OPEN ACCESS
MES: *Journal of Mathematics Education and Science*
ISSN: 2579-6550 (online) 2528-4363 (print)
Vol. 10, No. 1, Oktober 2024

Situs web: https://jurnal.uisu.ac.id/index.php/mesuisu
Email: jurnalmes@fkip.uisu.ac.id

PERBANDINGAN REGRESI LASSO DAN *PRINCIPAL COMPONENT REGRESSION* DALAM MENGATASI MASALAH MULTIKOLINEARITAS

Elsa Fadillah Nasution*
Universitas Sumatera Utara, Medan, Sumatera Utara, Indonesia, 20155

Rahmawati Pane
Universitas Sumatera Utara, Medan, Sumatera Utara, Indonesia, 20155

Abstrak. *Multikolinearitas merupakan suatu permasalahan yang sering muncul dalam analisis regresi. Keberadaan multikolinearitas dapat mengakibatkan estimasi parameter regresi pada metode Ordinary Least Square menjadi tidak efisien. Oleh karena itu, untuk mengatasi hal tersebut digunakan metode Regresi LASSO dan metode Principal Component Regression (PCR). Data yang digunakan dalam penelitian ini adalah data bangkitan berasal dari tingkat korelasi* (r) *rendah* $(0{,}1-0{,}3)$, *sedang* $(0{,}4-0{,}6)$, *dan tinggi* $(0{,}7-0{,}9)$ *dengan ukuran sampel berbeda* $(n = 20{,}40{,}120{,}200)$ *dari distribusi normal dengan* 30 *dan* 60 *variabel independen. Kinerja metode Regresi LASSO dan metode Principal Component Regression (PCR) dievaluasi menggunakan nilai Mean Square Error (MSE) dan nilai koefisien determinasi* (R^2). *Berdasarkan penelitian ini diperoleh metode Regresi LASSO memiliki efisiensi yang lebih baik dibandingkan metode Principal Component Regression (PCR) karena metode Regresi LASSO memperoleh nilai Mean Square Error (MSE) lebih kecil dan nilai koefisien determinasi* (R^2) *lebih tinggi dibandingkan metode Principal Component Regression (PCR).*
Kata Kunci: *Multikolinearitas, Principal Component Regression (PCR), Regresi LASSO*

Abstract. *Multicollinearity is a problem that often arises in regression analysis. If the assumption of the absence of multicollinearity is not met, researchers will have difficulty in identifying independent variables that have a significant effect in the regression model. The presence of multicollinearity can cause the estimation of regression parameters in the Ordinary Least Square (OLS) method to be inefficient. To overcome this, the LASSO Regression method and the Principal Component Regression (PCR) method are used. The data used in this study are generation data derived from low (0,1-0,3), medium (0,4-0,6), and high (0,7-0,9) correlation levels with different sample sizes (n=20,40,120,200) from normal distribution with 30 and 60 independent variables. The performance of LASSO Regression method and Principal Component Regression (PCR) method is evaluated using Mean Square Error (MSE) value and coefficient of determination* (R^2) *value. Based on this research, the LASSO Regression method has better efficiency than the Principal Component Regression (PCR) method because the LASSO Regression method obtains a smaller Mean Square Error (MSE) value and a higher coefficient of determination* (R^2) *value than the Principal Component Regression (PCR) method.*
Keywords: *Multicollinearity, Principal Component Regression (PCR), LASSO Regression*

Sitasi: Nasution, E.F., Pane, R. 2024. Perbandingan Regresi LASSO dan Principal Component Regression dalam Mengatasi Masalah Multikolinieritas. *MES* (*Journal of Mathematics Education and Science*), 10(1): 120-128.

Submit:	**Revise:**	**Accepted:**	**Publish:**
16 Juli 2024	18 Juli 2024	16 Agustus 2024	10 Oktober 2024

PENDAHULUAN

Analisis regresi adalah suatu kajian terhadap ketergantungan satu variabel, yaitu variabel dependen, terhadap satu atau lebih variabel lain yang disebut variabel independen (Gujarati & Porter, 2008). Untuk mengestimasi parameter regresi, salah satu metode yang sering digunakan adalah *Ordinary Least Square* (OLS). Dalam analisis regresi membutuhkan asumsi-asumsi

*Corresponding Author: *elsafdllhnst@gmail.com*

tertentu agar dapat menghasilkan sebuah model regresi yang baik. Asumsi-asumsi tersebut antara lain normalitas, tidak mengalami multikolinearitas, homoskedastisitas, dan tidak terjadi autokorelasi (Sarwoko, 2005). Jika satu atau lebih asumsi dilanggar, maka model yang ada tidak dapat diandalkan dalam memperkirakan parameter populasi.

Multikolinearitas merupakan suatu permasalahan yang sangat sering terjadi pada analisis regresi. Jika asumsi tidak mengalami multikolinearitas tidak terpenuhi, peneliti akan menghadapi kesulitan dalam mengidentifikasi variabel independen yang memiliki dampak yang signifikan dalam model regresi (Irwan, 2015). Keberadaan multikolinearitas dapat mengakibatkan estimasi parameter regresi pada metode OLS menjadi tidak efisien, karena kondisi tersebut dapat menghasilkan variansi parameter yang besar pada regresi berganda. Hal tersebut juga menyebabkan nilai variabel independen tidak memiliki pengaruh signifikan terhadap variabel dependen meskipun nilai koefisien determinasinya (R^2) sangat tinggi, sehingga dapat menghasilkan model yang tidak baik. Oleh karena itu, perlu dilakukan tindakan lebih lanjut untuk mengatasi masalah multikolinearitas.

Permasalahan multikolinearitas dapat diatasi dengan menyusutkan (*shrinkage*) koefisien yang diestimasi. Salah satu pendekatan *shrinkage* yang sering digunakan adalah metode *Least Absolute Shrinkage and Selection Operator* (LASSO). *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan suatu metode regresi yang dapat mereduksi koefisien regresi menjadi tepat nol. LASSO meminimumkan jumlah kuadrat eror yang bergantung pada jumlahan nilai mutlak dari koefisien (Tibshirani, 1996). Metode ini hanya memasukkan variabel-variabel independen terbaik ke dalam model sehingga metode ini dapat digunakan sebagai seleksi variabel pada model, yang mana hal ini juga bermanfaat untuk memudahkan proses interpretasi model regresi (Robbani et al., 2019). Selain itu, multikolinearitas juga dapat ditangani dengan metode *Principal Component Regression* (PCR). Metode ini menggabungkan antara analisis regresi dan *Principal Component Analysis* (PCA). Metode PCA merupakan suatu teknik statistik yang dapat membuat satu set variabel baru yang lebih kecil dan tidak berkorelasi dengan cara mengubah dari sebagian variabel asli yang digunakan dan saling berkaitan antara yang satu dengan lainnya (Delsen et al., 2017). Metode PCR adalah analisis regresi yang melibatkan variabel independen terhadap komponen-komponen utama yang tidak berkorelasi satu sama lain, yang mana setiap komponen utama merupakan gabungan linear dari semua variabel dependen (Draper & Smith, 1992).

METODE

Metode penelitian yang digunakan adalah studi literatur yaitu mengumpulkan dan mempelajari data-data pustaka melalui buku dan artikel penelitian terdahulu yang relevan dengan permasalahan dalam penelitian ini. Dalam penelitian ini, data yang digunakan adalah data bangkitan dengan menggunakan *software R*. Proses simulasi dimulai dengan membangkitkan data yang berasal dari tiga tingkat korelasi (r), yaitu rendah $(0{,}1-0{,}3)$, sedang $(0{,}4-0{,}6)$, dan tinggi $(0{,}7-0{,}9)$ dengan ukuran sampel berbeda ($n = 20{,}40{,}120{,}200$) dengan 30 dan 60 variabel independen yang berdistribusi $N{\sim}(0{,}1)$. Perbedaan koefisien korelasi mengindikasikan bahwa data bangkitan berasal dari tingkatan multikolinearitas yang berbeda, yaitu rendah, sedang, dan tinggi. Setelah membangkitkan data, data bangkitan yang mengandung masalah multikolinearitas tersebut akan diatasi dengan metode Regresi LASSO dan metode *Principal Component Regression* (PCR) untuk selanjutnya akan dianalisis metode manakah yang terbaik dalam menyelesaikan permasalahan multikolinearitas tersebut dengan menggunakan nilai *Mean Square Error* (MSE) dan nilai koefisien determinasi (R^2).

Regresi Linear Berganda

Regresi linear berganda digunakan untuk meramalkan dampak dua atau lebih variabel independen pada variabel dependen atau untuk menguji apakah terdapat hubungan fungsional antara dua variabel independen atau lebih dengan satu variabel dependen (Sartika et al., 2020).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (1)$$

atau

$$y = X\beta + \varepsilon \qquad (2)$$

di mana

y = vektor pengamatan yang berukuran $n\ x\ 1$,
X = matriks $n\ x\ p$ dari variabel independen,
β = vektor $p\ x\ 1$ dari koefisien regresi,
ε = vektor kolom dari galat yang berukuran $n\ x\ 1$.

Parameter $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ bisa didapatkan melalui *Ordinary Least-Square* (OLS) atau metode kuadrat terkecil. Dengan metode kuadrat terkecil dapat diperoleh estimator sebagai berikut:

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (3)$$

di mana $X'X$ adalah matriks simetris $p\ x\ p$ dan $X'y$ adalah vektor kolom $p\ x\ 1$ (Montgomery et al., 2012)

Multikolinearitas

Multikolinearitas adalah suatu bentuk kondisi yang tidak baik pada matriks $X'X$. Selain itu, masalahnya adalah pada salah satu derajatnya, yaitu setiap kumpulan data akan mengalami multikolinearitas sampai batas tertentu kecuali kolom X ortogonal ($X'X$ adalah diagonal matriks). Adanya multikolinearitas membuat analisis model regresi dengan metode kuadrat terkecil menjadi tidak signifikan. Multikolinearitas dapat disebabkan oleh beberapa faktor (Paul, 2006). Berikut empat faktor utama yang menjadi penyebab terjadinya multikolinearitas.

1. Metode pengumpulan data yang digunakan
2. Kendala pada model atau populasi yang dijadikan sampel
3. Spesifikasi model
4. Model yang *overdetermined*

Multikolinearitas dapat dideteksi melalui beberapa cara, salah satunya adalah dengan menggunakan pemeriksaan matriks korelasi. Koefisien korelasi dapat memberikan informasi mengenai keterkaitan antara dua variabel, x dan y, dan tidak digunakan untuk memperkirakan nilai y berdasarkan variabel x. Nilai koefisien korelasi didefinisikan sebagai berikut (Boediono, 2001):

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \qquad (4)$$

Selain itu, multikolinearitas juga dapat dideteksi dengan nilai *Varians Inflation Factor* (VIF). Nilai VIF dapat didefinisikan sebagai berikut:

$$VIF_j = C_{jj} = \left(1 - R_j^2\right)^{-1} \qquad (5)$$

Nilai VIF yang melebihi 5 atau 10 menunjukkan bahwa adanya indikasi bahwa koefisien regresi terkait tidak dapat diestimasi dengan baik karena adanya multikolinearitas (Paul, 2006).

Metode LASSO

Metode LASSO yang dikembangkan oleh Tibshirani, pertama kali diperkenalkan pada tahun 1996 sebagai suatu teknik untuk menyeleksi variabel dengan mengestimasi parameter. Metode ini menyusutkan koefisien regresi dari variabel yang berkorelasi tinggi dengan galat. Tujuannya adalah supaya koefisien regresi tersebut mendekati nol atau bahkan menjadi nol. Oleh karena itu, metode LASSO tidak hanya berfungsi sebagai alat untuk menyeleksi variabel, tetapi juga efektif dalam mengatasi masalah multikolinearitas (Sartika et al., 2020). Regresi LASSO diperoleh dengan meminimumkan jumlah kuadrat galat dengan persamaan sebagai berikut:

$$\hat{\beta} = \underset{\beta}{argmin} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \quad (6)$$

dengan

y_i = nilai variabel independen ke-i
β_j = koefisien regresi variabel independen ke-j
x_{ij} = pengamatan ke- i variabel independen ke-j
i = $1,2,\ldots,n$; n adalah banyaknya pengamatan
j = $1,2,\ldots,k$; k adalah banyaknya variabel independen

dengan kendala $\sum_{j=1}^{k}|\beta_j| \leq t$ dengan $t \geq 0$ dan $t = \sum_{j=1}^{k}|\hat{\beta}_j|$, $\hat{\beta}_j$ adalah penduga LASSO pada setiap tahapan seleksi variabel. Nilai t merupakan parameter *tuning* yang mengontrol penyusutan koefisien regresi pada metode LASSO (Sartika et al., 2020).

Estimasi pada regresi LASSO juga dapat dituliskan ke dalam persamaan Lagrange sebagai berikut (Mait et al., 2021):

$$\hat{\beta} = \underset{\beta}{argmin} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \quad (7)$$

Dalam persamaan (7), kendala $\lambda \sum_{j=1}^{k}|\beta_j|$ disebut sebagai kendala L_1 karena $\sum_{j=1}^{k}|\beta_j| = \|\beta\|_1$ dengan $\lambda \geq 0$ yang merupakan parameter *tuning*.

***Least Angle Regression* (LARS)**

Least Angle Regression (LARS) adalah suatu metode regresi yang dapat diubah menjadi algoritma komputasi untuk menerapkan metode LASSO. Modifikasi LARS untuk LASSO menghasilkan algoritma yang efisien dalam mengestimasi koefisien LASSO. Proses LARS dimulai dengan mengatur semua koefisien menjadi nol, kemudian secara bertahap menemukan variabel independen yang paling berkorelasi dengan variabel dependen. Langkah-langkah dalam memperkirakan koefisien LASSO menggunakan algoritma LARS dapat dijelaskan sebagai berikut (Sartika et al., 2020):

1. Mencari suatu vektor yang sebanding dengan vektor korelasi $(\hat{c})$ antara variabel independen dan galat dari setiap variabel independen, yaitu:

   $$\hat{c} = X'(Y - \hat{\mu}) \quad (8)$$

   Selanjutnya menentukan nilai korelasi absolut terbesar $(\hat{C})$, melalui

   $$\hat{C} = \max|\hat{c}_j| \quad (9)$$

   sehingga diperoleh $s_j = sign\{c_j\}$ untuk $j \in A$.

2. Menetapkan X_A, di mana A adalah adalah kumpulan indeks dari variabel independen yang aktif yang dipilih berdasarkan nilai korelasi absolut terbesar. Didefinisikan dengan matriks berikut:
$$X_A = \left(\dots, s_j X_j, \dots\right) \quad (10)$$
dengan s_j bernilai ± 1.
3. Melakukan perhitungan *equingular vector* (u_a), yang mana *equingular vector* ini membagi sudut dan kolom-kolom X_A menjadi sama besar dan sudutnya kurang dari 90°. Nilai *equingular vector* diperoleh melalui prosedur sebagai berikut:
$$u_a = X_A\, w_a \quad (11)$$
di mana $w_a = A_A G_A^{-1}\, 1_A$, $A_A = (1_A' G_A 1_A)^{-\frac{1}{2}}$, $G_A = X' X_A$ yang mana 1 merupakan matriks dengan entrinya 1.
4. Melakukan perhitungan vektor *inner product*, yaitu:
$$a \equiv X' u_a \quad (12)$$
5. Menghitung nilai $\hat{\mu}_A$ dengan menggunakan rumus sebagai berikut:
$$\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma} u_a \text{ dan } A_+ = A - \{j\} \quad (13)$$
Jadi $\hat{\gamma}$ dapat diperoleh dengan persamaan sebagai berikut:
$$\hat{\gamma} = min^{+}_{j\epsilon A^c}\left\{\frac{\hat{C} - \hat{c}}{A_A - a_A}, \frac{\hat{C} + \hat{c}}{A_A + a_A}\right\} \quad (14)$$
$min^{+}_{j\epsilon A^c}$ berarti bahwa yang dipilih adalah nilai minimum positif dari j dan bukan merupakan himpunan A. Pada perhitungan akhir, nilai $\hat{\gamma}$ dapat diperoleh dengan rumus $\hat{\gamma} = \frac{\hat{c}_m}{A_m}$.
6. Algoritma LARS untuk LASSO mengharuskan tanda dari koordinat bukan nol $\hat{\beta}_j$ sama dengan tanda $\hat{c}_j$ dengan rumus berikut:
$$sign(\hat{\beta}_j) = sign(\hat{c}_j) = s_j \quad (15)$$
7. Langkah keenam bisa didapat jika $\hat{\gamma} = \underbrace{min}_{\gamma_j > 0}\{\gamma_j\}$, yang mana $\gamma_j = \frac{-\hat{\beta}_j}{s_j\ w_{aj}}$. Jika $\tilde{\gamma} < \hat{\gamma}$, maka $\beta_j(\gamma)$ bukan merupakan solusi untuk LASSO karena pembatasan tanda tidak dipenuhi dan hilangkan variabel j dari *equingular vector* selanjutnya. Variabel j dapat dimasukkan kembali ke dalam perhitungan LARS pada tahap selanjutnya.
8. Langkah-langkah yang sama diulang sehingga semua variabel independen telah terseleksi.

Metode Principal Component Regression (PCR)

Principal Component Regression (PCR) merupakan salah satu metode estimasi yang sering diterapkan ketika terdapat multikolinieritas dalam data. Metode ini pertama kali dibahas oleh Hotelling pada tahun 1933 (Küçükakçah & Gözükara Bağ, 2022). *Principal Component Regression* (PCR) menggabungkan antara analisis regresi dengan *Principal Component Analysis* (PCA). Analisis regresi digunakan untuk menguji ada tidaknya korelasi antara variabel independen dengan variabel dependen, sedangkan PCA digunakan untuk mengubah sejumlah besar variabel asli yang awalnya saling berkorelasi menjadi seperangkat variabel baru yang lebih kecil dan independen satu sama lain dengan cara menyusutkan (mereduksi) dimensinya (Pendi, 2021). PCR dapat dinyatakan dalam persamaan sebagai berikut (Johnson & Wichern, 2007):

$$Y = w_0 + w_1K_1 + w_2K_2 + \cdots + w_mK_m + \varepsilon \tag{16}$$

dengan

w_0 = Konstanta *principal component regression*
$w_1, w_2, \dots, w_m$ = Parameter *principal component regression*
$K_1, K_2, \dots, K_m$ = Komponen utama

Komponen utama diperoleh dari variabel yang dibakukan, yaitu:

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}$$
$$Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}$$
$$\vdots$$
$$Z_k = \frac{(X_k - \mu_k)}{\sqrt{\sigma_{kk}}} \tag{17}$$

Bentuk kombinasi linear dari variabel baku Z, yaitu:

$$K_1 = a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{k1}Z_k$$
$$K_2 = a_{12}Z_1 + a_{22}Z_2 + \cdots + a_{k2}Z_k$$
$$\vdots$$
$$K_m = a_{1m}Z_1 + a_{2m}Z_2 + \cdots + a_{km}Z_k \tag{18}$$

Mean Square Error **(MSE)**

Mean Square Error (MSE) digunakan untuk memberikan gambaran seberapa jauh nilai yang diprediksi oleh model dari nilai yang sebenarnya. Semakin kecil nilai MSE, maka semakin baik kinerja model dalam memprediksi data (Montgomery et al., 2012).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{19}$$

Koefisien Determinasi

Koefisien determinasi (R^2) adalah sebuah ukuran statistik yang digunakan untuk menilai seberapa baik model regresi menjelaskan variabilitas dalam data dependen berdasarkan data independen. Nilai R^2 yang besar menunjukkan bahwa model memiliki kesalahan pengujian yang kecil. Koefisien determinasi (R^2) didefinisikan dengan persamaan sebagai berikut (Montgomery et al., 2012):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{20}$$

HASIL DAN PEMBAHASAN

Berdasarkan proses simulasi yang telah dilakukan, kinerja metode Regresi LASSO dan *Principal Component Regression* (PCR) akan dibandingkan menggunakan nilai *Mean Square*

Error (MSE) dan nilai koefisien determinasi (R^2) yang telah diperoleh dari kedua metode tersebut.

Tabel 1. Nilai MSE dari metode Regresi LASSO dan metode *Principal Component Regression*

p	n	r	MSE		
			OLS	LASSO	PCR
30	20	0,1-0,3	-	0,078408	1,749236
		0,4-0,6	-	0,002432	1,165618
		0,7-0,9	-	0,181061	2,875545
	40	0,1-0,3	1,238178	0,053682	1,124746
		0,4-0,6	5,414862	0,794523	1,525849
		0,7-0,9	6,747325	0,417170	2,644469
	120	0,1-0,3	3,231679	0,013101	1,648587
		0,4-0,6	2,793727	0,184987	2,007803
		0,7-0,9	7,813376	0,145227	1,045711
	200	0,1-0,3	2,872047	0,021214	1,192972
		0,4-0,6	2,274964	0,067259	1,553494
		0,7-0,9	3,872368	0,044483	2,900932
60	20	0,1-0,3	-	0,013530	0,979488
		0,4-0,6	-	1,001528	1,495950
		0,7-0,9	-	0,650159	3,000079
	40	0,1-0,3	-	0,010670	2,505946
		0,4-0,6	-	0,002380	1,448417
		0,7-0,9	-	0,010357	4,141754
	120	0,1-0,3	1,356697	1,020952	1,054839
		0,4-0,6	5,130962	0,251359	1,960178
		0,7-0,9	3,308732	0,014663	1,213007
	200	0,1-0,3	5,302995	0,255024	2,086238
		0,4-0,6	2,025857	0,178668	1,289228
		0,7-0,9	8,685313	0,026593	3,351584

Berdasarkan Tabel 1 dapat dilihat bahwa *Ordinary Least Square* (OLS) tidak baik digunakan ketika jumlah variabel lebih besar dari ukuran sampel ($p > n$) karena tidak dapat menghasilkan nilai *Mean Square Error* (MSE) dan memiliki estimasi parameter yang buruk dengan adanya multikolinearitas bahkan ketika ($p < n$). Oleh karena itu, multikolinearitas dapat diatasi dengan menggunakaan metode Regresi LASSO dan metode *Principal Component Regression* (PCR). Metode Regresi LASSO membuktikan efisiensinya dengan memperoleh nilai *Mean Square Error* (MSE) yang lebih kecil dibandingkan metode *Principal Component Regression* (PCR).

Pemilihan metode terbaik juga menggunakan nilai koefisien determinasi (R^2) yang telah diperoleh dari kedua metode. Berdasarkan Tabel 2 diperoleh bahwa metode Regresi LASSO memiliki nilai *koefisien* determinasi (R^2) yang lebih tinggi pada setiap banyaknya variabel independen, ukuran sampel, dan tingkat korelasi yang berbeda. Hal tersebut menunjukkan jika ditinjau dari nilai koefisien determinasi (R^2), Regresi LASSO tetap menjadi metode yang lebih baik daripada PCR.

Tabel 2. Nilai R^2 dari metode Regresi LASSO dan metode *Principal Component Regression* (PCR)

p	n	r	R^2		
			OLS	LASSO	PCR
30	20	0,1-0,3	-	0,997595	0,946355
		0,4-0,6	-	0,999888	0,946814
		0,7-0,9	-	0,993155	0,891299
	40	0,1-0,3	0,629742	0,998335	0,965124
		0,4-0,6	0,449330	0,981206	0,963908
		0,7-0,9	0,422967	0,989507	0,933487
	120	0,1-0,3	0,797940	0,999561	0,944770
		0,4-0,6	0,029371	0,993324	0,902288
		0,7-0,9	0,639025	0,991408	0,938136
	200	0,1-0,3	0,585604	0,999433	0,941424
		0,4-0,6	0,093014	0,997880	0,919543
		0,7-0,9	0,661944	0,999555	0,971024
60	20	0,1-0,3	-	0,999756	0,982339
		0,4-0,6	-	0,985814	0,978811
		0,7-0,9	-	0,960296	0,816792
	40	0,1-0,3	-	0,999843	0,963239
		0,4-0,6	-	0,999969	0,981473
		0,7-0,9	-	0,999539	0,815732
	120	0,1-0,3	0,435129	0,984458	0,979376
		0,4-0,6	0,646886	0,995786	0,967138
		0,7-0,9	0,196127	0,999840	0,965157
	200	0,1-0,3	0,203278	0,995802	0,965660
		0,4-0,6	0,467821	0,995528	0,967734
		0,7-0,9	0,434569	0.999450	0,930801

Metode Regresi LASSO merupakan metode yang paling ideal dalam mengatasi permasalahan multikolinearitas jika dibandingkan dengan metode *Principal Component Regression* (PCR), baik *berdasarkan* nilai *Mean Square Error* (MSE) maupun nilai koefisien determinasi (R^2). Dalam hal ini, metode Regresi LASSO lebih efektif dan efisien untuk setiap banyaknya variabel independen, ukuran sampel, dan tingkat korelasi pada permasalahan tersebut.

KESIMPULAN

Metode *Ordinary Least Square* (OLS) adalah metode yang tidak baik digunakan ketika jumlah variabel lebih besar dari ukuran sampel ($p > n$) dan memiliki estimasi parameter yang buruk dengan adanya multikolinearitas bahkan ketika ($p < n$) karena menghasilkan kesimpulan yang salah. Metode Regresi LASSO adalah metode yang lebih baik daripada metode *Principal Component Regression* (PCR) dalam mengatasi permasalahan multikolinearitas pada setiap tingkat korelasi, ukuran sampel, dan variabel independen yang berbeda-beda karena memperoleh nilai *Mean Square Error* (MSE) lebih kecil dan nilai koefisien determinasi (R^2) lebih tinggi.

DAFTAR PUSTAKA

Boediono. (2001). Statistik dan Probabilitas. PT Remaja Rosdakarya.

Delsen, M. S. N. Van, Wattimena, A. Z., & Saputri, S. D. (2017). Penggunaan Metode Analisis Komponen Utama untuk Mereduksi Faktor-Faktor Inflasi di Kota Ambon. Jurnal Ilmu Matematika Dan Terapan, 11(2), 109–118.

Draper, N. R., & Smith, H. (1992). Applied Regression Analysis (Second). John Wiley.

Gujarati, D. N., & Porter, D. C. (2008). Basic Economterics (Fourth). McGraw-Hill.

Irwan, M. (2015). Least Square and Ridge Regression Estimation. Jurnal MSA, 3(2), 7–13.

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis (Sixth). Prentice Hall.

Küçükakçah, Z., & Gözükara Bağ, H. (2022). Comparison of Ordinary Least Squares and Principal Components Regression Analyses. The Journal Of Cognitive Systems, 7(22), 7–11. https://doi.org/10.52876/jcs.1117961

Mait, Y. A., Salaki, D. T., & Komalig, H. A. H. (2021). Kajian Model Prediksi Metode Least Absolute Shrinkage and Selection Operator (LASSO) pada Data Mengandung Multikolinearitas. Jurnal Matematika Dan Aplikasi, 10(2), 69–75. https://doi.org/10.35799/dc.10.2.2021.34909

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (Fifth). John Wiley & Sons.

Paul, R. K. (2006). Multicollinearity: Causes, Effects and Remedies. IASRI, 1(1), 58–65.

Pendi. (2021). Analisis Regresi dengan Metode Komponen Utama dalam Mengatasi Masalah Multikolinearitas. Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster), 10(1), 131–138. http://dx.doi.org/10.26418/bbimst.v10i1.44750

Robbani, M., Agustiani, F., & Herrhyanto, N. (2019). Regresi Least Absolute Shrinkage and Selection Operator (LASSO) pada Kasus Inflasi di Indonesia Tahun 2014-2017. Jurnal Eurekamatika, 7(2). https://doi.org/10.17509/jem.v7i2.22130

Sartika, I., Debataraja, N. N., & Imro'ah, N. (2020). Analisis Regresi dengan Metode Least Absolute Shrinkage and Selection Operator (LASSO) dalam Mengatasi Multikolinearitas. Buletin Ilmiah Math. Stat. Dan Terapannya, 9(1), 31–38. https://dx.doi.org/10.26418/bbimst.v9i1.38029

Sarwoko. (2005). Dasar-Dasar Ekonomterika. ANDI.

Tibshirani, R. (1996). Regression Shinkrage and Selection Via The LASSO. Journal of the Royal Statistical Society Series B (Methodological), 58(1), 267–288. https://www.jstor.org/stable/2346178