

ANALISIS KEMISKINAN MENGGUNAKAN REGRESI *RIDGE*, *LASSO* DAN *ELASTIC-NET* PROVINSI SUMATERA UTARA TAHUN 2025

Tengku Mashitah Crisanty^{1*}, Pismia Sylvi²

¹²Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

*Penulis korespondensi: tengkumashitahcs@gmail.com

ABSTRAK

Salah satu prioritas utama dalam tujuan Pembangunan Berkelanjutan (SDGs) dan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2025-2029 yaitu pengentasan kemiskinan. Pada tahun 2025, angka kemiskinan di Provinsi Sumatera Utara tercatat sebesar 7,24 persen. Dalam menganalisis kemiskinan yang memiliki banyak faktor penyebab dan sering kali terjadi pelanggaran asumsi klasik multikolinearitas sehingga menyebabkan model regresi linear (OLS) tidak lagi bersifat *Best Linear Unbiased Estimator* (BLUE). Tujuan dari penelitian ini yaitu membandingkan metode yang terbaik dari metode OLS dan regresi regularisasi, yaitu OLS, *Ridge*, *LASSO*, dan *Elastic Net* serta variabel yang memengaruhi persentase kemiskinan di 33 kabupaten/kota Provinsi Sumatera Utara tahun 2025. Dalam penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) tahun 2025 dengan 18 variabel independen. Analisis dilakukan menggunakan perangkat lunak RStudio dengan dan menentukan parameter penalti λ yang optimal. Hasil penelitian menunjukkan bahwa *LASSO* merupakan model terbaik dengan nilai *Mean Squared Error* (MSE) terkecil sebesar 5,610, dibandingkan *Elastic Net* (5,706) dan *Ridge* (6,502). *LASSO* berhasil mereduksi 12 variabel dan menyisakan 6 variabel signifikan, yaitu rata-rata lama sekolah (X3), pengeluaran perkapita (X4), angka partisipasi sekolah (X11), pertumbuhan ekonomi (X12), persentase rumah tangga dengan sanitasi tidak layak (X14), dan rasio jenis kelamin (X18). Variabel rata-rata lama sekolah (X3) dan pertumbuhan ekonomi (X12) menjadi faktor paling dominan dilihat dari pengaruh negatif yang besar terhadap kemiskinan.

Kata kunci: Kemiskinan, Regularisasi, *Ridge*, *LASSO*, *Elastic Net*.

1 PENDAHULUAN

Salah satu target utama dalam Sustainable Development Goals (SDGs) yaitu pengentasan kemiskinan. Pengentasan kemiskinan secara berkelanjutan juga dijabarkan dalam salah satu sasaran pokok Rencana Pembangunan Jangka Menengah Nasional (RPJMN) dengan target penurunan kemiskinan pada tahun 2029 berada pada rentang 4,5 hingga 5 persen (Bappenas, 2024). Meskipun angka kemiskinan yang dikeluarkan BPS mengalami trend menurun tiap tahunnya namun terdapat beberapa wilayah yang memiliki angka kemiskinan diatas angka kemiskinan nasional. Sumatera Utara merupakan salah satu provinsi yang ada di pulau Sumatera yang memiliki persentase kemiskinan sebesar 7,24 persen pada September 2025 (BPS, 2026) dengan keragaman sumber daya alam yang melimpah, mulai dari sektor pertanian seperti kelapa sawit dan karet hingga produk pertambangan seperti emas dan nikel menjadikan Sumatera Utara menjadi salah satu pusat sumber daya alam yang potensial (Purba dkk., 2024). Potensi sumber daya alam yang dimiliki Sumatera Utara harus dimanfaatkan secara maksimal sehingga angka kemiskinan turun dan mencapai target nasional di tahun 2029.

Dalam menganalisis kemiskinan yang memiliki variabel multidimensi, indikator seperti pendidikan, kesehatan, dan infrastruktur dasar berperan dalam menentukan persentase kemiskinan (Widayatsari, A., Harlen, H., Damaiyanti, D., & Misdawita, M., 2025; Zam, 2025)). Selain itu, variabel kependudukan seperti tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, upah minimum, indeks pembangunan manusia, jumlah penduduk serta indikator ekonomi seperti pertumbuhan ekonomi juga memengaruhi kemiskinan (Ashari &

Athoillah, 2023). Kemudian, dalam menganalisis kemiskinan, diperlukan regresi regularisasi untuk mengatasi multikolinearitas yang tinggi. Metode terpilih yaitu metode LASSO dan variabel yang memengaruhi kemiskinan adalah angka harapan hidup, akses listrik PLN, dan proporsi tenaga kerja sektor pertanian (Jannah dkk., 2025). Beberapa penelitian sebelumnya yang telah menggunakan regresi regularisasi dalam menganalisis kemiskinan menunjukkan bahwa pendekatan regresi regularisasi mampu menyoroti faktor utama seperti pendidikan, pendapatan dan ketenagakerjaan (John Lu, 2010).

Kemiskinan merupakan masalah kompleks karena dipengaruhi oleh berbagai macam faktor sehingga dapat dikatakan bahwa kemiskinan merupakan fenomena yang bersifat multidimensi dengan karakteristik high dimension data (Kusuma & Wulansari, 2020a). Salah satu teknik statistika yang umum digunakan dalam melihat pengaruh variabel independen terhadap variabel dependen yaitu analisis regresi linear (Susanti & Saumi, 2022). Namun dalam analisis regresi linear, terdapat asumsi-asumsi yang harus dipenuhi sehingga model yang terbentuk dapat dipercaya dan dapat diandalkan sehingga bersifat Best Linear Unbiased Estimator (BLUE). Salah satu permasalahan yang sering muncul dalam analisis data multidimensi seperti kemiskinan yaitu adanya pelanggaran asumsi multikolinearitas dimana terdapat hubungan yang tinggi antar variabel dalam satu model dan menyebabkan model tersebut tidak bersifat BLUE sehingga koefisien regresi yang dihasilkan kurang akurat. Selain itu, perlunya seleksi variabel dalam menganalisis kemiskinan agar menghasilkan analisis ringkas. Dalam mengatasi kasus multikolinearitas dan perlunya penyeleksian variabel independent yaitu dengan beberapa metode yaitu regresi *Ridge*, *Least Absolute Shrinkage and Selection Operator* (LASSO) dan *regularisasi Elastic-net* (Kusuma & Wulansari, 2020a).

Berdasarkan latar belakang diatas, penelitian ini bertujuan untuk melakukan pendugaan persentase kemiskinan pada level kabupaten kota di Provinsi Sumatera Utara tahun 2025 dan mengetahui model terbaik dari metode OLS, Regresi *Ridge*, LASSO, dan *Elastic-net* dengan membandingkan hasil Mean Square Error (MSE) dan Root Mean Square Error (RMSE) terendah serta nilai tertinggi dari R-Squared (R²). Selanjutnya model terpilih beserta hasil seleksi variabel yang dihasilkan akan digunakan untuk mengetahui faktor-faktor yang memengaruhi kemiskinan di provinsi Sumatera Utara Tahun 2025.

2 METODE

2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dari website dan publikasi Badan Pusat Statistik. *Software* yang digunakan yaitu RStudio versi 4.5.3 (2026-03-11). Variabel dependen yang digunakan yaitu persentase kemiskinan Provinsi Sumatera Utara (Y) dengan data berjumlah 33 yang merupakan kabupaten kota di Provinsi Sumatera Utara.

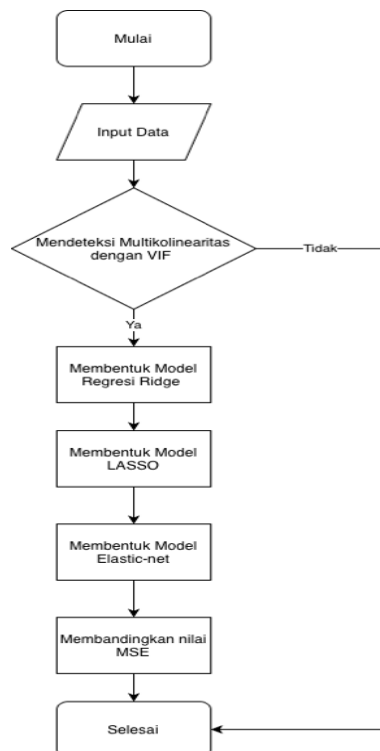
Tabel 1 . Variabel Penelitian

Variabel	Nama Variabel	Skala Pengukuran	Keterangan
Y	Miskin (%)	Rasio	Persentase Penduduk Miskin
X ₁	Umur Harapan Hidup	Rasio	Umur Harapan Hidup
X ₂	Harapan Lama Sekolah	Rasio	Harapan Lama Sekolah
X ₃	Rata-rata lama sekolah	Rasio	Rata-rata lama pendidikan/sekolah
X ₄	Pengeluaran per Kapita	Rasio	Pengeluaran per kapita (dalam Rupiah)
X ₅	Jumlah Angkatan Kerja	Rasio	Total penduduk usia kerja yang aktif secara ekonomi
X ₆	TPAK	Rasio	Tingkat Partisipasi Angkatan Kerja

X ₇	Jumlah Penduduk	Rasio	Proyeksi jumlah penduduk tahun 2025
X ₈	UMK	Rasio	Upah Minimum Kabupaten/Kota (dalam Rupiah)
X ₉	TPT	Rasio	Tingkat Pengangguran Terbuka
X ₁₀	KRT Pertanian	Rasio	Persentase Kepala Rumah Tangga sektor pertanian
X ₁₁	APS	Rasio	Angka Partisipasi Sekolah
X ₁₂	PE	Rasio	Pertumbuhan Ekonomi
X ₁₃	BB Masak	Rasio	Proporsi rumah tangga dengan bahan bakar masak tidak layak
X ₁₄	Sanitasi Tak Layak	Rasio	Persentase rumah tangga dengan sanitasi tidak layak
X ₁₅	Lantai Tak Layak	Rasio	Persentase rumah tangga dengan lantai tidak layak
X ₁₆	Keluhan Kesehatan	Rasio	Persentase penduduk dengan keluhan kesehatan
X ₁₇	Laju Pertumbuhan Penduduk	Rasio	Laju pertumbuhan penduduk
X ₁₈	Rasio Jenis Kelamin	Rasio	Rasio jenis kelamin (Sex Ratio)

2.2 Metode Analisis

Penelitian ini termasuk penelitian kuantitatif yang bertujuan membandingkan kinerja empat metode yaitu *ordinary least square* (OLS), regresi *Ridge*, LASSO dan *Elastic-net* dalam memodelkan tingkat kemiskinan Provinsi Sumatera Utara tahun 2025. Pendekatan keempat metode digunakan untuk mengatasi masalah multikolinearitas, perlunya seleksi variabel dan mengetahui variabel-variabel independen mana yang paling berpengaruh terhadap kemiskinan Provinsi Sumatera Utara.



Gambar 1. Langkah Analisis Data

2.3 Regresi Linear Berganda

Analisis regresi linear berganda merupakan salah satu metode statistika yang memprediksi nilai suatu variabel dependen (Y) berdasarkan dua atau lebih variabel independen (X) didalam satu persamaan tunggal (James dkk., 2023). Model *Ordinary Least Square* (OLS) merupakan pendekatan dasar dalam analisis regresi. Berikut rumus regresi linear berganda lebih dari satu variabel :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (1)$$

Keterangan :

Y_i = Variabel dependen

X_i = Variabel independen

β_0 = Intersep

β_i = Koefisien regresi

ε = residual

Estimasi parameter yang digunakan dalam analisis regresi adalah metode kuadrat terkecil (*Ordinary least Square*) yang mengestimasi parameter β dengan meminimumkan jumlah dari kuadrat eror persamaan regresinya. Estimasi parameter diperoleh dengan meminimalkan jumlah kuadrat perbedaan antara nilai yang diamati dan nilai yang diprediksi. Hal ini menghasilkan estimator yang tidak bias, efisien, dan konsisten jika asumsi klasik terpenuhi. Namun, OLS sensitif terhadap multikolinearitas (Jannah dkk., 2025). Estimasi parameter β menggunakan OLS adalah :

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2)$$

Asumsi yang harus dipenuhi dalam menggunakan analisis regresi adalah tidak adanya pelanggaran asumsi seperti multikolinearitas, normalitas dan heteroskedasitas (Putri & Suliadi, 2023). Uji multikolinearitas digunakan untuk mengetahui kondisi ada tidaknya hubungan linear antar variabel independen. Variasi parameter akan menjadi besar dan dapat menurunkan akurasi dalam mengestimasi jika terdapat pelanggaran asumsi multikolinearitas (Kutner, 2005). Salah satu cara untuk mengetahui apakah model mengalami multikolinearitas yaitu dengan *Variance Inflation Factors* (VIF) dan *Tolerance* (Draper & Smith, 1998).

$$VIF = \frac{1}{1-R_j^2} \quad (3)$$

Dimana $j = 1, 2, 3, \dots, p$ dan R_j^2 adalah nilai koefisien determinasi ketika X_j diregresikan dengan peubah prediktor lainnya. Apabila nilai $VIF > 10$ dan *tolerance* $< 0,10$ maka terjadi multikolinearitas. Ketika nilai VIF besar maka multikolinearitas dalam analisis tersebut juga semakin tinggi. Penyusutan variabel dengan metode regularisasi merupakan salah satu Teknik dalam mengatasi multikolinearitas (James dkk., 2013a).

2.4 Regresi Ridge (L2 Regularization)

Regresi *Ridge* merupakan salah satu metode dalam menstabilkan nilai koefisien akibat multikolinearitas. Metode ini dikenalkan oleh A. E. Hoerl pada tahun 1962. Regresi *Ridge* ini menghasilkan estimasi bias namun memiliki varian yang lebih kecil dibandingkan varians penduga OLS nya (Kutner, 2005). Hal ini dapat diketahui dengan MSE penduga *Ridge* lebih kecil dibandingkan MSE penduga OLS (Ohyver, 2011). Regresi *Ridge* mirip dengan OLS karena bertujuan untuk meminimalkan jumlah kuadrat sisaan (*Sum of Squares Error* atau SSE).

Bedanya, regresi *Ridge* menambahkan unsur penalti penyusutan (*shrinkage penalty*). Persamaan yang terbentuk adalah sebagai berikut:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (3)$$

Dimana $\lambda \geq 0$ adalah parameter penyusutan dan $\sum_{j=1}^p \beta_j^2$ adalah penalty penyusutan. Berbeda dengan OLS yang hanya menghasilkan satu set dugaan koefisien, regresi *Ridge* mampu menghasilkan berbagai variasi nilai dugaan yang bergantung pada besar kecilnya nilai λ . Oleh karena itu, penentuan nilai λ yang paling optimal biasanya dilakukan melalui metode validasi silang (*cross-validation*).

2.5 LASSO (*Least Absolute Shrinkage and Selection Operator*)

LASSO diperkenalkan oleh Tibshirani tahun 1996. Dalam metode ini, penduga koefisien dari regresi diturunkan atau disusutkan menuju nol diiringi peningkatan nilai λ . Perbedaan metode LASSO dengan regresi *Ridge* yaitu dalam regresi LASSO, adanya seleksi peubah atau variabel secara otomatis. Regresi *Ridge* merupakan proses kontinu yang mengecilkan koefisien sehingga lebih model yang dianalisis menjadi lebih stabil. Namun tidak menetapkan koefisien apapun menjadi nol sehingga tidak memberikan kemudahan dalam menginterpretasikan modelnya. Sedangkan LASSO dapat menyusutkan beberapa koefisien menjadi sama dengan nol. Sehingga LASSO dapat mempertahankan fitur dengan baik dan pemilihan subset dan regresi *Ridge* (Kusuma & Wulansari, 2020b).

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (4)$$

Regresi *Ridge* dan LASSO hampir sama, hanya penalti saja yang berbeda. Penalty penyusutan regresi *Ridge* adalah $\lambda \sum_{j=1}^p \beta_j^2$ sedangkan LASSO $\sum_{j=1}^p |\beta_j|$. Oleh karena itu, model LASSO ini dapat menyusutkan koefisien menjadi nol. Semakin besar nilai λ maka semakin banyak koefisien yang ditekan menjadi bernilai nol, sehingga hanya variabel yang paling signifikan yang dipertahankan dalam model (Panjaitan dkk., t.t.).

2.6 Elastic-net

Elastic-net pertama kali diperkenalkan oleh Zou dan Hastie pada tahun 2005. *Elastic-net* sama seperti regresi LASSO yang secara simultan melakukan seleksi variabel dan penyusutan koefisien penduga serta menangani kekurangan LASSO. *Elastic-net* mengkombinasikan penalti penyusutan dari regresi *Ridge* dan LASSO, sehingga persamaannya menjadi :

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (5)$$

Dengan $\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}, 0 \leq \alpha \leq 1$

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda [(1 - \alpha) \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j|] \quad (6)$$

Dimana α adalah parameter gabungan antara Regresi *Ridge* ($\alpha = 0$) dan LASSO ($\alpha = 1$). Jika $\alpha = 0$, maka penalti penyusutan pada *Elastic-net* menjadi penalti penyusutan regresi *Ridge*. Sedangkan ketika $\alpha = 1$, maka penalti penyusutan pada *Elastic-net* menjadi penalti penyusutan LASSO.

2.7 Validasi Silang (Cross Validation)

Metode yang umum dipakai untuk menduga kesalahan prediksi yaitu validasi silang. Dalam kondisi ideal di mana jumlah data mencukupi, peneliti dapat memisahkan suatu gugus data khusus sebagai data validasi untuk mengukur tingkat ketepatan model. Namun, keterbatasan ukuran sampel seringkali menjadi kendala untuk melakukan validasi secara langsung. Untuk mengatasi hal tersebut, validasi silang dilakukan dengan membagi data yang tersedia ke dalam dua kelompok utama, yakni data pemodelan dan data validasi (Hastie dkk., 2009). Pada metode validasi silang lipat (*fold cross-validation*), seluruh data observasi dibagi secara acak ke dalam kelompok dengan ukuran yang relatif sama. Satu kelompok ditetapkan sebagai data validasi, sedangkan kelompok lainnya digunakan untuk membentuk model. Proses ini diulang sebanyak k kali sehingga setiap kelompok pernah digunakan sebagai data validasi secara bergantian. Maka, penentuan parameter optimal yaitu pemilihan λ dan α menggunakan metode *fold cross-validation* dengan kriteria minimasi MSE. Sehingga *Cross Validation Error* (CVE) lipat- k diperoleh dengan cara:

$$CVE(k) = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (7)$$

Pemilihan model terbaik selain menggunakan MSE, juga menggunakan nilai terendah dari *Root Mean Squared Error* (RMSE) dan nilai tertinggi dari koefisien determinasi (R^2) untuk melihat akurasi yang telah diprediksi model dan kemampuan model menjelaskan variasi tingkat kemiskinan (James dkk., 2013b).

3 HASIL DAN PEMBAHASAN

Sebelum masuk ke regresi regularisasi, yang dilakukan pertama kali yaitu menganalisis model regresi linier dengan menggunakan metode estimasi OLS untuk mengetahui variabel-variabel yang memengaruhi kemiskinan dengan 18 variabel independen dan persentase rumah tangga miskin kabupaten/kota di Provinsi Sumatera Utara sebagai variabel dependen. Hasil menunjukkan bahwa model secara simultan memiliki pengaruh yang signifikan terhadap variabel dependen. Pada penilaian kebaikan model menggunakan R^2 dan R^2_{adj} dengan nilai masing-masing sebesar 94,52 persen dan 87,48 persen. Artinya model yang terbentuk sudah cukup baik dalam mengestimasi variabel independen. Dalam uji klasik model, Uji normalitas Kolmogorov-Smirnov menunjukkan bahwa data normal, tidak terdapat heteroskedastisitas dengan uji *Breusch-Pagan*, dan tidak terjadi autokorelasi dengan uji *Durbin-Watson*. Namun, terdapat pelanggaran uji klasik model yaitu multikolinearitas.

Tabel 2. Nilai VIF tiap variabel prediktor

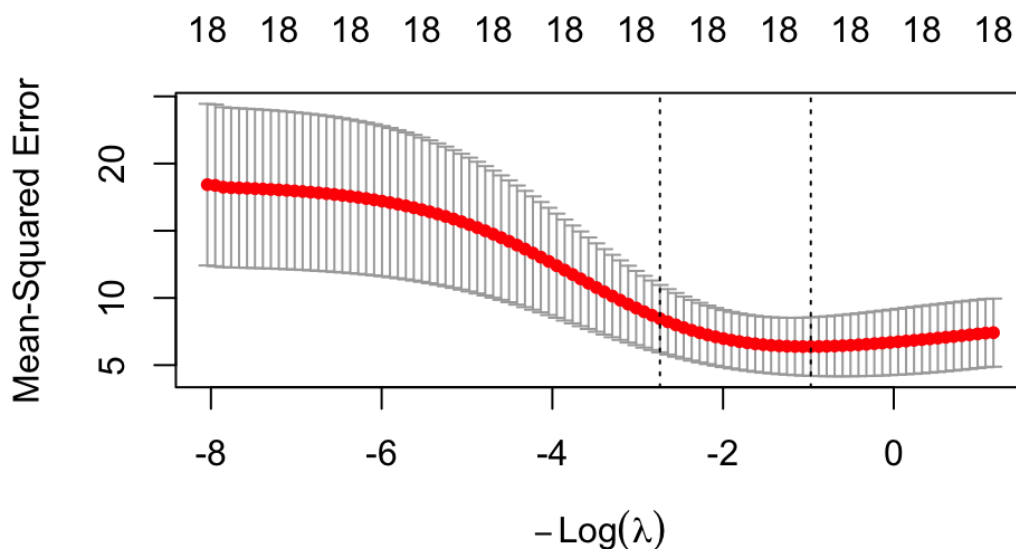
Variabel	VIF	Keterangan
X ₁	8,3042	Multikolinieritas Sedang (VIF < 10)
X ₂	6,4733	Multikolinieritas Sedang (VIF < 10)
X ₃	13,5467	Multikolinieritas Tinggi (VIF > 10)
X ₄	9,3817	Multikolinieritas Sedang (VIF dekat 10)
X ₅	3209,2061	Multikolinieritas Ekstrem (VIF > 10)
X ₆	5,8021	Multikolinieritas Sedang (VIF < 10)
X ₇	3247,6130	Multikolinieritas Ekstrem (VIF > 10)
X ₈	4,3451	Multikolinieritas Rendah (VIF < 5)
X ₉	7,7705	Multikolinieritas Sedang (VIF < 10)
X ₁₀	4,3236	Multikolinieritas Rendah (VIF < 5)
X ₁₁	11,2806	Multikolinieritas Tinggi (VIF > 10)

X ₁₂	3,1412	Multikolinieritas Rendah (VIF < 5)
X ₁₃	13,1135	Multikolinieritas Tinggi (VIF > 10)
X ₁₄	4,4863	Multikolinieritas Rendah (VIF < 5)
X ₁₅	2,1481	Bebas Multikolinieritas (VIF < 5)
X ₁₆	4,4106	Multikolinieritas Rendah (VIF < 5)
X ₁₇	4,1750	Multikolinieritas Rendah (VIF < 5)
X ₁₈	4,8099	Multikolinieritas Rendah (VIF < 5)

Berdasarkan tabel 2 diatas dapat dilihat bahwa terdapat 5 variabel yang memiliki nilai VIF > 10 sehingga terjadi multikolinearitas pada variabel independen. Maka diperlukan metode estimasi yang mampu mengatasi gangguan multikolinearitas sekaligus melakukan seleksi variabel dengan menggunakan metode regresi regularisasi yaitu Regresi *Ridge*, LASSO dan *Elastic-net*.

a. Nilai λ Optimum untuk Regresi *Ridge*

Penentuan parameter penyusutan (λ) yang optimal dalam regresi *Ridge* dilakukan melalui metode *ten-fold cross validation* (validasi silang lipat-10). Parameter λ yang dianggap paling optimal adalah nilai yang menghasilkan CVE terendah. Gambar 2 dibawah menyajikan hasil validasi silang (*cross-validation*) untuk menentukan parameter penyusutan (λ) pada model kemiskinan. Dua garis vertikal putus-putus dalam gambar diatas merupakan batas nilai optimum dengan *Cross-Validation Error* (CVE) minimal yang dapat digunakan. Dalam analisis ini, nilai λ_{1se} lebih direkomendasikan dan dijadikan λ optimum karena nilai mampu melindungi model terhadap *overfitting* dengan pemilihan nilai λ lebih besar dan juga dapat menghasilkan model dengan parameter yang lebih sedikit (dapat melakukan seleksi variabel). Nilai λ untuk model kemiskinan yaitu λ_{1se} sebesar 15,4771 dan $\log \lambda_{1se}$ sebesar 2,7394.

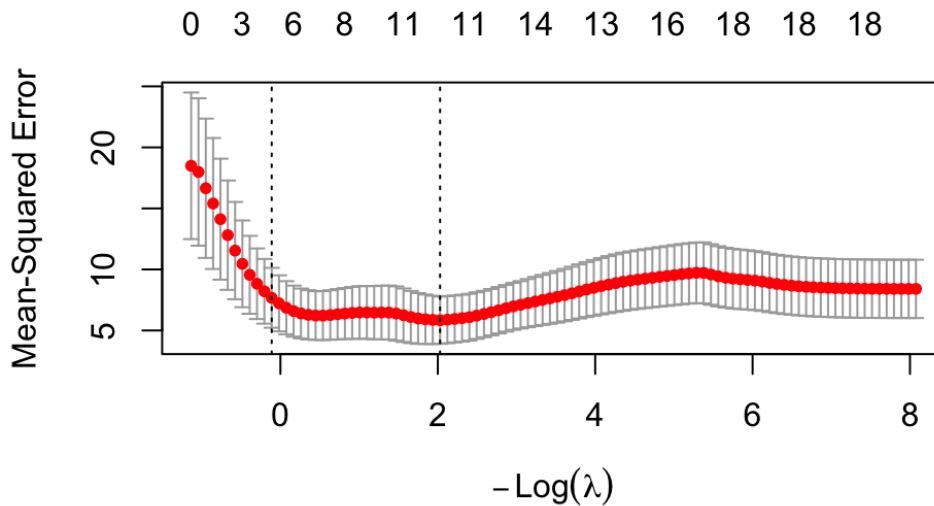


Gambar 2. Validasi silang pada parameter penyusutan (λ) Regresi *Ridge*

b. Nilai λ Optimum untuk LASSO

Parameter penyusutan (λ) dalam model LASSO diperoleh melalui metode validasi silang (*Cross Validation*) dengan tujuan menghasilkan nilai CVE minimum. Prosedur CV pada metode ini sama dengan CV pada Regresi *Ridge* yaitu dengan menggunakan *ten-fold*

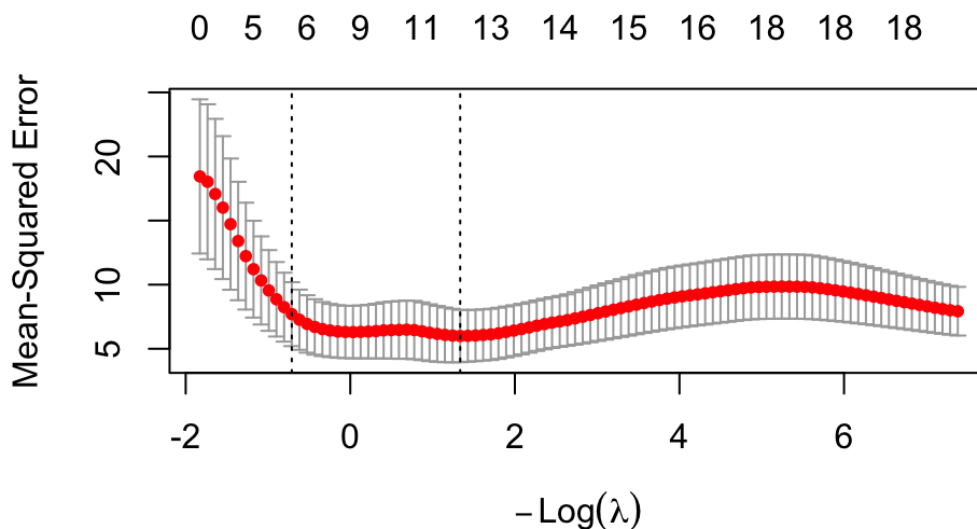
cross validation. Berdasarkan hasil validasi silang pada Gambar 3, hasil CV dengan nilai λ yang optimal diperoleh dalam model kemiskinan yaitu λ_{1se} sebesar 1,118 dan $\log \lambda_{1se}$ sebesar 0,111.



Gambar 3. Validasi silang pada parameter penyusutan (λ) LASSO

c. Nilai λ optimum untuk *Elastic-net*

Metode *Elastic-net* ini bekerja dengan menggabungkan dua jenis penalti penyusutan sebelumnya yaitu penalti regresi *Ridge* $\sum_{j=1}^p \beta_j^2$ dan penalti LASSO $\sum_{j=1}^p |\beta_j|$. Parameter λ pada regresi *Ridge* dikali dengan α sedangkan parameter λ LASSO dikali dengan $(1 - \alpha)$ dengan α bernilai nol hingga satu. Garis vertikal di sebelah kiri menunjukkan λ_{1se} yang menghasilkan model lebih ringkas dengan hanya menyisakan 6 variabel aktif. Berdasarkan hasil pengujian model kemiskinan dengan *Elastic-net*, diperoleh α optimum dengan λ_{1se} sebesar 2,037 dan $\log \lambda_{1se}$ sebesar 0,711.



Gambar 4. Validasi silang pada parameter penyusutan (λ) *Elastic-net*

Koefisien Regresi

Perbandingan regresi pada Regresi *Ridge*, LASSO, dan *Elastic-net* pada model kemiskinan

Tabel 3. Koefisien regresi pada Regresi *Ridge*, LASSO, dan *Elastic-net* pada model kemiskinan

Variabel	Nama Variabel	Metode			
		RLB	Regresi <i>Ridge</i>	LASSO	Elastic-net
<i>Intercept</i>	<i>Intercept</i>	40,62	37,1734	28,2761	30,0386
X ₁	Umur Harapan Hidup	0,4525	-0,0562	.	.
X ₂	Harapan Lama Sekolah	-0,8006	-0,3650	.	.
X ₃	Rata-rata lama sekolah	-1,8290	-0,2807	-0,6897	-0,5958
X ₄	Pengeluaran per Kapita	-0,0000	-0,0001	-0,0002	-0,0002
X ₅	Jumlah Angkatan Kerja	0,0000	-0,0000	.	.
X ₆	Tingkat Partisipasi Angkatan Kerja	0,0210	0,0080	.	.
X ₇	Jumlah Penduduk	-0,0000	-0,0000	.	.
X ₈	Upah Minimum Kabupaten/Kota (dalam Rupiah)	-0,0000	-0,0000	.	.
X ₉	Tingkat Pengangguran Terbuka	0,0804	-0,0479	.	.
X ₁₀	KRT Pertanian	-0,0000	-0,0000	.	.
X ₁₁	Angka Partisipasi Sekolah	0,2109	0,0792	0,0077	0,0600
X ₁₂	Pertumbuhan Ekonomi	-2,234	-0,6206	-1,7127	-1,4072
X ₁₃	Proporsi rumah tangga dengan bahan bakar masak tidak layak	-0,1485	0,0243	.	0,0028
X ₁₄	Persentase rumah tangga dengan sanitasi tidak layak	0,0261	0,0107	0,0030	0,0067
X ₁₅	Persentase rumah tangga dengan lantai tidak layak	0,1843	0,0205	.	.
X ₁₆	Persentase penduduk dengan keluhan kesehatan	-0,0455	-0,0152	.	.
X ₁₇	Laju Pertumbuhan Penduduk	1,103	0,2515	.	.
X ₁₈	Rasio Jenis Kelamin	-0,2628	-0,1370	-0,0245	-0,0851

d. Menentukan Model Terbaik

Pemilihan model terbaik dari metode regresi *Ridge*, LASSO dan *Elastic-net* menggunakan nilai MSE, RMSE yang terkecil dan R^2 yang terbesar. Berdasarkan Tabel 4 model yang terbaik yaitu LASSO dengan MSE sebesar 5,610, RMSE 2,3687 dan nilai R^2

sebesar 67,73%. Hal ini menunjukkan bahwa metode LASSO merupakan model yang paling akurat dan efisien dalam memprediksi tingkat kemiskinan di Provinsi Sumatera Utara.

Tabel 4. Perbandingan MSE, RMSE dan R^2

Metode	MSE	RMSE	R^2
Regresi Ridge	6,5024	2,5499	0,6260
LASSO	5,610	2,3687	0,6773
Elastic-net	5,7065	2,3889	0,6718

Berdasarkan hasil analisis Tabel. 4, model terbaik yang digunakan yaitu LASSO. Dari 18 variabel independen yang digunakan terdapat beberapa variabel yang mengalami penyusutan koefisien hingga tepat bernilai nol. Proses penyusutan ini menunjukkan bahwa metode LASSO berhasil mengatasi masalah multikolinieritas ekstrim pada data dengan cara menyeleksi variabel yang paling relevan. Dalam model ini, terdapat 12 variabel yang mengalami penyusutan hingga ke arah nol dan terseleksi keluar dari model. Variabel-variabel yang koefisien tetap bertahan merupakan variabel yang memiliki pengaruh nyata terhadap tingkat kemiskinan di Provinsi Sumatera Utara. Berdasarkan hasil estimasi tersebut, terdapat 6 variabel yang secara signifikan memengaruhi persentase penduduk miskin, yaitu rata-rata lama sekolah (X_3), pengeluaran perkapita (X_4), angka partisipasi sekolah (X_{11}), pertumbuhan ekonomi (X_{12}), persentase rumah tangga dengan sanitasi tidak layak (X_{14}), dan rasio jenis kelamin (X_{18}).

Variabel yang memengaruhi persentase kemiskinan Provinsi Sumatera Utara merupakan variabel dengan berbagai sektor yaitu infrastruktur dasar, kependudukan, ekonomi dan pendidikan. Pengaruh negatif yang besar pada variabel pertumbuhan ekonomi (X_{12}) dengan koefisien sebesar -1,7127 dan rata-rata lama sekolah (X_3) sebesar -0,6897. Hal ini menunjukkan bahwa pertumbuhan ekonomi dan peningkatan kualitas pendidikan dapat menurunkan angka kemiskinan. Hal ini sejalan dengan penelitian dilakukan oleh Jannah (2023) yang mengatakan bahwa pendidikan berperan penting dalam mengurangi kemiskinan jangka panjang. Kemudian penelitian Faritz (2020) yang dilakukan di Jawa Tengah dengan kesimpulan pertumbuhan ekonomi serta rata-rata lama sekolah berpengaruh signifikan dalam penurunan kemiskinan. Kemudian, variabel pengeluaran perkapita (X_4) yang signifikan mengindikasikan bahwa pentingnya menjaga harga bahan pokok stabil sehingga konsumsi masyarakat tidak mengalami penurunan. Selanjutnya, koefisien positif pada angka partisipasi sekolah (X_{11}) mengindikasikan bahwa pendidikan yang tidak diiringi dengan peningkatan kualitas dan kesesuaian kurikulum dengan dunia kerja belum mampu menurunkan angka kemiskinan. Selain itu, koefisien positif pada variabel sanitasi tidak layak (X_{14}) menunjukkan bahwa adanya keterbatasan infrastruktur yang tidak memadai seperti kebersihan dan kelayakan sanitasi yang bersih dapat menjadi salah satu faktor tingginya persentase kemiskinan. Sementara itu, signifikansinya variabel rasio jenis kelamin (X_{18}) mengindikasikan bahwa pentingnya ketersediaan kesempatan kerja formal serta ramah untuk kaum perempuan. Hal ini sejalan dengan penelitian yang dilakukan oleh Dewi (2023) yang menyimpulkan bahwa rasio jenis kelamin dapat memengaruhi kemiskinannya.

Upaya dalam menurunkan persentase kemiskinan Provinsi Sumatera Utara dapat dengan melihat koefisien yang bernilai negatif seperti pertumbuhan ekonomi, rata-rata lama sekolah, pengeluaran perkapita dan rasio jenis kelamin. Jika keempat variabel ini meningkat, maka persentase kemiskinan Provinsi Sumatera Utara akan turun.

4 KESIMPULAN

Model LASSO merupakan model terbaik dalam memprediksi persentase kemiskinan Provinsi Sumatera Utara Tahun 2025. Dari 18 variabel independen yang pada awalnya dimasukkan ke dalam model kemiskinan, terdapat 12 variabel yang mengalami penyusutan koefisien hingga tepat bernilai nol. Proses penyusutan ini menunjukkan bahwa metode LASSO berhasil mengatasi masalah multikolinieritas ekstrim pada data dengan cara menyeleksi variabel yang paling relevan untuk menghasilkan model yang paling efisien. Variabel yang mengalami penyusutan mengindikasikan bahwa variabel tersebut tidak memiliki peran yang cukup signifikan secara statistik dalam menjelaskan variasi persentase kemiskinan. Sebaliknya, variabel-variabel yang koefisiennya tetap bertahan merupakan variabel yang berpengaruh terhadap tingkat kemiskinan di Provinsi Sumatera Utara. Berdasarkan hasil estimasi tersebut, terdapat 6 variabel yang secara signifikan memengaruhi persentase penduduk miskin, yaitu rata-rata lama sekolah (X3), pengeluaran perkapita (X4), angka partisipasi sekolah (X11), pertumbuhan ekonomi (X12), persentase rumah tangga dengan sanitasi tidak layak (X14), dan rasio jenis kelamin (X18). Upaya dalam menurunkan persentase kemiskinan Provinsi Sumatera Utara Tahun 2025 dapat dilihat dengan koefisien yang bernilai negatif seperti pertumbuhan ekonomi, rata-rata lama sekolah, pengeluaran perkapita dan rasio jenis kelamin. Jika keempat variabel ini meningkat, maka persentase kemiskinan Provinsi Sumatera Utara juga akan mengalami penurunan. Peneliti selanjutnya diharapkan dapat menambahkan variabel makro dan mikro lainnya yang belum tercakup dalam penelitian ini serta mempertimbangkan efek spasial dengan dikombinasikan dengan metode regresi regularisasi sehingga menangkap keterkaitan antar wilayah.

UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terimakasih kepada keluarga serta dosen pembimbing yang telah membantu penulis dalam menyelesaikan artikel ilmiah ini. Penulis berharap karya ilmiah ini dapat memberikan kontribusi bagi pendidikan dan kebijakan di Provinsi Sumatera Utara.

DAFTAR PUSTAKA

- Ashari, R. T., & Athoillah, M. (2023). Analisis Pengaruh tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan kerja, Upah Minimum, Indeks Pembangunan Manusia, Pertumbuhan Ekonomi dan Jumlah Penduduk Terhadap Kemiskinan di kawasan Tadal kuda. *Journal of Development Economic and Social Studies*, 2(2), 313–326. <https://doi.org/10.21776/jdess.2023.02.2.08>
- Badan Pusat Statistik. 2026. Provinsi Sumatera Utara Dalam Angka 2026. Sumatera Utara. Badan Pusat Statistik
- Badan Pusat Statistik. 2026. Keadaan Angkatan kerja Provinsi Sumatera Utara. Sumatera Utara. Badan Pusat Statistik
- Badan Pusat Statistik. 2025. Indikator Kesejahteraan Rakyat Provinsi Sumatera Utara. Badan Pusat Statistik
- Dewi, I. R. (2023). Mengupas Kemiskinan di Provinsi Banten: Bagaimanakah Peran Faktor Kependudukan dan Ekonomi?. *Ecoplan*, 6(2), 100-117.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (1ed.). Wiley. <https://doi.org/10.1002/9781118625590>
- Faritz, M. N., & Soejoto, A. (2020). Pengaruh pertumbuhan ekonomi dan rata-rata lama sekolah terhadap kemiskinan di Provinsi Jawa Tengah. *Jurnal Pendidikan Ekonomi (JUPE)*, 8(1), 15-21.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013a). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013b). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear Regression. Dalam G. James, D. Witten, T. Hastie, R. Tibshirani, & J. Taylor, *An Introduction to Statistical Learning* (hlm. 69–134). Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0_3
- Jannah, M., & Sari, I. F. (2023). Analisis Pengaruh Rata-Rata Lama Sekolah, Angka Harapan Hidup dan Pengeluaran Perkapita Terhadap Kemiskinan Provinsi Nusa Tenggara Barat. *EKOMA: Jurnal Ekonomi, Manajemen, Akuntansi*, 3(1), 164-172.
- Jannah, S. R., Rusyana, A., & Ramadhani, E. (2025). *Perbandingan Metode Regresi OLS, Ridge, dan LASSO dalam Memodelkan Kemiskinan di Indonesia Tahun 2024*.
- John Lu, Z. Q. (2010). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 173(3), 693–694. https://doi.org/10.1111/j.1467-985X.2010.00646_6.x
- Kusuma, G. W., & Wulansari, I. Y. (2020a,b). Analisis Kemiskinan dan dan Kerentanan Kemiskinan dengan Regresi Ridge, LASSO dan Elastic Net di provinsi Jawa Tengah Tahun 2017. *Seminar Nasional Official Statistics*, 2019(1), 503–513. <https://doi.org/10.34123/semnasoffstat.v2019i1.189>
- Kutner, M. H. (Ed.). (2005). *Applied linear statistical models* (5. ed). McGraw-Hill Irwin.
- Ohyver, M. (2011). Metode Regresi Ridge Untuk Mengatasi Kasus Multikolinear. *ComTech: Computer, Mathematics and Engineering Applications*, 2(1), 451. <https://doi.org/10.21512/comtech.v2i1.2782>
- Panjaitan, A. Z., Badawi, A., Rambe, R. R., & Audina, R. (2025). *Perbandingan Metode Kuadrat Terkecil, LASSO, Ridge, dan Elastic Net dalam Memprediksi Harga Mobil*. 01(2).
- Purba, B., Tarigan, D. R., & Ginting, R. O. (2024). Analisis Kebijakan Sumber Daya Alam Provinsi Sumatera Utara. *Economic Reviews Journal*, 3(3). <https://doi.org/10.56709/mrj.v3i3.246>
- Putri, S. B., & Suliadi, S. (2023). Penerapan Metode Regresi Ridge Parsial untuk Mengatasi Masalah Multikolinearitas untuk Memodelkan Faktor-Faktor yang Mempengaruhi Kemiskinan di Jawa Tengah pada Tahun 2020. *Bandung Conference Series: Statistics*, 3(1). <https://doi.org/10.29313/bcss.v3i1.5578>
- Susanti, I., & Saumi, F. (2022). *Penerapan Metode Analisis Regresi Linear Berganda untuk mengatasi Masalah Multikolinearitas Pada Kasus Indeks Pembangunan Manusia (IPM) di kabupaten Aceh Tamiang*. 4.
- Widayatsari, A., Harlen, H., Damaiyanti, D., & Misdawita, M. (2025). Mengurai Dimensi Kemiskinan di Riau: Pendekatan Multidimensional dan Dampaknya pada Pembangunan. *Jurnal Ekonomi dan Pembangunan Indonesia*, 25(1). <https://doi.org/10.7454/jepi.v25i1.1728>
- Zam, M. Z. (2025). *Dampak Kebijakan Pembangunan Infrastruktur Terhadap Pengurangan Kemiskinan di Daerah Pedesaan*.