

Penerapan LASSO Least Trimmed Squares untuk Mengidentifikasi Peubah yang Berpengaruh Penyebaran Penyakit di Sulawesi Selatan

Trigarcia Maleachi Randa^{1,a)}, Georgina Maria Tinungki^{1,b)}, dan Nurtiti Sunusi^{1,c)}

¹Departemen Statistika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Hasanuddin, Makassar, Indonesia

^{a)}georgina@unhas.ac.id

Abstrak. Tuberkulosis merupakan penyakit menular kronis yang masih menjadi masalah kesehatan masyarakat di dunia. Indonesia merupakan negara dengan beban Tuberkulosis tertinggi ketiga setelah India dan Cina. Sulawesi Selatan merupakan salah satu provinsi yang menyumbang tingginya angka kasus tuberkulosis di Indonesia pada tahun 2020. Analisis regresi linier dapat diterapkan pada data tuberkulosis untuk mengetahui variabel-variabel yang mempengaruhi jumlah kasus tuberkulosis di Sulawesi Selatan. Masalah yang sering muncul dalam analisis regresi adalah masalah multikolinearitas dan pencilan pada data. Salah satu metode yang dapat digunakan untuk menyelesaikan masalah multikolinearitas dan pencilan adalah regresi LASSO LTS. Regresi LASSO LTS merupakan modifikasi dari metode regresi LASSO berdasarkan penduga LTS dari regresi kekar. Variabel pada data tuberkulosis di Sulawesi Selatan memiliki permasalahan multikolinearitas dan terdapat pencilan sehingga dalam penelitian ini digunakan pendekatan dengan metode LASSO LTS untuk mengatasinya. Hasil penelitian menunjukkan metode LASSO LTS dapat mengatasi masalah multikolinearitas dan pencilan dalam menduga parameter regresi yang dibuktikan dengan nilai koefisien determinasi tertinggi sebesar 89.41%.

Kata Kunci: multikolinearitas, pencilan, regresi LASSO, regresi LASSO LTS, tuberkulosis

Abstract. Tuberculosis is a chronic infectious disease which is still a public health problem in the world. Indonesia is a country with the third highest tuberculosis burden after India and China. South Sulawesi is one of the provinces that contributes to the high number of tuberculosis cases in Indonesia in 2020. Linear regression analysis can be applied to tuberculosis data to determine the variables that affect the number of tuberculosis cases in South Sulawesi. Problems that often arise in regression analysis are multicollinearity problems and outliers in the data. One method that can be used to solve multicollinearity and outlier problems is the LASSO LTS regression. The LASSO LTS regression is a modification of the LASSO regression method based on the LTS estimator of joint regression. The variables in the tuberculosis data in South Sulawesi have multicollinearity problems and there are outliers, so in this study an approach with the LASSO LTS method was used to overcome them. The results showed that the LASSO LTS method could overcome multicollinearity and outlier problems in estimating regression parameters as evidenced by the highest coefficient of determination of 89.41%.

Keywords: multicollinearity, outliers, LASSO regression, LASSO LTS regression, tuberculosis

PENDAHULUAN

Di seluruh dunia, penyakit tuberkulosis (TB) menjadi penyebab kematian urutan ke-13 dan pembunuh menular urutan kedua setelah Coronavirus Disease 2019 (WHO, 2020). Penanganan penyakit TB apabila perlakuan dari pengobatannya tidak terselesaikan dapat menimbulkan berbagai macam penyakit hingga kehilangan nyawa seseorang (Prisma *et al.*, 2020). Situasi tersebut yang menjadikan penyakit TB sampai saat ini sebagai isu kesehatan dunia khususnya di Indonesia. Hasil publikasi Global TB Report oleh WHO pada tahun 2020, saat ini Indonesia menduduki urutan ke-2 dari 8 negara di dunia yang menjadi penyumbang terbesar (2/3 dari total)

penderita TB setelah India. Berdasarkan peta data TB per 16 April 2021 yang dipublikasikan oleh Kemenkes RI, estimasi insiden TB tahun 2020 di Indonesia kasusnya sebesar 845.000 dan Provinsi Sulawesi Selatan termasuk dalam provinsi dengan jumlah semua kasus TB yang diobati dan dilaporkan di antara 100,000 penduduk yang tinggi (Kemenkes RI, 2021). Tepatnya angka semua kasus TB yang diobati dan dilaporkan di antara 100.000 penduduk di Sulawesi Selatan sebesar 135 per 100,000 penduduk dan capain tersebut menempatkan posisi Sulawesi Selatan urutan ke-11 di Indonesia (BPS, 2020). Posisi Sulawesi Selatan urutan ke-11 menjadi salah satu capaian terbaik, namun jika dilihat dari angka keberhasilan pengobatan kasus TB di Sulawesi Selatan ditemukan sebanyak 38.8% kabupaten/kota masih belum memenuhi dari target 80%.

Salah satu metode yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi penyebaran penyakit TB adalah analisis regresi linier berganda dengan menggunakan Metode Kuadrat Terkecil (MKT). Sebuah model regresi dapat disebut sebagai model yang baik jika model tersebut dapat memenuhi asumsi yaitu normalitas, autokorelasi, heteroskedastisitas dan multikolinieritas (Montgomery *et al.*, 2012).

Pelanggaran asumsi yang sering terjadi adalah multikolinieritas dimana terdapat hubungan linier antara beberapa atau semua variabel prediktor dalam model regresi. Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan salah satu metode yang digunakan mengatasi multikolinieritas (Tibshirani, 2011). Terjadinya pelanggaran asumsi pada data biasanya disebabkan oleh adanya pencilan. Kondisi dimana terdapat data yang nilainya berbeda jauh dengan data lainnya disebut sebagai pencilan (Dzikrullah, 2022). Metode yang dapat menganalisis data yang terdeteksi sebagai pencilan adalah regresi kekar (robust). Regresi kekar memiliki beberapa estimator salah satu estimator yang populer adalah *Least Trimmed Squares* (LTS).

Studi awal penelitian menggunakan regresi linier berganda dengan estimasi parameter MKT, diketahui pada data TB terdapat pelanggaran asumsi multikolinieritas dan terindikasi terdapat beberapa pencilan. Untuk menangani permasalahan tersebut diperlukan metode yang hasil estimasi parameternya stabil terhadap multikolinieritas dan pencilan. Metode tersebut adalah regresi LASSO LTS yang merupakan gabungan dari metode regresi LASSO dan regresi LTS (Alfons *et al.*, 2013). Penelitian ini membahas analisis data jumlah kasus TB tiap kabupaten/kota di Sulawesi Selatan tahun 2020 dengan menggunakan metode regresi LASSO LTS.

METODE PENELITIAN

Data Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi Dinas Kesehatan dan Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. Unit amatan pada penelitian ini adalah 24 kabupaten/kota di Provinsi Sulawesi Selatan pada Tahun 2020. Peubah respon pada penelitian ini adalah jumlah kasus TB. Terdapat 11 peubah bebas yang diduga berpengaruh terhadap jumlah kasus TB. Pemilihan peubah pada penelitian berdasarkan penelitian terdahulu yang telah dilakukan. Peubah respon dan peubah penjelas yang digunakan dalam penelitian ini disajikan dalam Tabel 1.

TABEL 1. Peubah respon dan penjelas yang digunakan

Peubah	Nama Peubah	Satuan
Y	Jumlah Kasus TB	Orang
X ₁	Persentase Balita 0-59 Bulan Gizi Kurang	Persen
X ₂	Persentase Keluarga Dengan Akses Terhadap Jamban Sehat	Persen

X ₃	Persentase Cakupan Desa/Kelurahan <i>Universal Child Immunisation</i> (UCI)	Persen
X ₄	Persentase Rumah Tangga yang Memiliki Akses Terhadap Sumber Air Minum Layak	Persen
X ₅	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak	Persen
X ₆	Kepadatan Penduduk	Rasio
X ₇	Persentase Jumlah Tempat Pengelolaan Makan (TPM) yang Memenuhi Syarat Kesehatan	Persen
X ₈	Persentase Pemberian ASI Eksklusif	Persen
X ₉	Persentase Pemuda yang Merokok Dalam Sebulan Terakhir	Persen
X ₁₀	Rasio Jenis Kelamin	Rasio
X ₁₁	Rasio Puskesmas per 100,000 Penduduk	Rasio

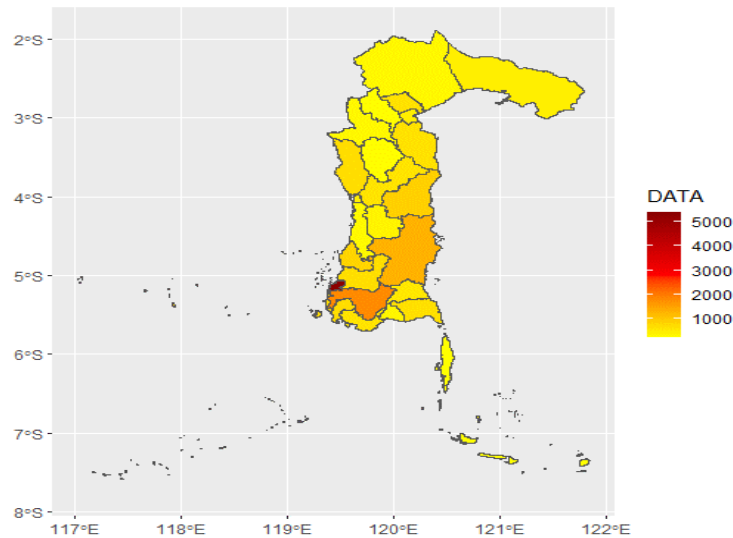
Langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Pengumpulan data jumlah kasus TB dan faktor-faktor yang mempengaruhinya di Provinsi Sulawesi Selatan Tahun 2020
2. Melakukan eksplorasi data
3. Pemodelan regresi linier dengan metode kuadrat terkecil (MKT)
4. Melakukan uji multikolinearitas dengan melihat nilai *variance inflation factor* (VIF)
5. Mendeteksi adanya pencilan dengan menggunakan metode *difference in fits* (DFFITS)
6. Pemodelan regresi linier dengan metode LASSO
7. Pemodelan regresi linier dengan metode LASSO LTS
8. Menentukan model terbaik berdasarkan nilai koefisien determinasi (R^2)
9. Melakukan interpretasi dari model terbaik yang didapatkan
10. Membuat simpulan berdasarkan masalah yang telah dibahas

HASIL DAN PEMBAHASAN

Eksplorasi Data

Sebaran jumlah kasus TB di Sulawesi Selatan pada tahun 2020 disajikan pada Gambar 1. Batasan warna dalam gambar, didapat dengan menggunakan nilai kuartil 1, median dan kuartil 3 sebagai pembatas. Jumlah kasus TB dengan nilai dibawah kuartil 1 (warna kuning) menyebar cukup merata dari utara menuju tengah Sulawesi Selatan. Jumlah kasus TB dengan nilai antara kuartil 1 dan median (warna oranye), cukup acak tersebar yang berada pada kabupaten/kota disisi barat daya dan selatan Sulawesi Selatan. Jumlah kasus TB dengan nilai lebih dari kuartil 3 (warna merah) hanya terdapat pada satu bagian di Sulawesi Selatan.



GAMBAR 1. Jumlah kasus TB di Sulawesi Selatan Tahun 2020

Salah satu syarat lain dalam analisis regresi linier yang harus dipenuhi adalah tidak terjadi multikolinearitas pada data. Multikolinearitas pada data dapat diperkirakan melalui keeratan hubungan antar peubah bebasnya. Keeratan hubungan antar peubah bebas dapat dilihat pada hasil nilai korelasi antar peubah bebasnya. Korelasi antar peubah bebas disajikan dalam bentuk plot korelasi pada Tabel 2.

TABEL 2. Plot korelasi antarpeubah bebas

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
Y	1											
X ₁	-0.27	1										
X ₂	0.06	-0.02	1									
X ₃	-0.26	0.06	0.06	1								
X ₄	-0.23	-0.21	-0.02	-0.1	1							
X ₅	0.18	-0.12	-0.01	0.02	-0.28	1						
X ₆	0.93	-0.07	0.08	-0.24	-0.2	0.13	1					
X ₇	0.13	0.34	-0.09	0.01	-0.03	-0.15	0.29	1				
X ₈	0.19	-0.16	0	-0.21	-0.12	-0.33	0.05	-0.2	1			
X ₉	0.22	-0.07	-0.12	0.16	-0.07	0.29	0.22	0.04	-0.36	1		
X ₁₀	-0.09	-0.04	-0.12	0.47	-0.02	-0.22	-0.01	0.54	-0.2	0.09	1	
X ₁₁	-0.24	0.27	0.23	-0.08	-0.04	-0.36	-0.13	0.08	0.14	-0.09	0.04	1

Pada Tabel 2 terlihat bahwa terdapat peubah bebas yang memiliki korelasi cukup besar dengan peubah bebas lainnya yang memiliki korelasi lebih dari 0.5. Peubah bebas X₇ yaitu persentase jumlah TPM yang memenuhi syarat kesehatan memiliki korelasi cukup besar dengan peubah bebas X₁₀ yaitu rasio jenis kelamin. Nilai korelasi positif menunjukkan hubungan yang searah. Artinya jika suatu peubah bertambah nilainya, maka peubah yang memiliki korelasi positif dengan peubah tersebut juga akan bertambah nilainya. Nilai korelasi negatif menunjukkan hubungan yang berlawanan arah. Artinya jika suatu peubah berkurang nilainya, maka peubah

yang memiliki korelasi negatif dengan peubah tersebut akan bertambah nilainya dan sebaliknya. Nilai korelasi yang cukup besar antar peubah bebas dapat mengindikasikan terjadinya multikolinearitas pada data yang digunakan.

Regresi Linier Berganda dengan MKT

Analisis data menggunakan analisis regresi linier berganda dengan penduga MKT menghasilkan model dengan nilai F-hitung sebesar 15.35 dan nilai-p sebesar 0.000027. Dengan demikian, model yang dihasilkan signifikan pada taraf nyata 5% atau terdapat pengaruh secara bersama-sama (simultan) dari peubah bebas terhadap peubah respon. Hasil analisis ragam dengan menggunakan MKT disajikan pada Tabel 3

TABEL 3. Analisis ragam yang dihasilkan MKT

Sumber Keragaman	db	Jumlah Kuadrat	Kuadrat Tengah	F-hitung	Nilai-p
Regresi	11	23699787	2154526	15.35	0.000027
Sisaan	12	1684202	140350		
Total	23	25383989	2294876		

Pendugaan koefisien regresi secara parsial ditampilkan pada Tabel 4. Dari 11 peubah bebas yang digunakan, terlihat bahwa terdapat peubah yang signifikan atau berpengaruh nyata terhadap model, yaitu peubah bebas X_6 . Hal tersebut bisa dilihat dari nilai-p peubah selain X_6 tidak ada yang kurang dari taraf nyata 5%.

TABEL 4. Koefisien regresi yang dihasilkan MKT

Peubah	Dugaan Koefisien	t-hitung	Nilai-p
Intersep	2105.66	0.45	0.66
X_1	-54.85	-1.88	0.09
X_2	0.61	0.04	0.97
X_3	-0.67	-0.16	0.88
X_4	-6.44	-0.97	0.35
X_5	0.15	0.01	0.99
X_6	0.56	9.27	0.0000008
X_7	0.53	0.12	0.90
X_8	10.28	1.34	0.20
X_9	21.90	0.63	0.54
X_{10}	-17.84	-0.51	0.62
X_{11}	-54.40	-1.07	0.31

Mendeteksi Multikolinearitas

Pemeriksaan multikolinearitas dapat dilakukan dengan melihat nilai VIF dari masing-masing peubah bebas. Peubah dengan nilai VIF > 10 mengindikasikan adanya multikolinearitas sehingga

perlu dilakukan penanganan (Ihsan *et al.*, 2020). Terdapat peubah dengan nilai VIF lebih besar dari 10 yang artinya memiliki masalah multikolinearitas serius. Peubah dengan nilai VIF yang lebih besar dari sepuluh yaitu peubah X_6 . Tabel 5 menunjukkan Nilai VIF untuk setiap peubah bebas.

TABEL 5. Nilai VIF untuk setiap peubah bebas

Peubah	VIF
X_1	2.308453
X_2	1.210979
X_3	1.800889
X_4	1.552407
X_5	1.813868
X_6	12.562985
X_7	2.424406
X_8	1.934635
X_9	1.447574
X_{10}	2.853879
X_{11}	1.554821

Mendeteksi Pencilan

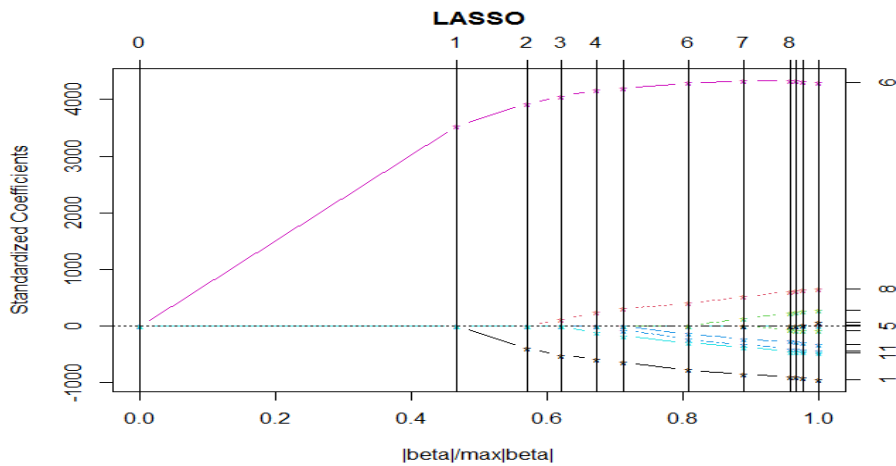
Untuk mendeteksi pencilan dapat menggunakan metode DFFITS untuk mengidentifikasi pencilan di peubah respon. Karena jumlah data ($n = 24$) termasuk gugus data yang berukuran kecil, maka kriteria disebut pencilan jika nilai $|DFFITS_i| > 1$ (Deria *et al.*, 2019). Data yang mempunyai nilai $|DFFITS_i| > 1$ adalah data ke-1, ke-2, ke-8, ke-13, ke-20 dan ke-22, sehingga data tersebut merupakan pencilan. Nilai DFFITS dapat dilihat pada Tabel 6.

TABEL 6. Nilai DFFITS

Pengamatan	DFFITS _{<i>i</i>}	$ DFFITS_i $
1	2.168284	2.168284
2	-21.931580	21.931580
8	-1.041095	1.041095
13	1.017621	1.017621
20	-3.846738	3.846738
22	8.475870	8.475870

Regresi Linier Berganda dengan LASSO

Metode LASSO juga dapat digunakan untuk mengatasi masalah multikolinearitas yang terjadi pada analisis regresi linier berganda. Penduga koefisien LASSO diperoleh melalui komputasi dengan memodifikasi algoritma *Least Angle Regression* (LAR) sehingga menghasilkan algoritma yang lebih efisien dibandingkan pemrograman kuadratik. Plot tahapan pendugaan koefisien LASSO menggunakan algoritma LAR dapat dilihat pada Gambar 2.

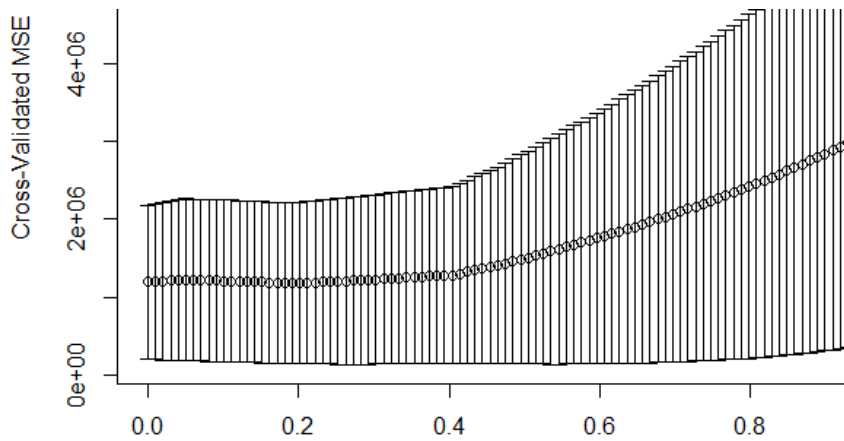


- **GAMBAR 2.** Plot pergerakan penduga koefisien regresi LASSO
- Pada metode LASSO, tahap inisiasi diawali dengan menetapkan semua koefisien dengan angka nol sehingga pada tahap ini belum ada peubah bebas yang masuk ke dalam model. Tabel 7 menampilkan beberapa informasi penting seperti banyaknya tahapan komputasi LASSO untuk penelitian ini, nilai parameter *tuning* baku (λ) yang digunakan, tahapan masuknya peubah bebas, dan nilai koefisien untuk setiap peubah bebas yang masuk ke dalam model. Berikut adalah tahapan peubah bebas yang masuk ke model LASSO:

TABEL 7. Tahapan peubah bebas yang masuk ke dalam model LASSO

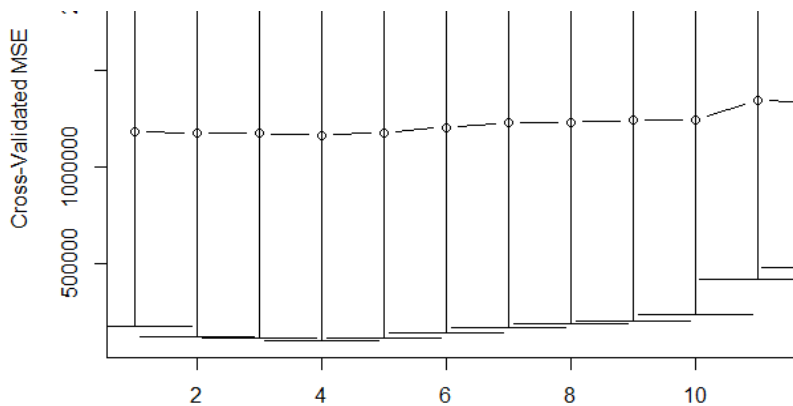
TAHAP	Peubah Bebas Yang Masuk Ke Dalam Model										
1	X ₆										
2	X ₆	X ₁									
3	X ₆	X ₁	X ₈								
4	X ₆	X ₁	X ₈	X ₁₁							
5	X ₆	X ₁	X ₈	X ₁₁	X ₄						
6	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀					
7	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀	X ₉				
8	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀	X ₉	X ₃			
9	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀	X ₉	X ₃	X ₅		
10	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀	X ₉	X ₃	X ₅	X ₇	
11	X ₆	X ₁	X ₈	X ₁₁	X ₄	X ₁₀	X ₉	X ₃	X ₅	X ₇	X ₂

Pemilihan model terbaik dalam metode LASSO dilakukan dengan menggunakan kriteria validasi silang, yaitu dengan menggunakan mode *fraction* dan mode *step*. Pada mode *fraction*, nilai validasi silang dihitung berdasarkan $\sum |\hat{\beta}_j| / \max \sum |\hat{\beta}_j|$. Pada Gambar 3 nilai sisaan validasi silang minimum terjadi saat nilai parameter tuning berada diantara nilai 0,1 – 0,4.



• **GAMBAR 3.** Plot validasi silang LASSO dengan menggunakan mode *fraction*

Mode step menghitung nilai validasi silang pada setiap tahapan dimana satu peubah masuk dalam model. Pada Gambar 4 terlihat bahwa plot validasi silang menunjukkan nilai sisaan validasi silang minimum pada saat tahap ke-4.

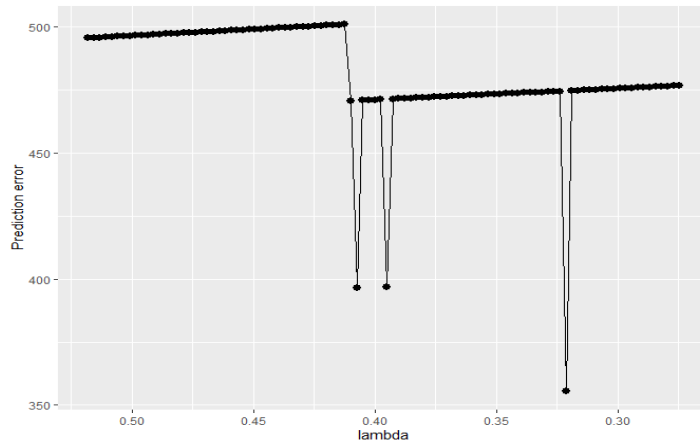


GAMBAR 4. Plot validasi silang LASSO dengan menggunakan mode *step*

Berdasarkan kedua metode validasi silang tersebut, maka model terbaik LASSO yang terpilih pada data ini adalah model pada tahap ke-4. Terdapat tiga peubah bebas yang masuk ke dalam model pada tahap ke-4, yaitu X1, X6, dan X8.

Regresi Linier Berganda dengan LASSO LTS

Untuk mengatasi masalah masalah multikolinearitas dan pencilan secara bersamaan, maka regresi LASSO digabungkan dengan salah satu penduga regresi kekar yang paling populer yaitu penduga LTS untuk membentuk penaksir LASSO yang kekar yaitu penduga LASSO LTS. Metode LASSO LTS merupakan metode LASSO yang menggunakan data terpankaskan yang diperoleh dengan algoritma LTS. Pemilihan model terbaik dalam metode LASSO dilakukan dengan menggunakan kriteria validasi silang, yaitu dengan menggunakan mode *fraction*. Pada Gambar 5 terlihat bahwa *fraction* nilai λ sebesar 0.3213741 merupakan nilai prediksi sisaan validasi silang minimum.



GAMBAR 5. Plot validasi silang LASSO dengan menggunakan mode *fraction*

Perbandingan Model

Adanya parameter *tuning* baku (λ) pada LASSO dan LASSO LTS menyebabkan dugaan koefisien regresi kedua metode tersebut menyusut. Perbedaan nilai λ antara LASSO dan LASSO LTS menghasilkan dugaan koefisien LASSO cenderung menyusut menjadi nol dibandingkan dugaan koefisien LASSO LTS. Perbandingan koefisien regresi hasil analisis dengan menggunakan MKT, LASSO, dan LASSO LTS dapat terlihat pada Tabel 8.

TABEL 8. Koefisien regresi hasil MKT, LASSO, dan LASSO LTS

Peubah	MKT	LASSO	LASSO LTS
Intersep	2105.6598	548.2878	-668.7902
X ₁	-54.8483	-29.9735	-16.4864
X ₂	0.6102	0.0000	-10.8057
X ₃	-0.6728	0.0000	-0.2479
X ₄	-6.4399	0.0000	-5.2996
X ₅	0.1487	0.0000	5.9320
X ₆	0.5551	0.5237	0.6316
X ₇	0.5330	0.0000	-2.7721
X ₈	10.2813	1.9533	5.3632
X ₉	21.8998	0.0000	-7.1354
X ₁₀	-17.8379	0.0000	16.9548
X ₁₁	-54.3981	0.0000	13.1968

Ukuran Kebaikan Model

Hasil analisis LASSO dan LASSO LTS memiliki perbedaan. Salah satunya adalah banyak peubah yang digunakan dalam model LASSO jauh lebih sedikit dibandingkan LASSO LTS. Hal ini akan berpengaruh pada nilai koefisien determinasi (R^2), dimana nilai R^2 dari LASSO LTS akan lebih besar dibandingkan dengan LASSO. Nilai R^2 adalah seberapa besar keragaman dari peubah respon dapat dijelaskan oleh peubah penjelas, sehingga semakin besar ukuran ini, akan

semakin baik. *Adjusted R²* (R_{adj}^2) merupakan nilai R^2 dengan mempertimbangkan banyak peubah penjelas yang masuk ke dalam model. Terlihat dari Tabel 9, bahwa model LASSO LTS lebih baik di nilai R^2 dan R_{adj}^2 yang lebih tinggi. Ukuran kebaikan kedua model ini disajikan dalam Tabel 9.

TABEL 8. Ukuran kebaikan LASSO dan LASSO LTS

Ukuran	LASSO	LASSO LTS
R_{adj}^2	77.38%	79.70%
R^2	88.20%	89.41%

KESIMPULAN

Hasil perhitungan R^2 dan R_{adj}^2 model regresi menggunakan metode LASSO LTS lebih besar yaitu secara berturut-turut sebesar 89.41% dan 79.70% dibandingkan dengan R^2 dan R_{adj}^2 model regresi menggunakan metode yaitu secara berturut-turut sebesar 88.20% dan 79.70% sehingga model regresi LASSO LTS dikatakan lebih baik dibandingkan dengan model regresi menggunakan LASSO dalam mengatasi masalah multikolinearitas dan pencilan.

Model regresi untuk faktor-faktor yang mempengaruhi jumlah kasus TB di Sulawesi Selatan berdasarkan LASSO LTS yaitu

$$\hat{y} = -668.7902 - 16.4864 X_1 - 10.8057X_2 - 0.2479X_3 - 5.2996X_4 + 5.9320X_5 + 0.6316X_6 - 2.7721X_7 + 5.3632X_8 - 7.1354X_9 + 16.9548X_{10} + 13.1968X_{11}$$

DAFTAR PUSTAKA

- Alfons, A., Croux, C., & Gelper, S. (2013). Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*, 226-248.
- BPS. (2020). *Sulawesi Selatan Dalam Angka 2020*. Makassar: Badan Pusat Statistik Provinsi Sulawesi Selatan.
- Deria, A. D., Hoyyi, A., & Mustafid, M. (2019). Regresi Robust Estimasi-M Dengan Pembobot Andrew, Pembobot Ramsay Dan Pembobot Welsch Menggunakan Software R. *Jurnal Gaussian*, 377-388.
- Dzikrullah, A. A. (2022). Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet Dengan Metode Centroid Linkage. *JMathCos (Journal of Mathematics, Computations, and Statistics)*, 48-57.
- Ihsan, H., Sanusi, W., & Ulfadwiyanti, R. (2020). Model Generalized Poisson Regression (GPR) dan Penerapannya pada Angka Pengangguran bagi Penduduk Usia Kerja di Provinsi Sulawesi Selatan. *JMathCos (Journal of Mathematics, Computations, and Statistics)*, 109-117.
- Kemenkes RI. (2021). *Profil Kesehatan Indonesia 2020 (Indonesian Health Profile 2020)*. Jakarta: Kementerian Kesehatan Republik Indonesia.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. USA: John Wiley & Sons.
- Prisma, R. H., Pramesti, W., & Gangga, A. (2020). Penggunaan Analisis Klaster K-Means Dalam Pemodelan Regresi Spasial Pada Kasus Tuberkulosis Di Jawa Timur Tahun 2007. *Indonesian Journal of Statistics and Its Applications*, 164-178.

- Tibshirani, R. (2011). Regression Shrinkage And Selection Via The Lasso: A Retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* , 273-282.
- WHO. (2020). *Global tuberculosis report 2020*. Geneva: World Health Organization.