18, no. 1, pp. 21–27, 2013.

[19] S. Sardy, J. Diaz-Rodriguez, and C. Giacobino, "Thresholding tests based on affine LASSO to achieve non-asymptotic nominal level and high power under sparse and dense alternatives in high dimension," *Comput Stat Data Anal*, vol. 173, pp. 1–14, Sep. 2022, doi: 10.1016/j.csda.2022.107507.

[20] Kusnaeni, A. M. Soleh, F. M. Afendi, and B. Sartono, "Function Group Selection of Sembung Leaves (Blumea Balsamifera) Significant to Antioxidants Using Overlapping Group Lasso," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 2, pp. 721–728, Jun. 2022, doi: 10.30598/barekengvol16iss2pp721-728.

[21] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, no. 1, pp. 49–67, 2006.

[22] S. Zheng and C. Ding, "A group lasso based sparse KNN classifier," *Pattern Recognit Lett*, vol. 131, pp. 227–233, Mar. 2020, doi: 10.1016/j.patrec.2019.12.020.

[23] H. F. Zhang, "Minimum Average Variance Estimation with group Lasso for the multivariate response Central Mean Subspace," *J Multivar Anal*, vol. 184, Jul. 2021, doi: 10.1016/j.jmva.2021.104753.

[24] W. Wang, H. Yuan, J. Han, and W. Liu, "PCLassoLog: A protein complex-based, group Lasso-logistic model for cancer classification and risk protein complex discovery," *Comput Struct Biotechnol J*, vol. 21, pp. 365–377, Jan. 2023, doi: 10.1016/j.csbj.2022.12.005.

[25] M. Yunus, A. Saefudin, and A. M. Soleh, "Characteristics of group LASSO in handling high correlated data," *Applied Mathematical Sciences*, vol. 11, pp. 953–961, 2017, doi: 10.12988/ams.2017.7276.

[26] H. Chen and Y. Xiang, "The Study of Credit Scoring Model Based on Group Lasso," in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 677–684. doi: 10.1016/j.procs.2017.11.423.

[27] C. Shang, H. Ji, X. Huang, F. Yang, and D. Huang, "Generalized grouped contributions for hierarchical fault diagnosis with group Lasso," *Control Eng Pract*, vol. 93, Dec. 2019, doi: 10.1016/j.conengprac.2019.104193.

[28] Q. Kang, Q. Fan, and J. M. Zurada, "Deterministic convergence analysis via smoothing group Lasso regularization and adaptive momentum for Sigma-Pi-Sigma

neural network," *Inf Sci (N Y)*, vol. 553, pp. 66–82, Apr. 2021, doi: 10.1016/j.ins.2020.12.014.

[29] Y. Sun and Q. Wang, "An adaptive group LASSO approach for domain selection in functional generalized linear models," *J Stat Plan Inference*, vol. 219, pp. 13–32, Jul. 2022, doi: 10.1016/j.jspi.2021.11.003.

[30] Alan. Agresti, *Analysis of Ordinal Categorical Data*, Second. Canada: John Wiley & Sons, Inc, 2010.

[31] P. McCullagh and J. A. Nelder Frs, *Generalized Linear Models*, II. London: Chapman and Hall, 1989.

[32] V. R. S. Nastiti, Y. Azhar, and R. S. Putri, "Logistic Regression Using Hyperparameter Optimization on COVID-19 Patients' Vital Status," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 1, pp. 681–687, 2023, doi: 10.29207/resti.v7i34868.xxx.

[33] I. Haq, M. N. Aidi, A. Kurnia, and E. Efriwati, "A Comparison of Logistic Regression and Geographically Weighted Logistic Regression (GWLR) on Covid-19 Data in West Sumatra," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 3, pp. 1749–1760, Sep. 2023, doi: 10.30598/barekengvol17iss3pp1749-1760.

[34] Kemendes PDTT RI, "Peringkat Indeks Desa Membangun Tahun 2018," Jakarta, 2018.

[35] Kemendes PDTT RI, "Status Indeks Desa Membangun Provinsi Kabupaten Kecamatan Tahun 2019," Jakarta, 2019.

[36] Kemendes PDTT RI, "Peringkat Status Indeks Desa Membangun (IDM)," Jakarta, 2020.

[37] Kemendes PDTT RI, "Peringkat Indeks Desa Membangun Tahun 2021," Jakarta, 2021.

[38] Kemendes PDTT RI, "Peringkat Nilai Rata-Rata Indeks Desa Membangun Tahun 2022," Jakarta, 2022.

[39] BPS, "STATISTIK POTENSI DESA INDONESIA (VILLAGE POTENTIAL STATISTICS OF INDONESIA)," Jakarta, 2021.

[40] N. R. Draper and H. Smith, *Applied Regression Analysis*, Third Edition. Canada: A Wiley-Interscience publication, 1998.

[41] A. Zeinal, "Generalized two-parameter estimator in linear regression model," *Journal of Mathematical Modeling*, vol. 8, no. 2, pp. 157–176, Mar. 2020, doi: 10.22124/jmm.2020.14903.1353.

[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Second. California: Springer, 2008.

[43] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann Stat*, vol. 32, no. 2, pp. 407–451, Apr. 2004, doi: 10.1214/009053604000000067.

[44] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," *arXiv: Statistics Theory*, Jan. 2010, [Online]. Available: http://arxiv.org/abs/1001.0736

[45] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A SPARSE-GROUP LASSO," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[46] Kemendes PDTT RI, "Standar Oprasional Prosedur (SOP) Update Data Indeks Desa Membangun Tahun 2020," Jakarta, 2020.

[47] BPS, "Kuesioner Pemuktahiran Data Perkembangan Desa 2020," Jakarta, 2020.

[48] P. K. Dunn and G. K. Smyth, *Generalized Linear Models With Examples in R*. New York: Springer, 2018. [Online]. Available: http://www.springer.com/series/417

[49] Y. E. Prasetya, A. R. T. Hidayat, and D. Dinanti, "Village Development Index of Probolinggo Coastal Villages Case study: Bhinor Village, Paiton District," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Oct. 2019. doi: 10.1088/1755-1315/328/1/012056.

[50] P. M. A. Saputra, "Understanding the Dynamics of Village Economic Activities and Development in a Developing Country: A Case Study in Java Island, Indonesia," *Sodality: Jurnal Sosiologi Pedesaan*, vol. 11, no. 1, pp. 43–58, May 2023, doi: 10.22500/11202344252.

