

FAKTOR PENENTU STATUS GIZI DENGAN METODE REGRESI REGULARISASI RIDGE, LASSO, DAN ELASTIC NET (STUDI KASUS: POSYANDU DESA SEJAHTERA PROVINSI SUMATERA UTARA)

Ruth Aryati Manihuruk^{1*}, Fonda Leviany²

^{1,2} Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

*Penulis korespondensi: rutharyatimanihuruk30@gmail.com

ABSTRAK

Gizi balita mengacu pada makanan yang telah disesuaikan dengan usia sehingga mampu memenuhi kebutuhan energi, protein, vitamin, mineral, dan nutrisi lainnya yang dibutuhkan balita. Gizi balita yang tepat pada balita membantu mencegah penyakit, disabilitas, dan masalah kesehatan lainnya serta meningkatkan perkembangan otak dan kesehatan fisik. Tujuan penelitian ini adalah untuk mengidentifikasi faktor-faktor pasti yang mempengaruhi status gizi balita dan sejauh mana pengaruhnya. Data penelitian menggunakan data balita yang berkunjung ke posyandu di Desa Sejahtera, Kecamatan Siantar, Kabupaten Simalungun, Provinsi Sumatera Utara. Metodologi penelitian menggunakan metode regresi ridge, lasso, dan elastic network untuk menganalisis faktor-faktor yang mempengaruhi, mengatasi masalah data seperti multikolinearitas dan overfitting, serta membandingkan keandalan ketiga metode melalui nilai Mean Squared Error (MSE) untuk mengetahui variable yang menjadi penentu status gizi balita. Variabel yang digunakan dalam metode ini berupa satu variable terikat (Y) yaitu Berat Badan menurut Berat Badan dan Tinggi Badan (BB/TB) dengan sebelas variable bebas. Berdasarkan hasil nilai MSE ketiga metode yang diteliti diperoleh bahwa metode lasso memiliki nilai MSE terkecil yaitu sebesar 0,000008135 yang mengartikan bahwa metode lasso sebagai metode terbaik dalam penelitian ini. Metode ini mendapatkan hasil 8 variabel dengan variabel jenis kelamin, status ekonomi, berat badan lahir dan berat badan memiliki koefisien positif signifikan. Variabel usia, pengeluaran per individu, kegiatan ibu dan tinggi badan mengidentifikasikan bahwa semakin tinggi usia, pengeluaran per individu, ibu yang memiliki kegiatan (bekerja) dan tinggi badan akan menyebabkan kecenderungan status gizi yang rendah begitu pula sebaliknya. Sementara itu variabel pendapatan orang tua, pendidikan ibu dan jumlah anggota keluarga tidak menunjukkan pengaruh yang signifikan sehingga ketiga variabel tersebut dieliminasi dalam model ini.

Kata Kunci : Multikolinieritas, Regresi Ridge, Regresi Lasso, Regresi Elastic Net, Status Gizi Balita

1 PENDAHULUAN

Nutrisi mencakup zat gizi penting dalam makanan sehari-hari, termasuk karbohidrat, protein, lemak, vitamin, mineral, dan air. Zat gizi ini diubah menjadi glukosa melalui glikolisis, menghasilkan adenosin trifosfat (ATP) sebagai sumber energi utama tubuh untuk aktivitas sehari-hari. Nutrisi merupakan faktor penting dalam mendorong pertumbuhan dan perkembangan sekaligus menjaga fungsi fisiologis. Organisasi Kesehatan Dunia (WHO) menekankan bahwa nutrisi penting tidak hanya untuk pertumbuhan tetapi juga untuk menjaga fungsi fisiologis dan kesehatan umum. Gizi balita adalah makanan yang sesuai dengan usia dan memenuhi kebutuhan energi, protein, vitamin, mineral, dan nutrisi balita (Ruswandi dkk,2024).

Penelitian terkait faktor penentu status gizi di Indonesia sudah banyak dilakukan untuk mengidentifikasi faktor apa saja yang mempengaruhi status gizi di Indonesia. Salah satu penelitian tersebut dilakukan oleh Pujiati, dkk (2017) yang mengidentifikasi kasus kekurangan gizi pada anak di bawah usia lima tahun di Kota Makasar. Di antara kasus kurang gizi tersebut, proporsi berjenis kelamin perempuan lebih besar (57,8%) dan usia mereka antara 25-36 bulan (42,2%) dalam kasus kurang gizi yang terjadi di Makasar tahun 2017. Setelah menyesuaikan semua faktor perancu, malnutrisi berat pada masa kanak-kanak secara signifikan berhubungan dengan status ekonomi rendah ($OR = 9,514$; $CI95\% = 1,219-74,230$), penyakit infeksi ($OR = 5,744$; $CI95\% = 2,615-12,613$), kelahiran prematur ($OR = 3,613$; $CI95\% = 1,452-8,987$), berat badan lahir rendah ($OR = 2,957$; $CI95\% = 1,387-6,302$) dan pemberian ASI eksklusif ($OR = 3,537$; $CI95\% = 1,481-8,446$). Penggunaan fasilitas kesehatan (posyandu) tidak berhubungan dengan malnutrisi berat penyakit infeksi, berat badan lahir rendah, dan pemberian ASI eksklusif merupakan faktor risiko terhadap malnutrisi berat pada anak usia di bawah lima tahun di Kota Makassar (Pujiati dkk, 2017).

Berdasarkan survei gizi nasional tahun 2022, prevalensi malnutrisi di Indonesia tercatat sebesar 21,6%, dengan angka 21,1% di Provinsi Sumatra Utara dan 17,4% di Kabupaten Simalungun. Pemenuhan gizi yang memadai selama masa kanak-kanak sangatlah penting untuk mencegah penyakit serta mendukung perkembangan fisik dan kognitif; sebaliknya, malnutrisi tidak hanya membebani layanan kesehatan secara berat, tetapi juga meningkatkan kejadian penyakit menular, memperparah kemiskinan, serta mengancam kesejahteraan dan masa depan bangsa. Jika tidak segera ditangani, masalah ini dapat berkembang menjadi krisis nasional yang serius.

Studi ini membandingkan model Ridge, Lasso, dan Elastic Net dalam analisis data (yang mencakup informasi numerik maupun deskriptif) untuk mengidentifikasi penyebab malnutrisi pada anak, sekaligus mengatasi masalah seperti overfitting dan multikolinearitas. Metode Ordinary Least Squares (OLS) memiliki keterbatasan terkait prediksi dan penentuan signifikansi statistik. Model Ridge membatasi koefisien menggunakan "penalti kuadrat", Lasso menggunakan "penalti beta absolut" untuk estimasi, sedangkan Elastic Net menggabungkan kedua pendekatan tersebut. Masing-masing model ini efektif dalam menangani multikolinearitas dan meningkatkan pemahaman terhadap data. (Holiawati,2025).

Oleh karena itu, untuk mengatasi akar penyebab gizi buruk pada anak di Indonesia khususnya di Sumatera Utara, Simalungun, Kecamatan Siantar, dan Desa Sejahtera sangat penting untuk mengembangkan kebijakan yang efektif. Penelitian ini bertujuan untuk mengidentifikasi penyebab gizi buruk pada anak usia 0-5 tahun dan mengusulkan strategi spesifik untuk mengatasinya dengan menggunakan metode regularisasi (regresi ridge, regresi lasso dan regresi elastic net) guna mengatasi multikolinieritas, outlier dan overfitting. Penurunan nasional dalam gizi buruk usia 0-5 tahun akan menunjukkan langkah signifikan dalam pembangunan nasional. Dalam konteks ini, peneliti tertarik untuk meneliti status gizi yang berawal dari wilayah terkecil yaitu desa dan

memilih Desa Sejahtera Kecamatan Siantar Kabupaten Simalungun Provinsi Sumatera Utara sebagai bagian dari penelitiannya guna upaya menyongsong Indonesia emas dan bebas gizi buruk di tahun 2045 guna untuk memberikan solusi yang tepat dalam menangani kasus gizi buruk di Desa Sejahtera.

2 METODE

2.1 Pengumpulan dan Persiapan Data

Data yang digunakan merupakan data sekunder yang diperoleh dari kader posyandu Desa Sejahtera, Kecamatan Siantar, Kabupaten Simalungun, Provinsi Sumatera Utara. Data bersifat cross-sectional berdasarkan balita yang berkunjung memeriksa anak-anak di Posyandu pada bulan April 2026. Pengumpulan data dilakukan melalui prosedur sebagai berikut.

Tabel 1. Variabel Penelitian

Kode	Variabel	Satuan	Tipe Data
Y	Berat Badan menurut Berat Badan dan Tinggi Badan (BB/TB)	Kg/m ²	Numerik
X ₁	Usia	Bulan	Numerik
X ₂	Jenis Kelamin	- Perempuan - Laki-laki	Kategorik
X ₃	Pendapatan Orang tua	Rupiah	Numerik
X ₄	Status Ekonomi	- Rendah - Tinggi	Kategorik
X ₅	Pengeluaran per Individu	Rupiah	Numerik
X ₆	Kegiatan Ibu	- Tidak Bekerja - Bekerja	Kategorik
X ₇	Pendidikan Ibu	- SD - SMP - SMA - Perguruan Tinggi	Kategorik
X ₈	Jumlah Anggota Keluarga	Orang	Numerik
X ₉	Berat Badan Lahir	Kg	Numerik
X ₁₀	Tinggi Badan	Cm	Numerik
X ₁₁	Berat Badan	Kg	Numerik

2.2 Status Gizi

Status gizi balita ditentukan dengan membandingkan tinggi dan berat badan berdasarkan standar antropometri dari Peraturan Kemenkes RI No. 2 Tahun 2020, yang meliputi empat indeks: BB/U, PB/U, BB/PB, dan IMT/U. Penelitian ini menggunakan variabel berat badan/tinggi badan (BB/TB). Menurut WHO, status gizi anak dibagi menjadi tiga kategori: stunting, kurang berat badan, dan kelebihan berat badan. Cara menghitung Berat Badan menurut Berat Badan dan Tinggi Badan (BB/TB) massa tubuh pada balita:

$$\text{Berat Badan menurut BB/TB} = \frac{\text{Berat Badan (Kg)}}{\text{Tinggi Badan (m)}} \quad (1)$$

dengan nilai Z-Score

$$\text{Z-Score} = \frac{\text{Nilai Pengukuran} - \text{Nilai Median Baku Rujukan}}{\text{Nilai Simpangan Baku (SD) Rujukan}} \quad (2)$$

Dengan kategori dan ambang batas Status Gizi Anak sebagai:

- ≤ 3 SD = Gizi Buruk (severely wasted);
- 3 SD s/d <-2 SD = Gizi Kurang (wasted);
- 2 SD s/d +1 SD = Gizi Baik (Normal);
- $>+1$ SD s/d +2 SD = Beresiko Gizi Lebih (possible risk of overweight);
- $>+2$ SD s/d +3 SD = Gizi Lebih (overweight);
- $>+3$ SD = Obesitas (obese)

(Peraturan Kementerian Kesehatan Republik Indonesia Nomor 2, 2020).

2.3 Statistika Deskriptif

Statistika deskriptif merupakan metode statistika yang digunakan dalam mendeskripsikan, menganalisis dan menyajikan data secara ringkas dalam bentuk numerik dan/atau visual. Dalam penelitian, statistika deskriptif digunakan dalam memberikan gambaran secara umum tentang variable-variabel yang ingin diteliti. Gambaran umum yang biasanya seperti perhitungan nilai rata-rata (mean), nilai tengah (median), nilai yang sering muncul (modus), standar deviasi, sampai distribusi dan kemiringan suatu data. (Dirna dkk, 2025).

Tujuan dari statistika deskriptif adalah untuk menyederhanakan, merangkum, dan menyajikan data sehingga dapat dengan mudah dipahami dalam proses analisis awal. Statistika deskripsi memiliki peran penting dalam penelitian ilmiah dalam mengambil pengambilan kebijakan karena memiliki kemampuan menyederhakan data yang kompleks menjadi informasi yang mudah dipahami. Data mentah yang sulit dipahami dan dibaca dapat diringkaskan dan divisualisasikan sehingga gambaran nyata terhadap fenomena yang terjadi dalam suatu data. (Dirna dkk, 2025).

2.4 Analisis Korelasi

Analisis korelasi adalah suatu analisis yang digunakan untuk mengukur hubungan linier antara dua variabel (variabel dependen dan variabel independen) hubungan tersebut bisa berupa hubungan yang negatif atau hubungan yang positif. Namun pengukuran kekuatan hubungan tersebut tidak menunjukkan hubungan fungsional atau bukan untuk membedakan antara variable bebas dan variable terikat (Umami, 2025).

2.5 Uji Multikolinieritas dan Outlier

Multikolinearitas adalah kondisi di mana data menunjukkan hubungan yang kuat antara variabel independen, menghasilkan model regresi linier. Berikut adalah dampak multikolinearitas pada analisis statistik:

1. Ketidakstabilan koefisien regresi membuat interpretasi model menjadi sulit dan tidak dapat diandalkan;
2. Estimasi varians yang tinggi menimbulkan bias dan meningkatkan kemungkinan kesalahan Tipe I dan Tipe II;
3. Dampak pada prediksi;
4. Kesulitan dalam pengambilan keputusan bisnis.

Koefisien inflasi varians (VIF) dapat dihitung secara manual menggunakan rumus berikut:

$$VIF = \frac{1}{(1-R^2_j)}; \quad j=1,2,\dots,k \quad (3)$$

Keterangan :

VIF = Angka Variance Inflation Factor (VIF)

j = Jumlah Sample 1,2,...k

R^2_j = Koefisien determinasi variable bebas ke-j dengan variable lain;

Selain itu dilakukan deteksi outlier dan observasi berpengaruh untuk melakukan identifikasi data ekstrem yang mengakibatkan distorsi hasil estimasi. Deteksi ini dilakukan dengan menggunakan analisis boxplot (Nisfiannoor, 2009).

2.6 Regresi Ridge

Ridge regression adalah teknik regresi linear yang mengurangi *overfitting* dan meningkatkan kinerja model melalui penggunaan penalti serta validasi silang, sehingga mencegah masalah yang terkait dengan banyaknya variabel atau multikolinearitas. Regresi ridge, atau pengelompokan L2, menambahkan penalti berupa kuadrat nilai koefisien (β^2) ke fungsi kerugian. Penalti ini mendorong model untuk mengurangi nilai koefisien, terutama ketika terdapat korelasi antar variabel (multikolinearitas). Namun, metode ridge tidak sepenuhnya menghilangkan koefisien variabel, sehingga meskipun kontribusinya kecil, koefisien tersebut tetap ada. Berkat analisis data dan metode penelitian yang menggunakan kumpulan data yang relatif besar dan sejumlah besar variabel, metode ini tetap stabil dan dapat memberikan generalisasi yang baik. Rumus untuk fungsi kerugian dalam regresi ridge adalah sebagai berikut (Diyasa dkk, 2025),

$$\text{Loss Ridge} = \text{RSS} + \lambda \sum_{j=1}^p \beta_j^2 \quad (4)$$

dengan:

RSS = Residual Sum of Squares (Jumlah kuadrat kesalahan terkecil);

λ = parameter regulasi untuk mengontrol besarnya penalti;

β_j^2 = koefisien regresi.

2.7 Regresi Lasso

Regresi Lasso (regularisasi L1) melakukan analisis dengan penambahan nilai absolut dari koefisien fungsi kerugian ($|\beta|$). Metode ini menghilangkan koefisien dengan menetapkan menjadi nol sehingga tidak seperti regresi Ridge, yang mengembalikan koefisien untuk variabel, bahkan jika efeknya kecil (Simarmata, 2025). Hal ini berarti bahwa Lasso secara selektif menghilangkan variabel dengan efek minimal. Metode ini membantu mengidentifikasi variabel yang paling berpengaruh dan mengurangi risiko overfitting. Fungsi kerugian dalam regresi Lasso adalah sebagai berikut:

$$\text{Loss Lasso} = \text{RSS} + \lambda \sum_{j=1}^p |\beta_j| \quad (5)$$

Parameter λ merupakan parameter penalti yang bernilai positif, $\lambda > 0$. Seiring meningkatnya nilai λ , estimasi parameter pada regresi LASSO akan terus menyusut menuju 0. Kemudian, jika λ cukup besar, beberapa koefisien akan tepat sama dengan 0 (Diyasa dkk, 2025).

2.8 Regresi Elastic Net

Metode elastis net merupakan kombinasi dari metode ridge (L2) dan lasso (L1), yang menggabungkan kedua metode penalti untuk mengurangi kesalahannya. Metode ini juga akan memberikan penyusutan terhadap koefisien dan eliminasi variable untuk mengatasi kekurangan Metode Ridge dan Metode Lasso. Pemberian kendala pada parameter pinalti terhadap L1-norm dan L2-norm akan menjadi terkontrol dan menjadi elastitis. Seperti dua metode di atas, metode ini cocok digunakan pada kumpulan data besar dengan banyak variabel dan korelasi tinggi, yang seringkali menghasilkan pemilihan variabel yang berkorelasi tinggi (Simon Simarmata, 2025). Adapapun rumus untuk melakukan analisis regresi elastic net adalah sebagai berikut:

$$\text{Elastic Net} = \frac{1}{2p} \sum_{j=1}^P (y_j - x_j \beta)^2 + \lambda \sum_{j=1}^P |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^P \beta_j^2 \quad (6)$$

Parameter $0 < \lambda < 1$ akan mengontrol parameter elastis dan pembeda antara regresi Ridge dan Lasso. (Diyasa, dkk, 2025)

3. HASIL DAN PEMBAHASAN

Status gizi anak usia dini ditentukan oleh beberapa faktor, seperti usia, jenis kelamin, pendapatan orang tua, dan sebagainya. Banyaknya variabel yang terlibat dapat menyebabkan terjadinya multikolinieritas dan pencilan (outliers), yang memengaruhi stabilitas estimasi koefisien dalam model OLS. Pencilan teridentifikasi pada variabel "pendapatan orang tua" dan "pengeluaran per kapita," serta ditemukan adanya multikolinieritas di antara beberapa variabel. Mengingat asumsi standar model OLS tidak terpenuhi, prosedur regularisasi seperti ridge dan elastic net perlu digunakan untuk memperoleh estimasi parameter yang lebih stabil.

3.1 Statistika Deskriptif

Statistika deskriptif memiliki fungsi sebagai peringkas dan penggambar suatu karakteristik data mentah sehingga mudah untuk di pahami. Analisis ini adalah ringkasan suatu ukuran pemusatan seperti tabel di bawah ini:

Tabel 2. Statistika Deskriptif Variabel Numerik

Kode	Variabel	Mean	Sd	Median	Min	Max	Skew	Kurtosis
Y	Berat Badan menurut Berat Badan dan Tinggi Badan (BB/TB)	0,13	0,02	0,13	0,02	0,09	1,26	3,41
X ₁	Usia	36,07	15,01	36,00	17,79	6,00	-0,20	-1,03
X ₃	Pendapatan Orang tua	3997272,73	2866044,78	2650000,00	1704990,00	1000000,00	1,21	0,42
X ₅	Pengeluaran per Individu	1036987,01	817135,47	741666,67	452193,00	214285,71	1,41	1,32
X ₈	Jumlah Anggota Keluarga	4,11	0,93	4,00	3	7	0,53	0,57
X ₉	Berat Badan Lahir	3,23	0,44	3,20	0,30	1,00	-2,28	9,01
X ₁₀	Tinggi Badan	84,72	9,98	86,00	8,90	50,00	-0,66	1,07
X ₁₁	Berat Badan	11,16	2,55	11,00	1,48	5,00	0,84	1,84

Variabel menunjukkan variasi yang cukup besar, terutama pada pendapatan dan pengeluaran. Beberapa variabel seperti berat badan lahir menunjukkan distribusi yang tidak normal dan outlier

ekstrem. Skewness dan kurtosis memberikan gambaran tentang distribusi data yang bisa mempengaruhi analisis statistik lanjutan.

3.2 Tabel Statistika Deskriptif Variabel Kategori

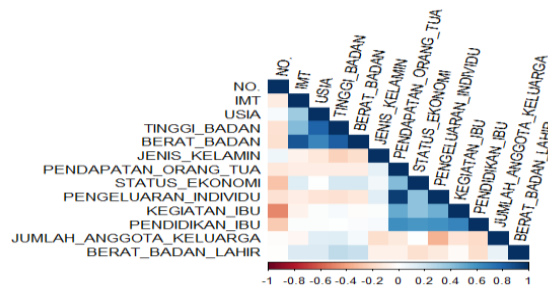
Tabel 3.Tabel Statistika Deskriptif Variabel Kategori

Kode	Variabel	Kategori	Frekuensi	Persentase
X ₂	Jenis Kelamin	1	55	50%
		2	55	50%
X ₄	Status Ekonomi	0	67	60,9%
		1	43	39,09%
X ₆	Kegiatan Ibu	0	63	57,27%
		1	47	42,72%
X ₇	Pendidikan Ibu	1	0	0
		2	4	3,63%
		3	62	56,36%
		4	44	40%

Populasi terbagi merata berdasarkan jenis kelamin. Mayoritas keluarga termasuk kategori ekonomi kurang mampu. Sebagian besar ibu tidak bekerja. Pendidikan ibu umumnya di tingkat SMA dan perguruan tinggi.

3.2 Analisis Korelasi

Analisis korelasi biasanya dilakukan untuk mengetahui hubungan linier antara dua variabel. Hubungan tersebut berupa hubungan yang positif dan negatif tergantung kepada pengaruh yang diberikan variabel bebas kepada variabel terikat. Berikut ditampilkan visualisasi tentang hubungan antara variabel yang mempengaruhi status gizi dengan aplikasi R-Studio.



Gambar 1. Hasil Analisis Korelasi

Hasil analisis menunjukkan adanya korelasi positif tingkat sedang antara massa tubuh dan indikator-indikator seperti usia serta tinggi badan. Hubungan antara jenis kelamin dan variabel-variabel lainnya ditemukan lemah. Pendapatan orang tua dan status ekonomi juga menunjukkan hubungan positif tingkat sedang. Tingkat pendidikan dan status pekerjaan ibu berkorelasi positif tingkat sedang dengan massa tubuh, sedangkan ukuran rumah tangga dan berat badan lahir menunjukkan korelasi yang lemah dan hampir tidak signifikan.

3.3 Uji Multikolinieritas dan Outlier

Uji multikolinieritas digunakan untuk menunjukkan hubungan yang terjadi diantara variabel independen yang dihasilkan oleh suatu model (dalam kasus ini yang melakukan uji multikolinieritas adalah variabel berupa numerik). Pengujian dengan menggunakan nilai Varians

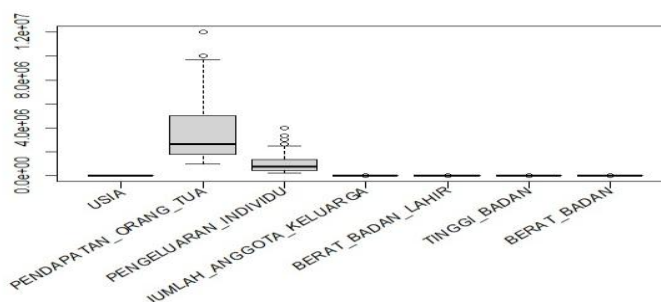
Inflation Factor (VIF), dengan batasan umum $VIF > 10$ menandakan adanya potensi masalah multikolinieritas kuat yang dapat menyebabkan estimasi koefisien tidak stabil pada model OLS.

Tabel 4. Nilai VIF Variabel Prediktor

Usia	Pendapatan Orang Tua	Pengeluaran /Individu	Jumlah Anggota Keluarga	Berat Badan Lahir	Berat Badan	Tinggi Badan
3,224781	19,078699	20,819460	2,465772	1,183798	3,203728	5,828551

Hasil tabel menunjukkan bahwa variabel dengan nilai $VIF > 10$ adalah variabel pendapatan orang tua dan pengeluaran/individu. Hal tersebut menandakan bahwa adanya multikolinieritas yang kuat. Variabel Tinggi badan, usia, jumlah anggota keluarga, berat badan, berat badan lahir dan tinggi badan tidak menunjukkan multikolinieritas yang kuat namun tetap memberikan hubungan yang cukup kuat karena memiliki nilai VIF yang cukup tinggi terutama variabel Tinggi badan dengan nilai 5,828551.

Deteksi anomali memainkan peran penting dalam analisis data, terutama untuk memastikan keakuratan dan validitas hasil. Selain itu, deteksi outlier juga penting untuk memastikan bahwa data yang dianalisis tidak mengandung nilai ekstrem yang tidak mencerminkan tren, sehingga hasil analisis dan model menjadi lebih akurat dan dapat diandalkan.

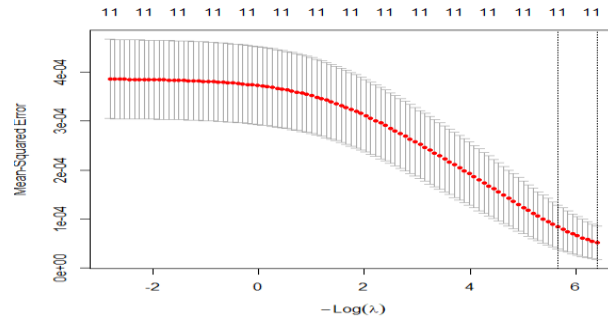


Gambar 2. Boxplot Variabel prediktor

Hasil visualisasi boxplot variabel prediktor (usia, pendapatan orang tua, pengeluaran per individu, jumlah anggota keluarga, berat badan lahir, tinggi badan, berat badan) menunjukkan bahwa terdapat dua variabel prediktor yang memiliki outlier atau dengan arti lain bahwa kedua variabel prediktor tersebut memiliki sejumlah observasi ekstrem yang terletak jauh terhadap rentang nilai utama. Titik tersebut menandakan bahwa pendapatan dan pengeluaran per individu jauh lebih tinggi dibandingkan dengan data dalam sampel. Keberadaan outlier ini sangat perlu diperhatikan karena akan memberikan dampak terhadap peningkatan nilai residual secara tajam serta mempengaruhi kinerja model Ordinary Least Squares (OLS).

3.4 Metode Ridge

Metode ridge adalah metode yang digunakan untuk mengurangi potensi multikolinearitas dalam model regresi linier. Metode ini menerapkan penalti kuadratik pada setiap koefisien variabel regresi untuk fungsi kerugian, sehingga menstabilkan estimasi koefisien dan mencegah overfitting (Diyasa dkk, 2026).

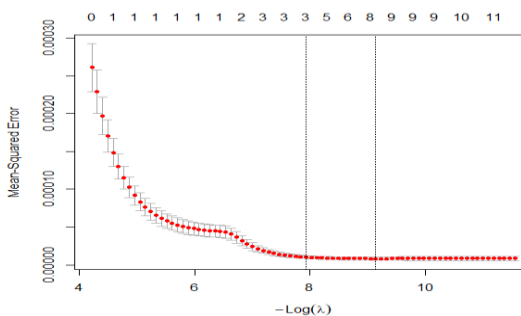


Gambar 3. Cross Validation Parameter λ pada Regresi Ridge

Gambar ini menunjukkan kurva Ridge yang memperlihatkan hubungan antara nilai koefisien regresi dan logaritma parameter lambda ($\text{Log}(\lambda)$) dalam metode normalisasi Ridge. Garis merah menunjukkan bahwa seiring meningkatnya nilai λ , margin menurun, yang mengindikasikan bahwa penalti Ridge semakin kuat, sehingga menghasilkan margin yang lebih kecil. Garis biru di belakangnya adalah interval kepercayaan atau dispersi estimasi koefisien pada setiap nilai λ . Dengan demikian, pada gambar tersebut, seiring meningkatnya derajat normalisasi, bobot variabel menurun hingga hampir nol, yang membantu mengatasi asimetri dan overfitting dalam model regresi.

3.5 Metode Lasso

Metode regresi Least Absolute Shrinkage and Selection Operator (LASSO) adalah suatu pendekatan regresi untuk melakukan pinalti (regularisasi L1) melakukan analisis dengan penambahan nilai absolut dari koefisien fungsi kerugian ($|\beta|$) atau dengan kata lain melakukan penyusutan koefisien dan seleksi terhadap variabel yang diteliti. Parameter yang digunakan pada metode ini adalah parameter elastis $\gamma = 1$. Penyusutan lambda dilakukan dengan 10-fold Cross Validation menggunakan fungsi `cv.glmnet` pada Program R Studio serta pendekatan one standard error untuk mendapatkan model stabil dan mampu menghindari overfitting (Diyasa dkk, 2026; Leviany dkk, 2025; Fitriana, 2025).



Gambar 4. Cross Validation Parameter λ pada Regresi LASSO

Berdasarkan hasil cross validasion pada gambar 4, menunjukkan garis yang vertikal dari kiri ke kanan dengan menunjukkan nilai $\log \lambda_{\min}$ dan nilai $\log \lambda_{1se}$. Gambar ini menunjukkan plot kesalahan kuadrat rata-rata (MSE) terhadap logaritma parameter lambda ($-\text{Log}(\lambda)$), yang kemungkinan diperoleh dari metode Ridge atau Lasso selama proses normalisasi. Titik merah menunjukkan estimasi MSE untuk nilai λ . Nilai MSE awalnya tinggi ketika λ kecil (di sebelah kiri) dan kemudian menurun secara signifikan seiring dengan penurunan λ . Setelah titik $-\text{Log}(\lambda) = 8$, nilai MSE stabil dan tidak banyak berubah, menunjukkan bahwa regularisasi tidak

meningkatkan model. Para penafsir mungkin awalnya menambahkan model ketika λ kecil (regresi non-regular), dan mengakibatkan MSE menurun. Seiring dengan peningkatan λ , MSE menurun ke titik optimal. Setelah titik ini, λ tidak lagi berguna dan model menjadi stabil.

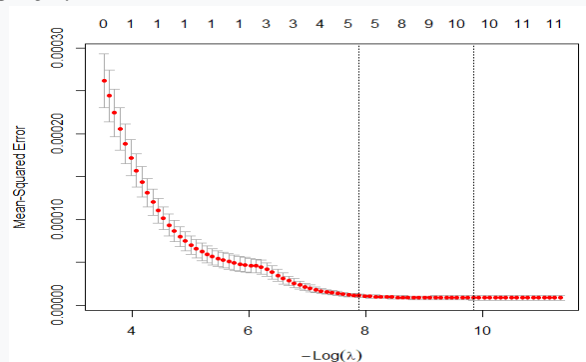
3.6 Metode Elastic Net

Pemodelan Elastic Net merupakan suatu model yang dilakukan mengoptimalkan parameter elastic γ dan λ dengan meningkatkan nilai sebesar 0,1. Metode 10-fold Cross Validasi (CV) untuk mengoptimalkan nilai γ dan pendekatan one standard error untuk mengoptimalkan nilai λ . Evaluasi performa model dari nilai Mean Squared Error (MSE) dan jumlah variabel yang tidak tereliminasi (p). Berikut tabel ringkasan performa model elastic net untuk setiap nilai γ yang diuji pada rentang 0,1 hingga 1.

Tabel 5. Performa Model berdasarkan Parameter Tuning Elastic Net

α	λ_{1se}	MSE	P
0,0	0,0038367574	5,256090e-05	11
0,1	0,0007531711	2,249640e-05	8
0,2	0,0007222573	2,342129e-05	8
0,3	0,0009234831	2,499258e-05	8
0,4	0,0010048619	2,750181e-05	6
0,5	0,0005540898	2,021218e-05	7
0,6	0,0008069065	2,281897e-05	5
0,7	0,0006916341	2,247150e-05	4
0,8	0,0008000121	2,269760e-05	2
0,9	0,0007804554	2,249584e-05	2
1	0,0007024099	2,306358e-05	2

Tabel 5 menunjukkan hasil optimal untuk model Elastic Net dengan nilai alfa sekitar 0,5, yang menghasilkan MSE terendah dengan menggunakan sekitar enam variabel. Jika kesederhanaan lebih diutamakan, nilai alfa di atas 0,8 menjadi alternatif yang layak, karena memungkinkan penggunaan lebih sedikit variabel namun tetap mempertahankan kinerja model yang kompetitif. Oleh karena itu, model yang menghasilkan MSE terendah dipilih untuk menjamin akurasi prediksi, sementara batasan jumlah variabel diterapkan untuk memperoleh model yang lebih sederhana dan lebih mudah diinterpretasikan.



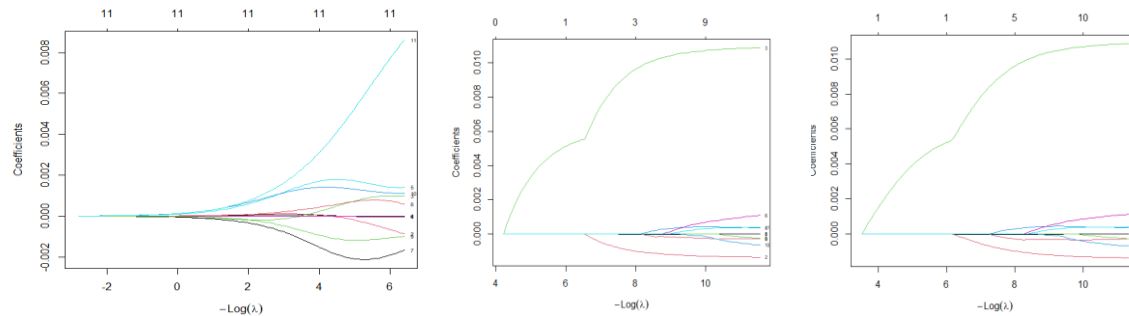
Gambar 5. Cross Validation Parameter λ pada Regresi ELASTIC NET

Gambar 5 memperlihatkan proses elastic net parameter λ dengan $\gamma=0,8$. Dua garis vertikal memperlihatkan bahwa nilai $\log(\lambda_{\min})$ dan $\log(\lambda_{1se})$. Ketika saat λ_{1se} menandakan Mse yang tinggi, hal tersebut menandakan bahwa model sangat kompleks dan terdapat overfitting. Grafik mulai datar/sedikit meningkat pada $\gamma=0,8$, mendekat pada titik minimum kurva serta memberikan

keseimbangan terbaik antara regularisasi dan performa model (dapat menjadi pilihan terbaik sebagai parameter optimal).

3.6 Koefisien Regresi dan Perbandingan Model

Nilai penyusutan yang terjadi pada λ akan memberikan nilai koefisien yang berbeda serta banyaknya variabel prediktor yang berbeda pula untuk setiap modelnya. Berikut gambar nilai penyusutan yang terjadi dari ketiga model yang diteliti.



Gambar 6. Pergerakan koefisien untuk setiap parameter penyusutan λ pada a).Regresi Ridge, b).Regresi Lasso dan c). Elastic Net.

Gambar 6 menampilkan lintasan koefisien untuk tiga metode regresi seiring dengan meningkatnya nilai $\log(\lambda)$. Metode Ridge mempertahankan kehalusan koefisien tanpa mereduksinya menjadi nol, sedangkan metode LASSO mengecilkan koefisien tertentu hingga menjadi nol, yang menunjukkan adanya proses pemilihan variabel secara otomatis. Metode Elastic-Net menunjukkan kombinasi dari kedua perilaku tersebut: sebagian koefisien menurun secara bertahap, sementara sebagian lainnya direduksi menjadi nol. Secara keseluruhan, lintasan-lintasan ini menggambarkan efek penalti koefisien serta kemampuan masing-masing metode dalam mengurangi nilai koefisien dan melakukan pemilihan variabel.

Tabel 6. Nilai Estimasi untuk Setiap Variabel Prediktor yang Terpilih

Kode	VARIABEL	Regresi Ridge $\gamma=0$	Regresi Lasso $\gamma=1$	Elastic-Net $0<\gamma<1$
	Intercept	1,075031e-01	1,075031e-01	1,197748e-01
X ₁	Usia	1,075031e-01	-6,575358e-06	-
X ₂	Jenis Kelamin	9,816170e-04	3,644444e-04	4,252660e-04
X ₃	Pendapatan Orang tua	4,484192e-11	-	2,440690e-10
X ₄	Status Ekonomi	1,408446e-03	2,616006e-04	8,225274e-04
X ₅	Pengeluaran per Individu	-1,165689e-09	-2,331974e-10	-1,417215e-09
X ₆	Kegiatan Ibu	-1,658042e-03	-1,775377e-04	-3,386340e-04
X ₇	Pendidikan Ibu	5,842870e-04	-	-
X ₈	Jumlah Anggota Keluarga	-9,955243e-04	-	-3,417799e-04
X ₉	Berat Badan Lahir	1,112072e-03	8,120937e-05	3,774585e-04

X_{10}	Tinggi Badan	-2,495658e-05	-1,210505e-03	-1,280917e-03
X_{11}	Berat Badan	8,602956e-03	1,047440e-02	1,066238e-02

Tabel 6 menyajikan estimasi nilai koefisien untuk setiap metode. Sementara metode Ridge mempertahankan semua variabel namun mengecilkan nilai koefisiennya, metode LASSO memilih delapan variabel utama dan mengesampingkan variabel lainnya. Metode Elastic-Net menggabungkan karakteristik metode Ridge dan LASSO, serta mengidentifikasi enam variabel. Secara keseluruhan, tanda koefisien menunjukkan konsistensi: status ekonomi, berat badan lahir, dan berat badan memiliki pengaruh positif terhadap status gizi, sedangkan konsumsi per kapita, tingkat aktivitas ibu, dan tinggi badan memiliki pengaruh negatif. Perbedaan utama di antara metode-metode tersebut terletak pada besaran koefisien dan jumlah variabel yang dipilih selama proses regularisasi.

Tabel 7. Perbandingan Kinerja Model

Metode	MSE
Regresi Ridge	0,00005601
Regresi Lasso	0,00000813
Elastic-Net	0,000008462

Berdasarkan tabel 7, diperoleh nilai Mean Squared Error (MSE) terkecil dihasilkan oleh metode lasso sebesar 0,00000813 dibandingkan dengan nilai elastic net yang tidak memiliki interval yang jauh sebesar 0,000008462 kemudian disusul oleh metode ridge yang memiliki interval yang cukup jauh sebesar 0,00005601, sehingga metode lasso terpilih sebagai metode terbaik dalam penelitian ini. Sehingga dapat disimpulkan bahwa metode lasso adalah metode yang paling tepat melakukan estimasi terhadap status gizi balita yang mampu mengatasi masalah multikolinieritas dan melakukan penyusutan koefisien dalam data gizi balita ini. Hal ini menunjukkan pada penelitian ini bahwa pendekatan dengan metode lasso mampu memberikan Tingkat akurasi prediski yang lebih baik di banding dua metode regularisasi lainnya. Model regresi terbaik yang diperoleh menggunakan metode lasso yaitu sebagai berikut:

$$\begin{aligned} \hat{y} = & 1,075031e^{-01} + (-6,575358e^{-06})\text{Usia} + (3,644444e^{-04})\text{Jenis Kelaim} \\ & + (2,616006e^{-04})\text{Status Ekonom} \\ & + (-2,331974e^{-10})\text{Pengeluaran per Individu} \\ & + (-1,775377e^{-04})\text{Kegiatan Ibu} + (8,120937e^{-05})\text{Berat Badan Lahir} \\ & + (-1,210505e^{-03})\text{Tinggi Badan} + (1,047440e^{-02})\text{Berat Badan} \end{aligned}$$

Berdasarkan persamaan diatas diketahui bahwa Jenis kelamin, status ekonomi, berat badan lahir, dan berat badan saat ini menunjukkan pengaruh positif terhadap status gizi, yang mengindikasikan bahwa peningkatan variabel-variabel tersebut cenderung memperbaiki status gizi anak balita. Sebaliknya, usia, pengeluaran per kapita, serta tingkat aktivitas dan kebugaran fisik ibu dikaitkan dengan penurunan status gizi. Di sisi lain, pendapatan orang tua, tingkat pendidikan ibu, dan ukuran keluarga ditemukan tidak memiliki dampak yang signifikan secara statistik; oleh karena itu, variabel-variabel tersebut dikeluarkan dari model.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Penelitian ini membandingkan metode regularisasi Ridge, Lasso, dan Elastic-Net untuk menentukan pendekatan terbaik dalam menganalisis luaran kesehatan pada anak balita di Posyandu Desa Sejahtera, Sumatera Utara. Hasil penelitian menunjukkan bahwa Lasso merupakan metode yang optimal karena menghasilkan *mean squared error* (MSE) terendah sebesar 0,00000813; metode ini secara efektif menangani berbagai variabel, pencilan (*outlier*), dan *overfitting* sekaligus memberikan prediksi yang lebih akurat. Lasso mengidentifikasi delapan variabel signifikan: jenis kelamin, status ekonomi, berat badan lahir, dan berat badan menunjukkan hubungan positif dengan luaran kesehatan, sedangkan usia, pengeluaran per kapita, pekerjaan ibu, dan tinggi badan menunjukkan hubungan negatif. Variabel seperti pendapatan orang tua, tingkat pendidikan ibu, dan ukuran keluarga ditemukan tidak signifikan sehingga tidak disertakan dalam analisis. Sebagai kesimpulan, metode Lasso merupakan pendekatan yang paling tepat untuk memprediksi status kesehatan anak balita di wilayah penelitian tersebut.

4.2 Saran

Berdasarkan hasil tersebut, disarankan agar penelitian selanjutnya menggunakan teknik regularisasi seperti Ridge dan Lasso. Metode-metode ini berperan penting dalam mengatasi masalah seperti multikolinieritas, pencilan (*outliers*), dan *overfitting*, sekaligus meningkatkan akurasi serta kemampuan interpretasi model. Selain itu, guna memastikan penerapan model yang lebih luas dan mengidentifikasi variabel yang mampu memprediksi status gizi balita secara akurat, disarankan untuk memperluas cakupan dataset serta variasi variabel yang dipertimbangkan.

DAFTAR PUSTAKA

- Dirna, M., dkk. (2025). *Statistika Deskriptif*. Star Digital Publishing. https://www.google.co.id/books/edition/Statistika_Deskriptif/Xb14EQAAQBAJ.
- Diyasa, dkk. (2025). *Dasar Perencanaan Proyek Sains Data*. Thalibul Ilmi Publishing & Education. <https://books.google.co.id/books?id=ULZ5EQAAQBAJ>.
- Diyasa, dkk. (2026). *Machine Learning*. Media Nusa Creative. https://books.google.com/books/about/MACHINE_LEARNING.html?id=9RbCEQAAQBAJ.
- Fitriana, I. N. L., Leviany, F., Kasmiarno, K. S., & Nuramaliyah. (2025). Identifying Workforce Size Determinants in Rebana and Arumanis Using Least Absolute Shrinkage and Selection Operator (LASSO) Regression. *Jurnal Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, 18(2), 1025–1040. <https://doi.org/10.36456/jstat.vol18.no2.a10790>.
- Holiawati. (2025). *Metodologi Penelitian*. Adab Indonesia. https://books.google.com/books/about/Metodologi_Penelitian.html?id=7qBCEQAAQBAJ.
- Kementerian Kesehatan Republik Indonesia. (2022). *Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Badan Kebijakan Pembangunan Kesehatan. <https://www.badankebijakan.kemkes.go.id/laporan-hasil-survei/>.
- Khoirunissa, H. A., Wijaya, A. R., Isnaini, B., & Ferawati, K. (2025). Analisis Faktor-Faktor Penyebab Inflasi di Indonesia Menggunakan Regresi Ridge, LASSO, dan Elastic-Net. *Indonesian Journal of Applied Statistics*, 7(2), 121–130. <https://doi.org/10.13057/ijas.v7i2.96921>.

- Levianny, F., Fitriana, I. N. L., & Amin, N. N. (2025). Feature Selection pada Indikator Indeks Ekonomi Hijau di Indonesia dengan Machine Learning. *SENTRI: Jurnal Riset Ilmiah*, 4(9), 1822–1835. <https://doi.org/10.55681/sentri.v4i9.4615>.
- Nisfiannoor, M., dkk. (2009). *Pendekatan Statistika Modern untuk Ilmu Sosial*. Salemba Humanika.
https://books.google.com/books/about/Pendekatan_statistika_Modern_untuk_Il.html?id=1j_O7aHTZD8C.
- Panjaitan, dkk. (2025). Perbandingan Metode Kuadrat Terkecil, LASSO, Ridge, dan Elastic Net dalam Memprediksi Harga Mobil. *Reliability: Journal of Industrial Statistics*, 1, 45–58. <https://journals.arcstics.id/reliability/article/view/25>.
- Peraturan Menteri Kesehatan Republik Indonesia. (2020). Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang Standar Antropometri Anak. <https://share.google/AlOgv15xrUgGI9Ajt>.
- Pujianti, Kurnia, et al. "Identifikasi Kasus Kekurangan Gizi pada Anak di Bawah Usia Lima Tahun di Kota Makassar." *Kes Mas: Jurnal Fakultas Kesehatan Masyarakat Universitas Ahmad Daulan*, vol. 11, no. 2, 2017, pp. 137-142, doi:10.12928/kesmas.v11i2.5787.
- Rifkhan. (2023). *Pedoman Metodologi Penelitian Data Panel Dan Kuesioner*. Adanu Abimata. <https://books.google.co.id/books?id=UN2vEAAAQBAJ>.
- Ruswandi, M. P., Ranifah, R. S., & Adawiah, R. (2024). Kekurangan Gizi Pada Balita. *Health Community Service*, 2(2), 59–66. <https://doi.org/10.47709/hcs.v2i2.3344>
- Sarni, Y., Hutagalung, V., Lestari, A. R., Usmaini, R., & Akbar, R. (2022). Peningkatan status gizi balita kekurangan gizi dari intervensi program Pemberian Makanan Tambahan (PMT) di wilayah kerja Puskesmas Klasaman Kota Sorong. *TROPHICO*, 2(1), 46–53. <https://doi.org/10.32734/trophico.v2i1.8658>.
- Septikasari, M. (2018). *Status Gizi Anak dan Faktor yang Memengaruhi*. UNY Press. https://books.google.com/books/about/STATUS_GIZI_ANAK_DAN_FAKTOR_YANG_MEMPENG.html?id=gjxsDwAAQBAJ.
- Simarmata, dkk. (2025). *Data Mining Menggunakan Python*. Eureka Media Aksara. https://books.google.com/books/about/Data_Mining_Menggunakan_Python_Teori_dan.html?id=AGLGEQAAQBAJ.
- Siregar, dkk. (2024). Penerapan Analisis Lasso dalam Menentukan Penyebab Utama Faktor Kematian Pasien Komorbid yang Terinfeksi Covid-19. (2024). *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika*, 7(1), 290–296. <https://doi.org/10.30605/proximal.v7i1.3456>.
- 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>.
- Umami. (2025). *Konsep Dasar Biostatistik*. Pelita Medika. https://www.google.co.id/books/edition/Konsep_Dasar_Biotatistik/XQYiEAAAQBAJ.