

ESTIMASI PARAMETER LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) MENGGUNAKAN ALGORITMA CYCLIC COORDINATE DESCENT

Erine Rosa Sherina Yanti Sagala
Universitas Sumatera Utara
e-mail: erinesagala12@gmail.com

INFORMASI ARTIKEL

Submitted : 2025-6-30
Review : 2025-6-30
Accepted : 2025-6-30
Published : 2025-6-30

KATA KUNCI

Cyclic Coordinate Descent (CCD), Least Absolute Shrinkage and Selection Operator (Lasso), Ordinary Least Square (OLS).

ABSTRAK

Regresi Least Absolute Shrinkage and Selection Operator (Lasso) merupakan satu metode yang dapat mengatasi masalah multikolinearitas dengan menyusutkan beberapa koefisien regresi menjadi tepat nol. Algoritma Cyclic Coordinate Descent secara efisien dan terstruktur digunakan untuk menentukan taksiran parameter terutama untuk dataset dengan banyak variabel independen. Tujuan dari penelitian ini adalah untuk memperoleh estimasi parameter yang akurat dan dapat diinterpretasikan, terutama dalam menghadapi permasalahan multikolinearitas. Untuk mengevaluasi performa metode ini, digunakan tiga dataset bangkitan dengan tingkat multikolinearitas yang berbeda: tinggi, sedang, dan rendah yang diperoleh dari bahasa pemrograman python. Hasil dari penelitian ini menunjukkan bahwa metode Lasso dengan algoritma CCD cocok digunakan pada data yang mengandung multikolinearitas tinggi, dimana terjadi penyusutan variabel menjadi nol pada variabel independen yang saling berkorelasi tinggi. Berdasarkan hasil penelitian dapat disimpulkan bahwa metode Lasso dengan algoritma CCD menghasilkan model yang lebih sederhana namun tetap memiliki akurasi prediksi yang tinggi pada data dengan multikolinearitas tinggi. Sedangkan untuk data dengan multikolinearitas sedang dan rendah, model regresi yang dihasilkan dengan metode OLS lebih unggul karena semua variabel masih berkontribusi signifikan, sehingga penalti Lasso justru dapat menurunkan performa model.

ABSTRACT

Least Absolute Shrinkage and Selection Operator (Lasso) regression is one method that can overcome multicollinearity problems by shrinking some regression coefficients to exactly zero. Cyclic Coordinate Descent algorithm is efficiently and structurally used to determine parameter estimates

Keywords: *Cyclic Coordinate Descent (CCD), Least Absolute Shrinkage and Selection Operator (Lasso), Ordinary Least Square (OLS).*

especially for datasets with many independent variables. The objective of this study is to obtain accurate and interpretable parameter estimates, especially in the face of multicollinearity problems. To evaluate the performance of this method, three generated datasets with different levels of multicollinearity: high, medium, and low were used and obtained from the python programming language. The results of this study show that the Lasso method with the CCD algorithm is suitable for data containing high multicollinearity, where there is a shrinkage of variables to zero in highly correlated independent variables. Based on the research results, it can be concluded that the Lasso method with the CCD algorithm produces a simpler model but still has high prediction accuracy on data with high multicollinearity. Meanwhile, for data with moderate and low multicollinearity, the regression model generated by the OLS method is superior because all variables still contribute significantly, so the Lasso penalty can actually reduce the model performance.

PENDAHULUAN

Analisis regresi merupakan teknik statistika untuk pemeriksaan dan pemodelan hubungan antarvariabel (Efendi et al., 2020). Analisis ini terdiri dari dua komponen yang dihubungkan, yaitu variabel dependen dan variabel independen. Variabel dependen (y) adalah variabel yang nilainya dipengaruhi oleh variabel lain, sedangkan independen (x) adalah variabel yang nilainya ditetapkan secara bebas dengan asumsi bahwa variabel tersebut memengaruhi nilai variabel dependen.

Salah satu pendekatan paling umum dalam regresi linier adalah Ordinary Least Squares (OLS), yang berupaya meminimalkan jumlah kuadrat dari selisih antara nilai observasi dan nilai prediksi. Namun, dalam kasus data yang mengandung multikolinearitas dimana terdapat korelasi tinggi antar variabel independen menyebabkan metode OLS menjadi kurang optimal dalam membentuk model prediksi yang akurat. Hal ini terjadi karena multikolinearitas menyebabkan estimasi koefisien menjadi tidak stabil, sulit diinterpretasikan, dan kurang dapat diandalkan. Oleh karena itu, diperlukan metode regresi yang mampu melakukan seleksi variabel serta regularisasi untuk meningkatkan akurasi prediksi dan interpretabilitas model, misalnya metode Least Absolute Shrinkage and Selection Operator (LASSO).

Regresi Lasso dikembangkan pertama kali oleh (Tibshirani, 1996). Regresi Lasso merupakan metode regresi berganda yang digunakan untuk menyusutkan koefisien taksiran menjadi nol dan menyeleksi variabel-variabel independen sehingga menghasilkan model dengan variabel terbaik. Dengan kata lain, regresi Lasso digunakan untuk mengatasi multikolinearitas pada analisis regresi linear berganda. Dalam penggunaannya, regresi Lasso melakukan pendugaan dengan penalti l_1 yang merupakan nilai absolut dari koefisien regresi. Dengan penalti l_1 , regresi Lasso melakukan regularisasi yang mendorong beberapa koefisien regresi menjadi nol.

Terdapat beberapa algoritma yang dapat digunakan dalam mengestimasi parameter regresi Lasso, diantaranya yaitu Least Angle Regression (LARS), Proximal

Gradient, Gradient Descent, Alternating Direction Method of Multipliers (ADMM), dan Cyclic Coordinate Descent (CCD). Meskipun algoritma LARS mampu menghasilkan lintasan solusi lengkap, serta Proximal Gradient dan ADMM efektif dalam konteks optimasi konveks, algoritma tersebut relatif kompleks dan memerlukan komputasi tinggi. Oleh karena itu, penelitian ini menggunakan algoritma Cyclic Coordinate Descent (CCD) karena bersifat sederhana, cepat, dan mudah diimplementasikan.

Cyclic Coordinate Descent (CCD) merupakan algoritma optimasi yang berguna untuk mengoptimalkan fungsi objektif dengan cara mengurangi nilai fungsi secara berulang-ulang, dengan setiap langkah mengoptimalkan fungsi terhadap satu koordinat pada satu waktu. Algoritma ini merupakan metode optimisasi yang efisien untuk masalah minimisasi yang terpisah, dimana algoritma memperbarui satu koefisien pada satu waktu sambil mempertahankan koefisien lainnya tetap konstan. Pendekatan ini sangat cocok untuk Lasso, di mana penalti l_1 menyebabkan non-linearitas dalam fungsi objektif.

Penelitian terdahulu terkait dengan model Regresi Lasso dilakukan oleh (Ningsih et al., 2023). Penelitian ini menggunakan regresi Lasso pada data tindak pidana kriminalitas di kota Tangerang. Hasil yang didapat menunjukkan tiga variabel yang berpengaruh terhadap tingkat kriminalitas yaitu, jumlah penduduk, laju pertumbuhan penduduk dan jumlah penduduk miskin. Selanjutnya Rachmawati et al., 2021 dalam penelitiannya yang berjudul “Lasso Regression for Daily Rainfall Modelling at Citeko Station, Bogor, Indonesia” menarik kesimpulan bahwa model regresi Lasso mampu memilih 9 dari 16 variabel sehingga model hanya memerlukan lebih sedikit variabel independen namun tetap mampu menjelaskan curah hujan lebih baik dibandingkan regresi linear berganda klasik.

Jamco et al., 2023 melakukan penelitian dalam menangani multikolinearitas pada distribusi persentase produk domestik regional bruto di Provinsi Maluku periode 1999-2021 menggunakan regresi Lasso. Hasilnya, faktor-faktor yang berpengaruh terhadap PDRB di Provinsi Maluku meliputi sektor pertanian, sektor pertambangan dan penggalian, sektor industri pengolahan, dan sektor perdagangan hotel dan restoran.

Dalam penelitian ini, penulis menggunakan algoritma CCD dalam konteks metode Lasso dalam penyelesaian permasalahan multikolinearitas. Dimana, algoritma CCD bekerja dalam estimasi parameter dalam membangun model regresi dan seleksi variabel.

METODE

Penelitian ini merupakan penelitian yang menggunakan pendekatan kuantitatif menerapkan regresi Lasso dengan algoritma Cyclic Coordinate Descent. Bantuan software yang digunakan adalah Microsoft Excel bahasa pemrograman python.

HASIL PEMBAHASAN

Penelitian ini menggunakan 3 data bangkitan yang mengandung multikolinearitas dengan tingkat tinggi, sedang, dan rendah dan diperoleh dari bahasa pemrograman python dengan source code dapat dilihat pada Lampiran 1. Penggunaan data bangkitan dengan tingkat multikolinearitas berbeda ini dilakukan untuk menguji dan membandingkan performa metode regresi OLS dan Lasso secara lebih menyeluruh dalam berbagai kondisi korelasi antarvariabel. Dengan menggunakan data yang memiliki multikolinearitas tinggi, peneliti dapat melihat sejauh mana metode OLS terganggu oleh korelasi antarvariabel yang sangat kuat, serta bagaimana Lasso dapat mengatasi permasalahan tersebut melalui proses regularisasi dan seleksi variabel.

Sementara itu, data dengan multikolinearitas sedang memungkinkan evaluasi terhadap performa kedua metode dalam situasi yang tidak terlalu ekstrem, di mana sebagian variabel saling berkorelasi tetapi tidak semuanya. Terakhir, data dengan multikolinearitas rendah digunakan sebagai pembanding, untuk mengamati apakah LASSO tetap memberikan keunggulan dalam kondisi ideal atau tanpa masalah korelasi tinggi.

Pada data bangkitan digunakan ukuran sampel $n=100$ dengan jumlah variabel independen 10, yang masing-masing disimulasikan dari distribusi normal standar $N(0,1)$. Nilai variabel dependen y juga disimulasikan dari distribusi normal $N(0,1)$. Data ini dimodelkan menggunakan regresi linear berganda dan regresi Lasso. Estimasi parameter regresi linear berganda ditentukan menggunakan metode Ordinary Least Squares (OLS), sedangkan untuk estimasi parameter Lasso digunakan algoritma Cyclic Coordinate Descent (CCD) untuk mengestimasi parameter Lasso dari data bangkitan tersebut. Sebelum dilakukan pemodelan, data X distandarisasi agar memiliki rata-rata nol dan standar deviasi satu pada setiap kolom (fitur), sedangkan y dipusatkan agar memiliki rata-rata nol.

Uji Multikolinearitas

Metode Lasso merupakan metode yang dapat digunakan untuk menangani multikolinearitas. Sebelum pembentukan model regresi Lasso terlebih dahulu dilakukan pengujian multikolinearitas dengan rumus VIF (Variance Inflation Factors) pada Persamaan (2.7). Kriteria pengambilan keputusan uji multikolinieritas adalah jika nilai VIF bernilai lebih kecil dari 10, maka variabel independen dinyatakan nonmultikolinieritas. Nilai VIF untuk masing-masing variabel independen dicari menggunakan bantuan python, yang ditampilkan pada tabel berikut:

Tabel 1 Nilai VIF Untuk Setiap Variabel Independen.

Variabel	VIF		
	Data 1	Data 2	Data 3
x_1	67,34624	3,75032	1,045063
x_2	273,5058	6,356483	1,037008
x_3	177,1809	2,64826	1,191592
x_4	196,1018	11,44997	7,032298
x_5	237,5661	13,73602	6,990275
x_6	326,8448	16,73775	10,53744
x_7	198,4884	12,93256	7,296674
x_8	89,17953	5,162947	1,855671
x_9	112,2019	7,065493	4,17551
x_{10}	167,5144	8,882424	4,538901

Berdasarkan Tabel 1 diatas, pada data 1 dapat dilihat bahwa 10 variabel independen dari data tersebut memiliki nilai $VIF > 10$. Semua variabel pada data memiliki nilai VIF yang sangat besar bahkan ada yang melebihi 100, hal ini mengindikasikan bahwa semua variabel independen memiliki permasalahan multikolinearitas, dimana terdapat hubungan yang kuat antarvariabel. Untuk data 2, 4 dari 10 variabel independen memiliki mengandung multikolinearitas dimana 4 dari 10 variabelnya memiliki nilai $VIF > 10$. Sedangkan untuk data terakhir yaitu data 3, terdapat 1 variabel yang mengandung multikolinearitas.

Adanya multikolinearitas dapat menyebabkan variabel menjadi tidak signifikan dalam OLS, karena model tidak dapat menentukan pengaruh independen dari setiap variabel. Permasalahan multikolinearitas dalam data akan ditangani dengan menerapkan metode Lasso dengan algoritma CCD. Metode ini efektif dalam mengatasi multikolinearitas karena pada metode ini terjadi penyusutan koefisien regresi terhadap parameter-parameter yang memiliki korelasi tinggi (Tibshirani, 1996).

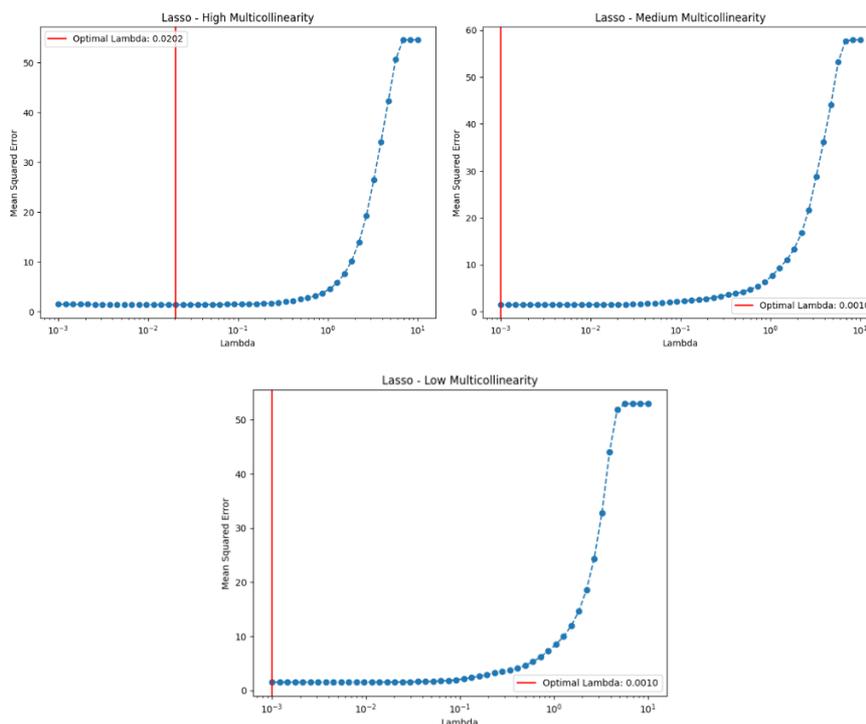
Validasi Silang

Metode Lasso mengatasi permasalahan adanya multikolinearitas dengan menyusutkan koefisien regresi menjadi nol dan sekaligus menyeleksi variabel independen. Hal penting yang diperlukan dalam pemakaian metode Lasso adalah pemilihan nilai λ yang optimal. Lambda optimal dapat diperoleh dengan melakukan validasi silang.

Validasi silang 5-fold (5-fold cross validation) adalah salah satu teknik yang umum digunakan untuk mengevaluasi kinerja model prediksi, khususnya dalam kasus regresi seperti model Lasso. Dalam metode ini, data dibagi secara acak menjadi lima bagian (disebut fold) yang kurang lebih seimbang. Proses pelatihan dan pengujian model dilakukan sebanyak lima kali, di mana pada setiap iterasi, empat fold digunakan untuk melatih model, sedangkan satu fold sisanya digunakan sebagai data pengujian. Fold yang digunakan untuk pengujian digilir sehingga setiap bagian data digunakan tepat satu kali sebagai data uji. Setelah lima iterasi selesai, nilai evaluasi model dari masing-masing iterasi dirata-ratakan untuk memperoleh estimasi kinerja model secara keseluruhan. Pendekatan ini membantu dalam mengurangi bias evaluasi dan memberikan gambaran yang lebih akurat terhadap kemampuan generalisasi model terhadap data baru.

Proses validasi silang 5-fold untuk menentukan nilai λ yang optimal dilakukan dengan bantuan software python dengan kode yang tertera pada Lampiran 3. Untuk setiap λ dihitung MSE pada Persamaan (2.14). Nilai λ dengan MSE terkecil dipilih sebagai parameter terbaik. Hasil validasi silang dapat dilihat pada gambar berikut ini.

Estimasi Parameter Least Absolute Shrinkage And Selection Operator (Lasso) Menggunakan Algoritma Cyclic Coordinate Descent



Gambar 1 Hasil Validasi Silang dengan MSE untuk Pemilihan Nilai λ Optimal

Berdasarkan Gambar 1 yaitu hasil proses validasi silang dimana proses ini membagi data menjadi beberapa bagian dan menguji performa model pada setiap bagian tersebut. Berdasarkan hasil dari validasi silang, nilai λ optimal dari proses validasi silang untuk semua data yang mengandung multikolinearitas dengan urutan tinggi, sedang, dan rendah adalah 0,0202, 0,0010, dan 0,0010. Pada gambar ditunjukkan bahwa nilai MSE minimum terletak pada titik ini. Nilai λ yang terlalu kecil menyebabkan tidak adanya regularisasi sehingga hasilnya akan mirip dengan OLS, sedangkan λ yang terlalu besar akan mengakibatkan semua koefisien mendekati nol. Oleh karena itu, metode regresi Lasso dengan algoritma CCD dengan menerapkan nilai λ yang optimal dalam koefisien modelnya.

Menentukan Model Regresi OLS

Dalam menentukan model regresi dengan metode OLS dapat dilakukan dengan penghitungan nilai koefisien β dengan penerapan rumus pada Persamaan (2.5) Namun, berdasarkan hasil dari uji multikolinearitas yang telah dilakukan disimpulkan bahwa data mengandung multikolinearitas sehingga koefisien β yang dihasilkan menjadi tidak dapat diandalkan dan sangat sensitif terhadap perubahan kecil dalam data. Hal ini terjadi karena matriks $X^T X$ mendekati singular atau memiliki nilai eigen yang sangat kecil. Untuk perhitungan koefisien β digunakan bantuan software python dengan menggunakan algoritma numerik yang dapat menangani kasus matriks mendekati singular dengan source code dapat dilihat pada Lampiran 2.

Tabel 2 Koefisien Model Regresi OLS

Variabel	Koefisien OLS		
	Data 1	Data 2	Data 3
β_0	0,040385	0,077081	0,091932
β_1	-0,55364	-0,9538	-1,03291
β_2	0,679337	0,910737	1,013138

β_3	1,284457	0,353136	0,283379
β_4	-1,35596	-1,36497	-1,43657
β_5	0,072415	-0,40331	-0,50928
β_6	0,724551	1,509045	1,595261
β_7	-3,06519	-2,56049	-2,4341
β_8	-2,02741	-0,72636	-0,57681
β_9	-0,04693	-0,58248	1,16684
β_{10}	-0,75055	0,647147	-0,92981

Berdasarkan tabel koefisien regresi OLS di atas, maka model regresi yang terbentuk untuk setiap data tersebut adalah sebagai berikut:

Model regresi dengan multikolinearitas tinggi (data 1), yaitu:

$$y = 0,040385 - 0,55364x_1 + 0,679337x_2 + 1,284457x_3 - 1,35596x_4 + 0,072415x_5 - 0,724551x_6 - 3,06519x_7 - 2,02741x_8 - 0,04693x_9 - 0,75055x_{10}$$

Model regresi dengan multikolinearitas sedang (data 2), yaitu:

$$y = 0,077081 - 0,9538x_1 + 0,910737x_2 + 0,283379x_3 - 1,43657x_4 - 0,50928x_5 + 1,509045x_6 - 2,56049x_7 - 0,72636x_8 - 0,58248x_9 + 0,647147x_{10}$$

Model regresi dengan multikolinearitas rendah (data 3), yaitu:

$$y = 0,091932 - 1,03291x_1 + 1,013138x_2 + 0,499215x_3 - 1,586212x_4 - 0,22393x_5 + 1,595261x_6 - 2,4341x_7 - 0,57681x_8 + 1,16684x_9 - 0,92981x_{10}$$

Koefisien OLS tidak mengalami regularisasi sehingga semua variabel memiliki koefisien berbeda dari nol.

Menentukan Model Regresi Lasso dengan Algoritma Cyclic Coordinate Descent

Metode Lasso tidak memiliki solusi analitik langsung seperti OLS karena fungsi objektif tidak dapat diturunkan langsung akibat penalti l_1 . Sehingga, algoritma Cyclic Coordinate Descent digunakan untuk menghitung koefisien. Lasso digunakan untuk menghasilkan model regresi yang tidak hanya memperkirakan parameter tetapi juga melakukan seleksi variabel.

Estimasi parameter Lasso dengan algoritma CCD dilakukan dengan mengikuti langkah-langkah berikut:

1. Inialisasi semua koefisien $\beta_j=0$ untuk semua variabel independen.
2. Untuk setiap variabel independen j , dihitung residual parsial berdasarkan nilai saat ini. Residual parsial adalah selisih antara nilai observasi dan prediksi tanpa kontribusi dari variabel ke- j . untuk setiap data ke- i , residual dapat dihitung sebagai:

$$r_{ij} = y_i - \sum_{p \neq j} x_{ip} \beta_p$$

3. karena β lainnya diinisialisasi 0, maka pada iterasi awal, $r_{ij}=y_i$
4. Hitung nilai korelasi parsial antara residual parsial dan variabel independen x_j :

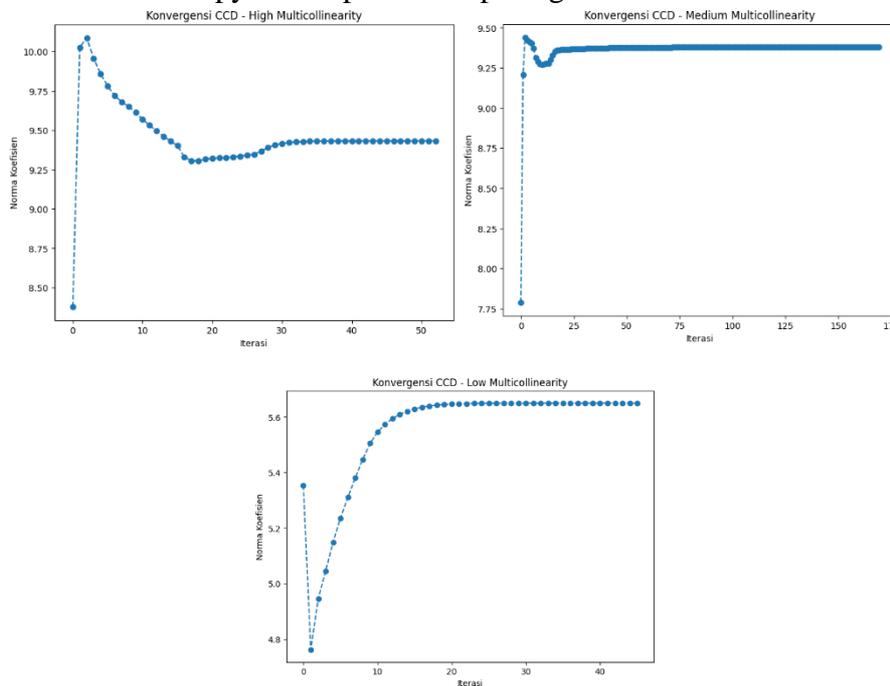
$$\rho_j = \sum_{i=1}^n x_{ij} r_{ij} = \sum_{i=1}^n x_{ij} (y_i - \sum_{p \neq j} x_{ip} \beta_p)$$

karena nilai awal $\beta_p=0$, maka perhitungan pertama $\rho_j = \sum_{i=1}^n x_{ij} y_i$

5. Lakukan update parameter β_j menggunakan operator soft-thresholding, yaitu:

$$\beta_j = \frac{S(\frac{1}{n} \sum_{i=1}^n x_{ij} y_i, \lambda)}{\frac{1}{n} \sum_{i=1}^n x_{ij}^2}$$

Proses ini akan terus berlangsung hingga koefisien konvergen, seperti yang dapat dilihat pada Lampiran 4. Proses konvergensi koefisien menggunakan algoritma CCD dengan bantuan software python dapat dilihat pada gambar berikut:



Gambar 2. Hasil Konvergensi Menggunakan Algoritma Cyclic Coordinate Descent

Dari gambar diatas dapat disimpulkan bahwa data dengan multikolinearitas tinggi konvergen setelah 580 iterasi, untuk data dengan multikolinearitas sedang menemukan solusi pada iterasi ke-180, sedangkan data dengan multikolinearitas rendah konvergen pada iterasi ke-134. Dengan didapatnya hasil dari konvergensi tersebut, maka koefisien model regresi Lasso menggunakan algoritma CCD dengan menerapkan nilai λ yang optimal:

Tabel 3 Koefisien Model Regresi Lasso

Variabel	Koefisien Lasso		
	Data 1	Data 2	Data 3
β_0	-0,18451	-0,04984	-0,45429
β_1	-0,46974	-2,10431	-0,95095
β_2	0,973278	2,738187	0,89604
β_3	0	0,684765	0,297226
β_4	-3,54919	-3,13963	-3,35247
β_5	0	-1,15651	-1,48183
β_6	0	3,096367	3,299628
β_7	-5,27353	-5,37177	-5,19827
β_8	-1,27106	-0,91717	-0,74635
β_9	0	1,214806	1,425621
β_{10}	0	-1,40823	-1,54891

Berdasarkan Tabel 3 untuk data 1, koefisien regresi Lasso mengalami penyusutan tepat pada nol terdapat pada 5 variabel, yaitu x_3 , x_5 , x_6 , x_9 dan x_{10} , variabel ini

dianggap tidak penting. Untuk koefisien variabel lainnya dkecilkan sehingga masih terdapat pada model namun lebih kecil dari variabel model OLS. Sedangkan untuk data 2 dan data terakhir tidak terjadi penyusutan koefisien tepat pada nol Penyusutan variabel pada data yang mengandung multikolinearitas tinggi ini, mengakibatkan model regresi menjadi lebih sederhana sekaligus dapat mengatasi masalah multikolinearitas karena pada variabel independen yang memiliki nilai korelasi yang tinggi akan terseleksi dari model. Sehingga model regresi Lasso menggunakan algoritma CCD dengan menerapkan nilai λ yang optimal yang terbentuk adalah:

Model regresi Lasso dengan multikolinearitas tinggi (data 1), yaitu:

$$y = -0,18451 - 0,46974x_1 + 0,973278x_2 - 3,54919x_4 - 4,16482x_6$$

Model regresi Lasso dengan multikolinearitas sedang (data 2), yaitu:

$$y = -1,57734 - 2,10431x_1 + 2,738187x_2 + 0,684765x_3 - 3,13963x_4 - 1,15651x_5 + 3,096367x_6 - 5,37177x_7 - 0,91717x_8 + 1,214806x_9 - 1,40823x_{10}$$

Model regresi Lasso dengan multikolinearitas rendah (data 3), yaitu:

$$y = -0,45429 - 0,95095x_1 + 0,89604x_2 + 0,297226x_3 - 3,35247x_4 - 1,48183x_5 + 3,299628x_6 - 5,19827x_7 - 0,74635x_8 + 1,425621x_9 - 1,54891x_{10}$$

Berdasarkan model regresi Lasso diatas, dapat dikatakan bahwa metode Lasso paling efektif digunakan pada dataset yang mengandung permasalahan multikolinearitas tinggi karena dapat mengubah 5 variabel tepat menjadi nol..

Evaluasi Keakuratan Model

Evaluasi keakuratan model regresi OLS dan Lasso dapat dilihat berdasarkan nilai MSE dan R^2 . Nilai MSE merupakan ukuran rata-rata kuadrat dari selisih antara nilai prediksi dan nilai aktual. Semakin kecil nilai MSE, semakin baik model dalam memprediksi data. Ini menunjukkan bahwa model semakin akurat dalam menangkap pola dalam data. Sedangkan R^2 menunjukkan proporsi variabilitas variabel dependen yang dapat dijelaskan oleh variabel independen dalam model. Semakin besar nilai R^2 , semakin baik model dalam menjelaskan variabilitas data. Ukuran keakuratan dari kedua model yang disajikan dalam tabel berikut didapat dengan menggunakan bantuan software python dengan source code yang terlampir pada Lampiran 5.

Tabel 4. Evaluasi Keakuratan Model OLS dan Lasso

Metode	MSE			R^2		
	Data 1	Data 2	Data 3	Data 1	Data 2	Data 3
OLS	1,491398	1,646965	1,481242	0,972476	0,970102	0,970816
Lasso	1,408573	1,813711	4,145364	0,974005	0,967075	0,918328

Berdasarkan hasil evaluasi model OLS dan Lasso terhadap tiga set data yang berbeda, dapat dilihat bahwa nilai Mean Squared Error (MSE) untuk semua model berada di atas 1. Hal ini menunjukkan bahwa rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual cukup besar atau jika model belum mampu sepenuhnya menangkap variasi dalam data.

Dari segi kinerja, pada Data 1 yang memiliki tingkat multikolinearitas tinggi, model Lasso memberikan hasil yang lebih baik dibandingkan OLS, dengan MSE lebih rendah dan nilai koefisien determinasi (R^2) lebih tinggi. Ini menunjukkan bahwa regularisasi Lasso efektif dalam menangani multikolinearitas dengan mengecilkan atau menghapus koefisien variabel yang saling berkorelasi tinggi. Sebaliknya, OLS yang tidak memiliki mekanisme penalti cenderung menghasilkan model yang overfitting dalam kondisi multikolinearitas tinggi. Untuk Data 2, performa OLS justru sedikit lebih

baik daripada Lasso dilihat dari MSE yang lebih rendah dan R^2 yang lebih tinggi. Hal ini menunjukkan bahwa penalti yang diterapkan oleh Lasso terlalu besar sehingga mengurangi efektivitas model dengan mengecilkan koefisien yang masih relevan, yang pada akhirnya menurunkan akurasi prediksi. Sedangkan pada Data 3 yang memiliki multikolinearitas rendah, kinerja Lasso menurun secara signifikan. Nilai MSE yang meningkat dan nilai R^2 menurun drastis dibandingkan OLS. Hal ini menunjukkan bahwa dalam data dengan kondisi dimana hubungan antarvariabel independen lemah, penggunaan regularisasi justru dapat merugikan karena Lasso menghapus atau mengecilkan koefisien variabel yang sebenarnya tidak bermasalah. Dalam kasus ini, OLS lebih unggul karena tidak melakukan penalti yang tidak diperlukan, sehingga mampu mempertahankan informasi dari semua variabel secara optimal. Sehingga dapat disimpulkan bahwa efektivitas Lasso sangat bergantung pada tingkat multikolinearitas dalam data. Lasso cenderung lebih unggul ketika multikolinearitas tinggi, tetapi bisa kurang optimal pada kondisi multikolinearitas sedang atau rendah.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang sudah dilaksanakan, maka diperoleh kesimpulan sebagai berikut:

1. Algoritma Cyclic Coordinate Descent (CCD) mampu mengestimasi parameter Least Absolute Shrinkage and Selection Operator (Lasso) secara efisien dan menghasilkan model regresi dengan seleksi variabel yang adaptif.
2. Efek penalti λ yang diterapkan melalui CCD pada metode Lasso menghasilkan model yang lebih stabil terhadap multikolinearitas dan lebih hemat dari sisi kompleksitas model dibandingkan dengan metode OLS yang mempertahankan semua variabel.
3. Berdasarkan model regresi yang dihasilkan, metode Lasso dengan algoritma CCD paling efektif digunakan pada data dengan multikolinearitas tinggi karena berhasil mengeliminasi 5 dari 10 variabel independen sehingga menghasilkan model yang lebih sederhana namun tetap memiliki akurasi prediksi yang tinggi. Hal ini menunjukkan bahwa Lasso mampu mengatasi permasalahan multikolinearitas dengan melakukan seleksi variabel dan penyusutan terhadap koefisien yang tidak signifikan. Sedangkan pada data dengan multikolinearitas sedang dan rendah, model regresi yang dihasilkan dengan metode OLS cenderung lebih baik karena semua variabel masih memberikan kontribusi yang relevan terhadap prediksi, sehingga penalti yang diterapkan oleh Lasso justru dapat menghilangkan informasi penting dan menurunkan performa model.

Saran

Diharapkan kepada peneliti selanjutnya untuk menggunakan data lain yang mengandung multikolinearitas. Penelitian ini juga dapat diperluas dengan membandingkan kinerja metode Lasso dengan metode lain seperti metode Ridge Regression atau metode Elastic Net.

DAFTAR PUSTAKA

- Armstrong, J. S. (2012). Illusions in Regression Analysis. SSRN Electronic Journal, July 2012. <https://doi.org/10.2139/ssrn.1969740>
- Badri, S. (2012). Metode Statistika untuk Penelitian Kuantitatif. Yogyakarta Ombak 2012.
- Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. John Wiley & Sons.
- Efendi, A., Wardani, N. W. S., Fitriani, R., & Sumarminingsih, E. (2020). Analisis Regresi: Teori dan Aplikasi dengan R. UB Press.

- Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 23* (8th ed.). Badan Penerbit UNDIP.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill Irwin.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer Science & Business Media.
- Jamco, J. C. S., Lembang, F. K., & Delsen, M. S. N. Van. (2023). Penanganan Multikolinearitas pada Regresi Linier Berganda Menggunakan Regresi Lasso (Studi Kasus: Distribusi Persentase Produk Domestik Regional Bruto di Provinsi Maluku Tahun 1999-2021). *02(02)*, 145–154.
- Montgomery, D. C., & Runger, G. C. (2010). *Applied Statistics and Probability for Engineers* (5th ed.). John Wiley & Sons.
- Ningsih, D. R., Keumala Intan, P., Yuliati, D., Prodi Matematika, M., Sains dan Teknologi, F., Sunan Ampel, U., & Prodi Matematika, D. (2023). Pemodelan Tindak Pidana Kriminalitas di Kota Tangerang Menggunakan Metode Regresi Lasso. *Estimasi: Journal of Statistics and Its Application*, *4(1)*, 2721–379. <https://doi.org/10.20956/ejsa.vi.24853>
- Rachmawati, R. N., Sari, A. C., & Yohanes. (2021). Lasso Regression for Daily Rainfall Modeling at Citeko Station, Bogor, Indonesia. *Procedia Computer Science*, *179(2020)*, 383–390. <https://doi.org/10.1016/j.procs.2021.01.020>
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58(1)*, 267–288. <https://www.jstor.org/stable/2346178>
- Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, *7*, 2541–2563