Jurnal Mantik, 6 (1) (2022) 271-277

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

DETERMINING FACTORS IDENTIFYING POVERTY RATE DUE TO COVID-19 PANDEMIC IN NORTH SUMATERA USING THE LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) METHOD

Nur Indah Sari[1], Hendra Cipta[2], Muhammad Fathoni[3]

[1.2]Department of Mathematics, Universitas Islam Negeri Sumatera Utara Medan
[3]Politeknik Unggul LP3M Medan

E-mail: nurindahsari1220@gmail.com, hendracipta@uinsu.ac.id, mhd.fathoni@gmail.com

ARTICLEINFO

Article history:
Received: 04/02/2022
Revised: 12/3/2022
Accepted: 27/04/2022

Keywords:

Covid-19, Poverty, Multicollinearity, LASSO, LARS

ABSTRACT

Province of North Sumatera is there in ranked 17th with the highest number of poor society in March 2020. Over the last three years, the poverty rate in North Sumatera has decreased in number and percentage. Then there was occur in increase poverty in terms of number and percentage of poor society in the March 2020 to September 2020 period caused by the covid-19 pandemic. This research was conducted to determining factors identifyin poverty due to the Covid-19 pandemic. LASSO is a method that can overcome the problem of multicollinearity by reducing the coefficient of the variable to zero or close to zero at the same time as a variable selection. The computation LASSO using the LARS algorithm. In this research, the factors that identifying poverty due to the Covid-19 pandemic are the open unemployment rate factor, the human development index, the district/city minimum wage and the number of unemployed 15 years and over with a classification accuracy of 98.1%.

1. Introduction

Covid-19 has been a public concern since its appearance was detected in China for the first time in early 2020 [1]. Covid-19 pandemic has an impact not only on the medical world, but also on the social, economic, and increasing of the poor society. The impact of the Covid-19 outbreak resulted in reduced labor supply, unemployment, reduced income, decreased consumption due to variations in consumer preferences for each item, as well as increased public sensitivity to disease and changes in economic conditions [2]. Social restrictions imposed by the government caused the entire community to be affected, especially the lower middle income class and day laborers. People who previously did not belong to poor society finally have the opportunity to become poor because of this wide scale restriction. As a result, the economic condition of the community is depressed and people lives become depressed [3].

According to Bappenas [5]-[6], poverty is a condition when an individual or group of individuals cannot fulfill their basic rights and build a noble life. These basic rights include fulfillment of food, welfare, training, business, shelter, clean water, protection, fixed assets and climate, the belief that everything is good from the treatment or danger of violence in the choice to take an interest in sociopolitical life.

Based on data from BPS North Sumatera [5], the poverty rate increased by 0.39 points, from 8.75% in March 2020 to 9.4% in September 2020. This poverty rate is equivalent to 1.36 million people in September 2020 or decreased about 73 thousand people in the last semester. While the percentage of poor society in Indonesia in September 2020 amounted to 10.19% increased by 0.41 points which was previously 9.78% in March 2020. The percentage increase in September 2020 is as a result of covid-19. North Sumatera occupies

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

the 17th position which has the highest poverty rate of 34 provinces. The percentage of poverty rate in Indonesia is 9.78% and in North Sumatera is 8.75% in March 2020.

Based on data from BPS North Sumatera [5], the number of poor society in North Sumatera in September 2020 reached 9.4% this figure increased when compared to 2019 which reached 8.63%. There are 3 districts with the highest poverty rate are West Nias Regency at 25.69%, North Nias Regency at 25.07%, and South Nias regency at 16.74%. Meanwhile, the 3 districts with the lowest percentage rate are Deli Serdang Regency at 3.88%, Binjai at 5.71% and Padangsidempuan at 7.40%. This explains that poverty rate in 25 districts and 8 cities in North Sumatera province is still uneven and some urban districts have a fairly high poverty rate. This shows that government's efforts to reduce poverty levels have not been evenly distributed in all districts. Therefore, it is necessary to seek factors that can affect poverty rate in all districts so that it can be used as a reference for each district with the aim of poverty alleviation problem.

Based on the description above, an analysis is needed to determine for identify what factors affect poverty due to the impact of the covid-19 pandemic in North Sumatera. The data will be used is data on the number of poor society in 33 districts in 2020. The regression analysis can explain the relationship between one dependent variable and two other independent variables. According to Gujarati [8]-[9], multiple regression analysis is used to see the influence of least two or more independent variables dependent or to indicate the presence or absence of a functional relationship between at least two independent variables and dependent variables. All assumptions of simple regression also apply in multiple regressions. In linear regression analysis, there is often a problem that a high correlation between multicollinearity)\. The impact of multicollinearity on the smallest square method specially the estimator value has variations and covariance that so large independent variables are not statistically measurable affecting dependent variables [7], [15]-[17].

From the number data of poor society in 33 districts of North Sumatera Province in 2020, there are independent variables that contain multicollinearity so that if used the ordinary least square model, it will produce large variances so that predictions on the model become inaccurate. In this reseach will be used analysis of the Least Absolute Shrinkage and Selection Operator (LASSO) method [21]-[22]. This method is a technique can shrink the regression coefficient on independent variables that have high correlations and errors with the intention of the regression coefficient becomes almost non existent or even equal to zero, so its method takes the role of variable selection technique while the problem has multicollinearity. Multicollinearity problems occur when one independent variable corresponds to another independent variable. If this happens, it can the resulting regression coefficient appraiser to be ineffective [23].

2. Methods

2.1 Place and Time of Research

This research was conducted at BPS of North Sumatera Province.Time this research was started from April 2021 to October 2021.

2.2 Types of Research

Quantitative research is used in this research by using secondary data obtained from BPS and the Manpower Office of North Sumatera Province.

2.3 Research Variable

Reseach variabel can be see below:

Table 1
Research Variable

Variable	Variable Name	Size
Y	Number of poor society in 2020	Percentage
X_1	Economic growth	Rupiah
X_2	Open unemployment rate	Percentage
X_3	Human development index	Percentage
X_4	Household consumption expenditures	Percentage
X_5	Number of people laid off	Person
X_6	Regency/city minimum wage	Rupiah
X_7	Number of unemployed 15 years above	Person

2.4 Research Procedures

The steps in this research as follows:

1) This research began by identifying problems arising from the covid-19 pandemic

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

2) Collecting data BPS and the Manpower Office
3) Create the descriptive statistics
4) Estimate models using Ordinary Least Square (OLS) method
5) Test independent variables by seeking VIF values to find out if multicollinearity issues occur
6) Standardize data on variables to distribute N(0,1)
7) Analyze data using Least Absolute Shrinkage and Selection Operator (LASSO) method with LARS Algorithm through R Studio software
8) Determine the best model using cross validation
9) Comparing the results of the Ordinary Least Square (OLS) regression model with the Least Absolute Shrinkage and Selection Operator (LASSO) model.

3. Results and Discussions

3.1 Data Description

Data description is described in the table below:

Table 2
Data Description

Variable	Symbol	Size
Economic growth	GDP	Rupiah
Open Unemployment Rate	OUR	Percentage
Human Development Index	HDI	Percentage
Household Consumption Expenditure	HCE	Percentage
Number of people laid off	TOE	Person
Regency/city minimum wage	MW	Rupiah
Number of unemployed 15 years above	NOU	Person

3.2 Descriptive Statistical Analysis

Descriptive analysis aims to see the characteristics or overview of the number people due to the covid-19 pandemic in North Sumatera in 2020. It results obtained through the SPSS program:

Table 3
Statistical Analysis Data

Variable	N	Minimum	Maximum	Mean	Std. Deviation
PM (Y)	34	4.59	1283.29	75.4879	216.02774
GDP (X_1)	34	22213405.00	105532865.0	45735005.97	19104063.17
OUR (X_2)	34	84	11.50	5.4956	2.83206
HDI (X_3)	34	61.51	80.98	70.7829	4.53845
HCE (X_4)	34	46.29	62.05	56.1518	4.03226
TOE (X_5)	34	67.00	681.00	161.2059	107.18853
MW (X_6)	34	2146073.00	3603246.00	2749852.119	281430.4063
NOU (X_7)	34	474.00	507805.00	29841.2059	88483.27707

Based on table 3 above resulting descriptive statistical analysis data include sums, minimum values, maximum values, average values and standard deviations. Next, calculate the regression equation using the Ordinary Least Square (OLS) method and obtained the best model with this equations:

$$\hat{y} = -46{,}21488192 + 2{,}68214\text{E} - 07X_1 - 5{,}207987885X_2 - 0{,}112932232X_3 + 2{,}171936174X_4 + 0{,}039156164X_5 - 2{,}00421\text{E} - 05X_6 + 0{,}002421028X_7 \quad (1)$$

3.3 Multicollinearity Test

The multicollinearity test is carried out to find out whether there is a multicollinearity problem or not by seeking the value of the Inflation Factor Variant (VIF). The output results of the multicollinearity test using the R Studio software

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Table 4
Multicollinearity Test

Variable	VIF Value
GDP	2.031505
OUR	2.606595
HDI	2.756987
HCE	3.605905
TOE	11.799837
MW	1.706213
NOU	9.591191

Based on Table 4, the results of multicollinearity testing by seeking at the VIF value in each independent variable obtained: 1 variable has value VIF > 10 namely variable layoffs, 1 variable almost 10 namely JPK variable and other variables have value VIF < 10 namely PDRB, TPT, HDI, KRT, and UMK variables. From the results of VIF test, this research indicates the occurrence of multicollinearity in the model which means the existence of linear relationships between independent variables, seen from the variable layoffs and JPK whose have serious multicollinearity problems with high VIF values. Because there is a multicollinearity problem in the data if the MKT model is used, it will produce a large variance so that the predictions on the model become inaccurate [12]. Therefore, the regression equation model can be solved using the Least Absolute Shrinkage and Selection Operator (LASSO) method.

3.4 Regression *Least Absolute Shrinkage and Selection Operator* (LASSO) with LARS Algorithm

Based on the VIF test, the data contained multicollinearity problems so that this problem could be solved using the LASSO method where calculated and assisted by the LARS package downloaded in R Studio software. Furthermore, the LASSO method stage is calculated using a numerical approach using the LARS algorithm. Before entering the LASSO method, The data must be standardized or normalizes by converting data to be standardized score.

Table 5
Variable Coefficients Calculated Using The Lars Algorithm

Phase	Free variables entered into the model						
	GDP	OUR	HDI	HCE	TOE	MW	NOU
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0.9043930
2	0	0	0	0	0	-0.001770782	0.9061638
3	0	0	-0.006759797	0	0	-0.006251921	0.9139309
4	0	-0.01425594	-0.017725185	0	0	-0.015318922	0.9362762
5	0	-0.03261742	-0.032337856	0.01333780	0	-0.031124071	0.9774892
6	0	-0.04292788	-0.059861076	0.02622472	0.07644015	-0.033862961	0.9366948
7	0.01235695	-0.04771933	-0.068652164	0.02717593	0.08853440	-0.040381134	0.9294595
R-squared: 0.981							

From table 5 indicates that to get a candidate the lasso coefficient value is carried out with 7 stages with all variable coefficients initially set to be worth 0. From the stage starting from point 0 to stage 7 the presumtion of LASSO coefficient will change when the selection of variables at each stage is carried out. Furthermore obtained the value R-squared (R^2) = 0.981, this indicates that 98.1% of factors affecting dependent variables can be explained by independent variable variables and the rest explained by other variable variables or other factors outside the model.

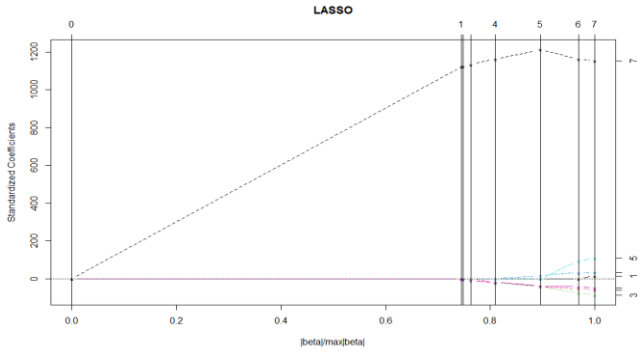

Figure 1 step of the LARS algorithm

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Based on figure 1 describes each step of the LARS algorithm in determining the lasso regression coefficients on the plot against $s = \frac{t}{\sum|\hat{\beta}_j^{OLS}|}$, dimana $t = \sum|\hat{\beta}_j^{LASSO}| . \sum|\hat{\beta}_j^{OLS}|$ is the absolute number of the estimated OLS coefficient and $\sum|\hat{\beta}_j^{LASSO}|$ is the absolute number of the presumptive LASSO coefficient while s in the plot is expressed by $\frac{|beta|}{max|beta|}$.

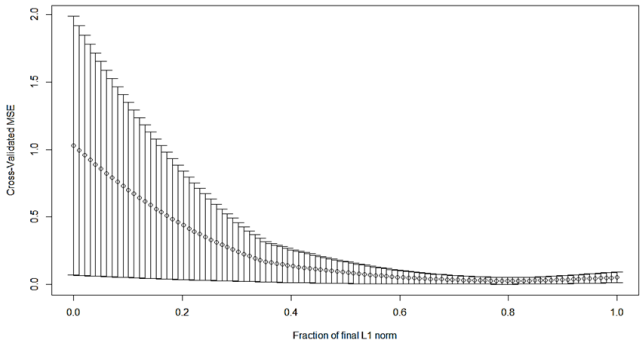

Figure 2 Determine of the interval LARS algorithm

Based on figure 2, the minimum s value is located at an interval of 0.7< s <0.9. The minimum CV value may vary each time the function is invoked. The minimum *s* value describes the stop of the iteration step to determine the coefficient of the best model of the LASSO method. The value of *s* at each stage is indicated:

Table 6
The Value Of At Each Stage Using LASSO Method

Phase	$\sum\|\hat{\beta}_j^{LASSO}\|$	$s = \frac{\sum\|\hat{\beta}_j^{LASSO}\|}{\sum\|\hat{\beta}_j^{OLS}\|}$
1	0.9043930	0,7447981
2	0,907934582	0,7477147154
3	0,926942618	0,7633684734
4	0,983576247	0,8100081748
5	1,086906347	0,8951039932
6	1,176011587	0,9684851602
7	1,214279408	1

Based on the minimum CV of plot in Figure 2 the minimum CV value is located between intervals of 0.7<s<0.9. So that, in Figure 2 can be known the optimal *s* value shows the best LASSO regression model, namely at stage of 4th when the value $s = 0{,}8100081748$ and will be calculated the value of the tuning parameter (t).

$$t = s \times \sum|\hat{\beta}_j^{OLS}|$$

$$t = 0{,}8100081748 \times 1{,}214279408 = 0{,}983576247$$

Then obtained the tuning parameter value (t) amounted to $0{,}983576247$. Based on calculations that have been done in cross validation criteria using fraction mode, the LASSO method produces the best model from the number of data poor society in North Sumatera in 2020 in the stage of 4th model with the best model:

$$PM = -46{,}21488192 - 0.01425594\ OUR - 0.017725185\ HDI - 0.015318922\ MW + 0.9362762\ NOU \quad (2)$$

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Table 7
VIF Value After Analysis Using Lasso Regression

Variable	VIF Value
HDI	1.481182
HCE	1.429451
MW	1.144511
NOU	1.085376

Based on Table 7, VIF value after analysis using LASSO regression decreases. This proves that LASSO regression can overcome multicollinearity problems while being able to be a variable selection. The resulting model is simpler and free from multicollinearity problems.

3.5 Comparison of OLS and LASSO Regression Model Coefficients

Tabel 8
Comparison Of OLS and LASSO Regression Model Coefficients

Variable	OLS	LASSO
Intersep	−46,21488192	−46,21488192
GDP	2,68214E − 07	0.0000000
OUR	−5,207987885	−0.01425594
HDI	−0,112932232	−0.017725185
HCE	2,171936174	0.0000000
TOE	0,039156164	0.0000000
MW	−2,00421E − 05	−0.015318922
NOU	0,002421028	0.9362762

Based on table 8, it is known that LASSO regression on selected variables can shrink the coefficient of variables towards zero such as the function in the LASSO regression itself. The coefficient resulting from lasso regression is better than the OLS coefficient. There are 3 variables that managed to shrink to exactly zero, namely GDP, HCE and the TOE variable, so these three variables are not included in the model. Through the selection of these variables, the model will be simpler and can overcome the problem of multicollinearity in the data.

4. Conclusion

The identification factors that affect poverty due to Covid-19 pandemic in North Sumatra in 2020 are determined through using the Least Absolute Shrinkage and Selection Operator (LASSO) method resulted in 4 variables that had a significant effect namely the Open Unemployment Rate (OUR), Human Development Index (HDI), Regency/City Minimum Wage (MW) and The Number of Unemployed 15 years above (NOU) with a classification accuracy of 98.1%. With the best model produced the Least Absolute Shrinkage and Selection Operator (LASSO) method using the LARS algorithm as follows:

$$PM = -46{,}21488192 - 0.01425594\ \text{OUR} - 0.017725185\ HDI - 0.015318922\ MW + 0.9362762\ NOU.$$

5. References

[1] Hanoatubun, S. Dampak Covid-19 Terhadap Perekonomian Indonesia". *Journal of Busines & Entrepreneurship.* vol 2. pp. 83-92. 2020.
[2] Kurniasih, E. P. "Dampak Pandemi Covid-19 Terhadap Penurunan Kesejahteraan Masyarakat Kota Pontianak. Prosiding Seminar Akademik Tahunan Ilmu Ekonomi dan Studi Pembangunan 2020". pp. 277-289. 2010.

[3] Robbani, M. "Faktor-Faktor Yang Mempengaruhi Inflasi Yang Terjadi di Indonesia Dari Tahun.2014-2017 dengan.Menggunakan Metode LASSO". *Jurnal EurekaMatika*. vol.7. 2019.
[4] Andana A.P, Safitri. D, Rusgiyono. A."Model Regresi Menggunakan *Least Absolute Shrinkage Selection Operator* (LASSO) pada Information Banyaknya Gizi Buruk Kabupaten/Kota di Jawa.Tengah*". Jurnal Gaussian*,. vol. 6. pp. 21-30. 2017.
[5] Bappenas. *Rencana Strategik Penanggulangan Kemiskinan di Indonesia*. Jakarta. 2004.
[6] BPS. *Profil Kemiskinan Provinsi Sumatera Utara Maret 2020.* Medan: BPS Sumatera Utara. 2020.
[7] Dewi. Y. S. "OLS, LASSO dan PLS Pada Data Mengandung Multikolinearitas". *Jurnal Ilmu. Dasar*. vol. 11. pp. 83-91.
[8] Efron, B. "Least Angle Regression". *The Annals of Statistics.* vol. 32. pp. 407-499. 2004.
[9] Hastie. T, Tibshirani. R, and Friedman. J.*The Elements of Statistical Learning: Data mining, Inference and Prediction.*. Springer, New York. 2008.
[10] Jolliffe, I.T. *Principle Component Analysis Second Edition.* New York: Springer-Verlag. 2002.
[11] Sartika, I, Debataraja, N., dan Imro'ah, N. "Analisis Regresi dengan Metode *Least Absolute Shrinkage Selection Operator* (LASSO) dalam Mengatasi Multikolinearitas". *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*. vol. 9. pp. 31-38. 2020.
[12] Ghozali, I. *Aplikasi Analisis Multivariat dengan Program IBM SPSS 19.* Semarang: Badan Penerbit Universitas Diponegoro. 2011.
[13] Prabowo, F.K, Rusgiyono, A, Wilandari. Y. "Pemodelan Pertumbuhan Ekonomi Jawa Tengah Menggunakan Pendekatann *Least Absolute Shrinkage Selection Operator* (LASSO)". *Jurnal Gaussian*. vol. 4. pp. 855-864. 2015.
[14] Zhao,.P., and .Yu, B. "On.Model Selection.Consistency of.Lasso. Journal.of Machine.Learning.Reaserch". vol. 7. pp. 2541-2562. 2006.
[15] R. Tibshirani. "Regression Shrinkage and Selection via The LASSO". *Journal Royal Statistic Soc B*, vol. 58, pp. 267 – 288. 1996.
[16] A. Mohammad, and H.S. Nghiemb. "Do Instructional Attributes Pose Multicollinearity Problems An Empirical Exploration." *Economic Analysis and Policy*, vol. 3, pp. 351–61. 2010.
[17] A. David. "Science Activity and Toxicity Modelling of Some NCI Selected Compounds against Leukemia P388ADR Cell Line Using Genetic Algorithm Multiple Linear Regressions" *Journal of King Saud University*. 2018.
[18] M. Dyar. "Spectrochimica Acta Part B Comparison of Partial Least Squares and Lasso Regression Techniques as Applied to Laser-Induced Breakdown Spectroscopy of Geological Samples". *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 70. pp. 51–67. 2012.
[19] Philippe. G. William. S, and Alain. B. "Sound Quality Prediction Based on Systematic Metric Selection and Shrinkage : Comparison of Stepwise, Lasso, and Elastic-Net Algorithms and Clustering Preprocessing." *Journal of Sound and Vibration* , vol. 400. pp. 134–53. 2017.
[20] Hastie. T, Efron. B, Johnstone. I, Tibshirani. R. "Least Angle Regression". *Annals of Statistics*, vol. 32. pp. 407-499. 2004.
[21] Kutner, M.H., Nachtsheim. C, and Neter. J. *Applied Linear Regression Models. 4th ed*. New York: Mc GrawHill Companies, Inc. 2017.
[22] Sermpinis. G, Serafeim. T, and Ping. Z. "Modelling Market Implied Ratings Using LASSO Variable Selection Techniques." *Journal of Empirical Finance*. vol. 48. pp. 19–35. 2018.
[23] Wang, S. "Predicting Ship Fuel Consumption Based on LASSO Regression." *Transportation Research*. 2017.