

Application of LASSO Regression and Robust Regression Analysis in Cases of Malnourished Babies in West Java in 2021

Penerapan Analisis Regresi LASSO dan Regresi Robust pada Kasus Bayi Gizi Buruk di Jawa Barat Tahun 2021

Cindy Indriyani^{1†}, Eliza Rahmadania¹, Fadhilah Yumna¹, Naiya Dzil Izzati¹, Sachnaz Desta Oktarina¹, Akbar Rizki¹, and Akmal Riza Wibisono¹

¹Department of Statistics, IPB University, Indonesia

[†]corresponding author: 1808cindy@apps.ipb.ac.id

Abstrak

Gizi buruk bertanggung jawab atas hampir 45% kematian bayi di Indonesia. Guna menekan tingginya persentase tersebut perlu diketahui peubah-peubah yang memengaruhi kasus gizi buruk pada bayi di Indonesia. Peubah-peubah yang diduga memengaruhi hal tersebut seringkali rentan terhadap multikolinearitas dan amatan pencilan sehingga diperlukan penanganan terhadapnya. Peubah yang digunakan pada penelitian ini diakses melalui laman Open Data Jabar, Badan Pusat Statistik (BPS) Provinsi Jawa Barat, dan profil kesehatan Provinsi Jawa Barat tahun 2021. Jawa Barat dipilih sebagai objek studi karena telah berhasil menurunkan angka prevalensi *stunting* pada tahun 2021 sehingga besar kemungkinan persentase bayi gizi buruk juga menurun. Delapan peubah prediktor yang digunakan dalam penelitian ini yaitu jumlah posyandu aktif (unit), berat badan lahir rendah (persen), inisiasi menyusui dini (persen), bayi penerima ASI eksklusif (orang), besaran upah minimum (rupiah), perilaku hidup bersih dan sehat pada rumah tangga (rumah tangga), akses penduduk terhadap air minum memenuhi syarat (persen), dan tempat pengelolaan makanan memenuhi syarat kesehatan (persen). Metode pengendalian multikolinearitas yang digunakan adalah *Ridge regression* dan *LASSO regression*. Model terbaik untuk menangani multikolinearitas didapatkan dari *LASSO regression* yang terdiri dari tiga peubah prediktor. Adapun penanganan kasus pencilan dikendalikan melalui *Robust regression Estimation-M* dengan pembobot *Huber*. Model akhir yang diperoleh setelah mengendalikan multikolinearitas dan amatan pencilan terdiri dari tiga peubah prediktor dengan *R-square* 0,422.

Kata Kunci: gizi buruk, multikolinearitas, regresi lasso, pencilan, regresi robust

1. Pendahuluan

Mutu sumber daya manusia (SDM) yang meningkat merupakan salah satu tujuan pembangunan nasional. Hal tersebut dapat dilakukan dengan memperhatikan proses tumbuh kembang anak dalam menciptakan SDM yang produktif, cerdas, dan sehat. Hal tersebut harus disertai dengan pemenuhan kebutuhan dasar anak, seperti makanan bergizi dan perawatan yang diberikan dengan baik (Irmah *et al.* 2021).

Gizi buruk terjadi saat tubuh tidak menerima nutrisi yang cukup untuk memenuhi kebutuhan dasar dalam pertumbuhan, perkembangan, dan fungsi tubuh. Hal ini sangat berbahaya bagi kondisi bayi karena pada umur di bawah 5 tahun adalah saat terpenting bagi perkembangan otak, organ tubuh, dan sistem kekebalan tubuh bayi. Gizi buruk adalah wujud terparah dari keadaan kurang gizi tahap berat dalam periode yang panjang, gizi buruk masuk ke dalam kategori kejadian luar biasa (KLB) dengan penemuan kasus baru yang harus dilaporkan oleh puskesmas ke jenjang yang lebih tinggi (Irmah *et al.* 2021). Menurut Kementerian Kesehatan, isu gizi buruk pada bayi merupakan salah satu tantangan dalam kesehatan global, terutama di negara berkembang seperti Indonesia. Kualitas sumber daya manusia dalam suatu negara sangat bergantung dengan keadaan gizi masyarakatnya. Administratif pemerintahan Provinsi Jawa Barat terbagi ke dalam 27 kabupaten/kota dengan jumlah penduduk sebesar 48.782.400 jiwa pada tahun 2021 berdasarkan Badan Pusat Statistik (BPS) dengan persentase bayi gizi buruk terbesar terdapat di Kota Banjar, yaitu sebesar 25,5%. Dinas Kesehatan (2023) menyatakan bahwa pada tahun 2021 angka prevalensi *stunting* di Provinsi Jawa Barat telah menurun sebesar 4,3%, sehingga besar kemungkinan persentase bayi gizi buruk di Jawa Barat pada tahun 2021 juga menurun.

Gizi buruk dapat disebabkan oleh berbagai faktor yaitu jumlah posyandu aktif, berat badan lahir rendah (BBLR), inisiasi menyusui dini, bayi penerima ASI eksklusif (Usada *et al.* 2021), besaran upah minimum (Nurmaliza dan Herlina 2019), perilaku hidup bersih dan sehat pada rumah tangga (Ummusalma dan Syafridah 2022), akses penduduk terhadap air minum memenuhi syarat (Ariesthi *et al.* 2020), dan tempat pengelolaan makanan (TPM) memenuhi syarat kesehatan (Triveni *et al.* 2023). Dampak dari gizi buruk diantaranya gangguan metabolisme tubuh, tingkat terinfeksi tinggi, serta menurunnya kemampuan kognitif dan produktivitas (Usada *et al.* 2021). Gizi bukan hanya tanggung jawab pemerintah atau tenaga kesehatan saja, tetapi kepentingan bersama yang melibatkan semua lapisan masyarakat.

Salah satu kendala yang dihadapi dalam penerapan analisis regresi adalah pencilan. Mengatasi pencilan dengan cara mengeluarkannya dari data bukanlah solusi yang ideal karena dapat mengakibatkan estimasi parameter yang kurang tepat dalam menggambarkan data. Alternatif yang bisa digunakan adalah regresi *Robust* karena model yang dihasilkan akan bersifat kekar terhadap keberadaan pencilan dan *resistance* (Husain dan Jamaluddin 2023). Analisis regresi harus memenuhi berbagai asumsi klasik seperti autokorelasi, heteroskedastisitas, dan multikolinearitas. Multikolinearitas dapat dikendalikan dengan regresi LASSO dan regresi *Ridge*. Regresi *Ridge* dapat menyusutkan koefisien mendekati nol, sedangkan koefisien regresi LASSO dapat menyusut tepat di nol (Siregar *et al.* 2024).

Penelitian ini bertujuan mendapatkan model terbaik dalam menggambarkan masalah persentase bayi gizi buruk di Provinsi Jawa Barat pada tahun 2021 dengan menerapkan regresi LASSO, regresi *Ridge*, dan regresi *Robust* Estimasi-M, serta menentukan peubah-peubah prediktor yang berpengaruh signifikan terhadap persentase bayi gizi buruk sehingga pemerintah dapat mengambil langkah terbaik

untuk menentukan program-program yang tepat dalam menurunkan persentase gizi buruk pada bayi baik di Provinsi Jawa Barat maupun Indonesia.

2. Metodologi

2.1 Bahan dan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Open Data Jabar, profil kesehatan Provinsi Jawa Barat tahun 2021, dan Badan Pusat Statistik Provinsi Jawa Barat terkait balita dengan gizi buruk. Data penelitian terdiri dari 8 peubah prediktor dan satu peubah respon dari masing-masing kabupaten/kota di Jawa Barat pada tahun 2021 yang berjumlah 27 amatan. Pemilihan peubah pada penelitian didasarkan pada penelitian yang telah dilakukan terlebih dahulu berdasarkan faktor yang memengaruhi peubah respon untuk dilakukan pendugaan.

Tabel 1: Daftar peubah yang digunakan.

Kode	Peubah	Satuan	Referensi
Y	Persentase balita gizi buruk pada bayi	Orang	
X ₁	Jumlah posyandu aktif	Unit	(Usada <i>et al.</i> 2021)
X ₂	BBLR (Berat badan lahir rendah)	Persen	(Usada <i>et al.</i> 2021)
X ₃	Inisiasi Menyusui Dini	Persen	(Usada <i>et al.</i> 2021)
X ₄	Bayi penerima ASI eksklusif	Orang	(Usada <i>et al.</i> 2021)
X ₅	Besaran upah minimum	Rupiah	(Nurmaliza dan Herlina 2019)
X ₆	Perilaku hidup bersih dan sehat pada rumah tangga	Rumah tangga	(Ummusalma dan Syafridah 2022)
X ₇	Akses penduduk terhadap air minum memenuhi syarat	Persen	(Ariesthi <i>et al.</i> 2020)
X ₈	TPM (Tempat pengelolaan makanan) memenuhi syarat kesehatan	Persen	(Triveni <i>et al.</i> 2023)

2.2 Metode Analisis Data

Tahapan dalam menganalisis data yang dilakukan dalam penelitian adalah sebagai berikut:

1. Melakukan eksplorasi data terhadap peubah respon dan peubah prediktor.

2. Melakukan pengujian untuk mendeteksi multikolinearitas antar peubah prediktor.

Multikolinearitas adalah kondisi ketika antar peubah prediktor memiliki korelasi atau hubungan linear yang kuat sehingga memiliki sifat tidak saling bebas. Jika terdapat korelasi yang kuat antar peubah prediktor, maka hubungan antar peubah prediktor dan peubah terikatnya menjadi terganggu. Analisis koefisien korelasi dengan menggunakan plot korelasi memiliki tujuan untuk melihat apakah terdapat hubungan antar dua peubah prediktor. Lemahnya koefisien korelasi antar peubah prediktor bertujuan untuk menunjukkan bahwa data penelitian tidak memiliki multikolinearitas. Selain itu, multikolinearitas dapat dideteksi dengan menggunakan faktor inflasi ragam atau VIF (*Variance Inflation Factor*). Multikolinearitas ditandai dengan nilai $VIF > 10$. VIF dapat dirumuskan sebagai berikut:

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

keterangan:

R_j^2 = koefisien determinasi peubah j terhadap peubah prediktor lainnya

3. Penanganan multikolinieritas dengan menggunakan regresi Ridge (Gulud) dan regresi LASSO.

- a. Regresi *Ridge* (Gulud).

Regresi *Ridge* dapat membuat koefisien regresi stabil karena adanya multikolinearitas. Penduga yang bias dengan varians yang kecil dapat mengurangi dampak dari multikolinearitas. Persamaan yang terbentuk adalah sebagai berikut:

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

$$\hat{\beta}_{gulud} = (X'X + \lambda I)^{-1} (X'y) \quad (3)$$

- b. Regresi LASSO.

Regresi LASSO (*Least Absolute Shrinkage and Selection Operator*) merupakan metode penyusutan yang dapat mengatasi permasalahan multikolinearitas. Regresi LASSO dapat melakukan seleksi peubah secara otomatis dengan menyusutkan dugaan koefisien mendekati nol. Persamaannya dapat dirumuskan sebagai berikut:

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

Nilai lambda (λ) yang cukup besar pada regresi LASSO dapat menjadikan dugaan koefisien tepat nol. Hal tersebut menyebabkan LASSO menghasilkan peubah prediktor dengan jumlah yang lebih sedikit.

4. Penentuan model terbaik untuk menangani multikolinearitas dengan AIC dan PRESS.

Salah satu kriterianya adalah *akaike information criteria* (AIC). AIC mengukur kualitas relatif pada pemodelan statistika untuk memberikan pemilihan model terbaik dari beberapa kombinasi model lainnya. AIC didasarkan pada metode *maximum likelihood estimation*.

$$AIC = n \log(\sigma)^2 + n \log[(n - p - 1)/n] + 2p + 4 \quad (5)$$

Formulasi AIC untuk membandingkan model yang satu dengan model lainnya dapat dirumuskan sebagai berikut:

$$AIC_c = AIC + \frac{2(p+2)(p+3)}{n-p-3} \quad (6)$$

Metode PRESS atau *Predicted Residual Error Sum of Squares* adalah bentuk validasi silang untuk memberikan ukuran kecocokan model dengan sampel pengamatan yang tidak digunakan dalam perkiraan model. Model terbaik ditentukan dengan nilai PRESS terkecil.

5. Melakukan pemeriksaan data dengan mendeteksi pencilan (*outlier*), titik *leverage*, dan amatan berpengaruh.

- a. Pencilan.

Outlier atau yang sering disebut dengan pencilan merupakan pengamatan dengan nilai mutlak sisaan yang besar. Suatu pencilan dapat dideteksi menggunakan rumus berikut:

$$r_i = \frac{e_i}{s\sqrt{1-h_{ii}}} \quad (7)$$

keterangan:

e_i = sisaan amatan ke-i

s = standar eror sisaan

h_{ii} = unsur diagonal ke-i matriks $H = X(X'X)^{-1}X'$

Keputusan: jika nilai $|r_i| > 2$, maka amatan tersebut merupakan pencilan.

- b. Titik *leverage*.

Titik *leverage* merupakan titik amatan yang nilai x-nya memiliki jarak yang jauh dari nilai-nilai x lainnya. Suatu amatan memiliki nilai *leverage* tinggi jika memiliki nilai peubah prediktor x “ekstrim”. Titik *leverage* dapat dideteksi menggunakan rumus berikut:

$$h_{ii} = x_i'(X'X)^{-1}x_i \quad (8)$$

keterangan:

h_{ii} = unsur diagonal ke-i pada matriks $H = X(X'X)^{-1}X'$

x_i = vektor baris ke-i pada matriks X

Keputusan: jika nilai $h_{ii} > \frac{2p}{n}$, maka amatan tersebut merupakan titik leverage dengan p banyaknya parameter dan n banyaknya observasi.

c. Amatan berpengaruh.

Amatan berpengaruh berhubungan dengan besarnya perubahan yang terjadi pada dugaan parameter regresi jika pengamatan tersebut dihilangkan. Adapun ukuran untuk menentukan amatan berpengaruh diantaranya adalah dengan menggunakan jarak COOK.

$$D_i = \frac{e_i^2}{ps^2} \left[\frac{h_{ii}}{(1-h_{ii})} \right]^2 \quad (9)$$

keterangan:

e_i^2 = sisaan pada amatan ke-i kuadrat

s^2 = dugaan bagi ragam $Y_i = KT_{sisaan}$

h_{ii} = unsur diagonal ke-i matriks $H = X(X'X)^{-1}X'$

p = banyaknya parameter

Keputusan: jika nilai $D_i > F_{(p, n-p; 1-\alpha)}$, maka amatan tersebut adalah amatan berpengaruh.

6. Estimasi model regresi *Robust Estimasi-M*.

Menurut Erdiani et al. (2021), *Robust regression* merupakan metode yang digunakan untuk menganalisis sebuah data yang tercemari dengan pencilan (*outlier*). Pencilan dapat diatasi dengan regresi *Robust Estimasi-M* yang menggunakan fungsi *Huber*. Estimasi pada tipe maksimum *likelihood* memiliki variansi minimum dan estimator tak bias yang disebut sebagai Estimasi-M (Nurbaroqah et al. 2022). Menurut Atamia et al. (2021), Estimasi-M merupakan bagian dari regresi *Robust* dengan meminimumkan fungsi ρ sebagai berikut:

$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^n \rho \left(y_i - \sum_{j=0}^k x_{ij} \beta_j \right)$	(10)
---	------

Fungsi pembobot *Huber* adalah sebagai berikut:

$$w(u_i) = \begin{cases} 1, & |u_i| \leq k \frac{k}{|u_i|} \\ |u_i| > k \end{cases} \quad (11)$$

$$\hat{\sigma} = \frac{\text{median}|e_i - \text{median}(e_i)|}{0,6745} \quad (12)$$

$$u_i = \frac{e_i}{\hat{\sigma}} \quad (13)$$

Koefisien regresi *Robust* Estimasi-M diduga dengan persamaan berikut:

$$\hat{\beta}_m = (X'WX)^{-1}X'WY \quad (14)$$

dengan W adalah matriks pembobot dan w_i sebagai unsur diagonalnya.

7. Pengujian asumsi Gauss-Markov.

Pendekatan statistik dalam melakukan analisis regresi harus memenuhi beberapa asumsi untuk mengidentifikasi apakah data normal atau tidak. Adapun asumsi-asumsi yang harus dipenuhi adalah sebagai berikut:

a. Nilai harapan sisaan sama dengan nol.

Pengujian asumsi ini dilakukan untuk mengetahui penduga parameter bias atau dengan menggunakan uji-t. Statistik uji-t adalah sebagai berikut:

$$t_0 = \frac{r - \mu}{(SE(r) / \sqrt{n})} \quad (15)$$

keterangan:

r = rata-rata sisaan

μ = nilai hipotesis nol yang ingin diuji

$SE(r)$ = standar deviasi dari r

n = jumlah observasi

Hipotesis:

H_0 : nilai harapan sisaan sama dengan nol [$E(\varepsilon) = 0$].

H_1 : nilai harapan sisaan tidak sama dengan nol [$E(\varepsilon) \neq 0$].

b. Sisaan saling bebas.

Pengujian asumsi ini dilakukan untuk mengetahui keberadaan autokorelasi antar amatan. Uji statistik yang digunakan adalah uji *runs test of randomness*. Jika tidak ada autokorelasi, maka sisaan akan terdistribusi secara acak.

Hipotesis:

H_0 : sisaan saling bebas [$E(\varepsilon_i, \varepsilon_j) = 0$].

H_1 : sisaan tidak saling bebas [$E(\varepsilon_i, \varepsilon_j) \neq 0$].

Statistik uji *runs test* adalah sebagai berikut:

$Z = \frac{R - E(R)}{S_R}$	(16)
$E(R) = \frac{2n_1n_2}{n} + 1$	(17)
$S_R = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n+1)}$	(18)

keterangan:

n = jumlah observasi

n_1 = jumlah observasi sisaan bernilai positif

n_2 = jumlah observasi sisaan bernilai negatif

R = jumlah run

Keputusan: jika nilai $|Z| < Z_{(1-\alpha)/2}$ atau $p\text{-value} > \text{taraf nyata}$, maka tak tolak H_0 .

c. Ragam sisaan homogen.

Pengujian asumsi ini dilakukan untuk mengetahui model menghasilkan ragam yang minimum atau tidak. Uji statistik yang digunakan adalah uji *Non-Constant Variance Test*. Uji ini lebih sering dikenal dengan uji *Studentized Breusch-Pagan*.

Hipotesis:

H_0 : ragam sisaan homogen [$Var(\varepsilon) = 0$].

H_1 : ragam sisaan tidak homogen [$Var(\varepsilon) \neq 0$].

$$BP = \left(\frac{1}{2}\right) f^T (Z^T Z)^{-1} Z^T f \quad (19)$$

keterangan:

n = jumlah observasi

f = vektor dari $\frac{e_i^2}{\alpha^2} - 1$ dengan e_i adalah sisaan untuk pengamatan ke- i

Z = matriks berukuran $n(p + 1)$ dengan unsur vektor yang dinormalkan.

Jika $BP < X_p^2$ atau $p\text{-value} > \text{taraf nyata}$, maka H_0 diterima,

d. Asumsi normalitas sisaan.

Uji statistik yang dapat digunakan untuk meninjau normalitas sisaan adalah uji normalitas *Kolmogorov-Smirnov*. Uji tersebut digunakan untuk menguji *goodness of fit* dari suatu set data dengan distribusi teoritis.

Hipotesis:

H_0 : sisaan menyebar normal.

H_1 : sisaan tidak menyebar normal.

Statistik uji untuk *Kolmogorov-Smirnov* adalah sebagai berikut:

$$D = \max |F_0(x) - S_N(x)| \quad (20)$$

keterangan:

$F_0(x)$ = fungsi distribusi frekuensi kumulatif yang ditentukan

$S_N(x)$ = distribusi frekuensi kumulatif

N = jumlah observasi

8. Uji signifikansi parameter dengan uji simultan (uji F) dan uji parsial (uji t).

Uji simultan (uji F) dapat digunakan untuk melihat pengaruh peubah prediktor terhadap peubah respon secara keseluruhan (Habibah *et al.* 2023). Hipotesis yang digunakan adalah sebagai berikut:

$$H_0 : \beta_2 = \beta_6 = \beta_8 = 0.$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0 \text{ untuk } j = 2, 6, 8.$$

Keputusan untuk menolak H_0 dalam uji simultan diambil jika nilai $p\text{-value} < \alpha$ (taraf nyata) dan $F_{hitung} > F_{tabel}$.

Uji parsial (uji t) berguna meninjau pengaruh masing-masing peubah prediktor terhadap peubah respon (Habibah *et al.* 2023). Hipotesis yang digunakan adalah sebagai berikut:

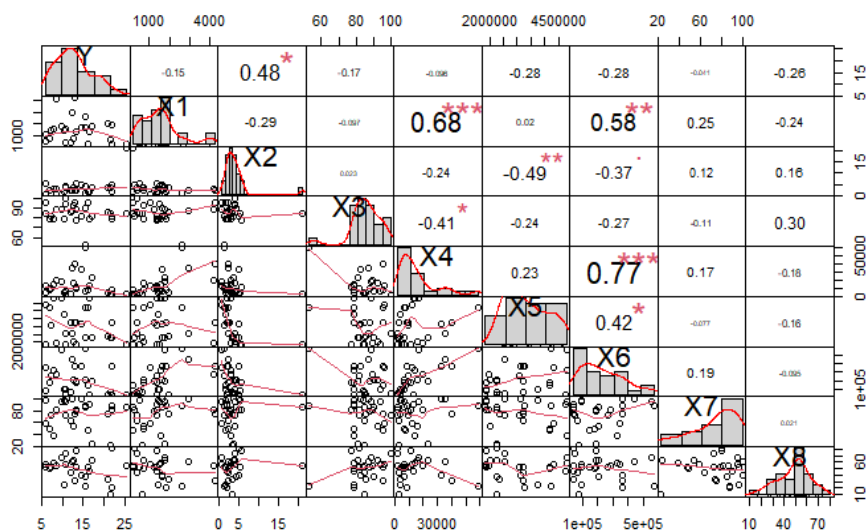
$$H_0 : \beta_j = 0.$$

$$H_1 : \beta_j \neq 0.$$

Keputusan untuk menolak H_0 dalam uji parsial diambil jika nilai $p\text{-value} < \alpha$ (taraf nyata).

3. Hasil dan Pembahasan

3.1 Eksplorasi Data



Gambar 1: Korelasi antar peubah, histogram, dan *scatter plot*.

Gambar 1 menunjukkan kekuatan hubungan antara peubah respon (Y) terhadap peubah prediktor (X), hubungan antar peubah prediktor, dan sebaran data masing-masing peubah. Gambar 1 juga menunjukkan bahwa terdapat peubah respon dan peubah prediktor yang memiliki korelasi positif maupun korelasi negatif dengan peubah lainnya. Peubah prediktor X_4 (Bayi penerima ASI eksklusif) memiliki

korelasi linear positif dengan peubah prediktor X_6 (Perilaku hidup bersih dan sehat pada rumah tangga) dan menghasilkan nilai korelasi paling tinggi, yaitu 0,77. Sebaran nilai data yang mendekati sebaran normal dimiliki oleh peubah respon, peubah prediktor X_2 (Berat badan lahir rendah), dan peubah prediktor X_8 (Tempat pengelolaan makanan memenuhi syarat kesehatan).

3.2 Pendeteksian Multikolinearitas

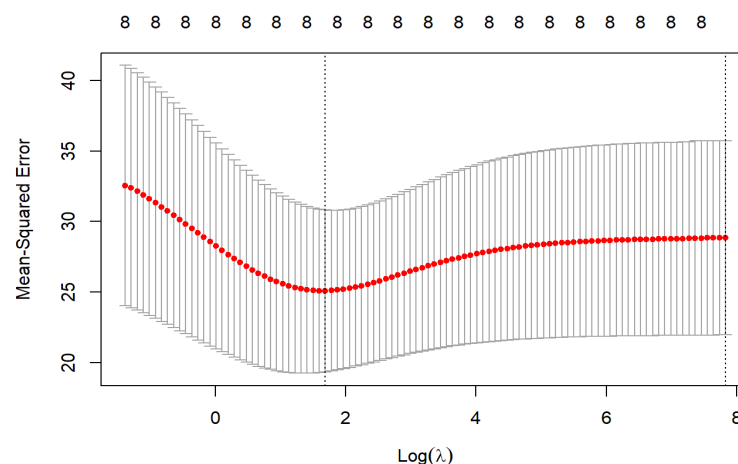
Tabel 2: Nilai VIF peubah prediktor.

Peubah	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	2,581	1,583	1,500	3,735	1,773	3,217	1,163	1,247

Tabel 2 menunjukkan bahwa nilai $VIF < 10$ untuk setiap peubah prediktor, artinya tidak terdapat multikolinearitas. Multikolinearitas juga dapat ditinjau dari plot korelasi antar peubah prediktor pada gambar 1. Korelasi antar peubah prediktor yang cukup kuat terdeteksi di antara bayi penerima ASI eksklusif (X_4) dan perilaku hidup bersih dan sehat pada rumah tangga (X_6) sebesar 0,77 atau 77%. Hal tersebut dapat menjadi indikasi adanya multikolinearitas pada penelitian ini.

3.3 Penanganan Multikolinearitas

3.3.1 Regresi Ridge (Gulud)

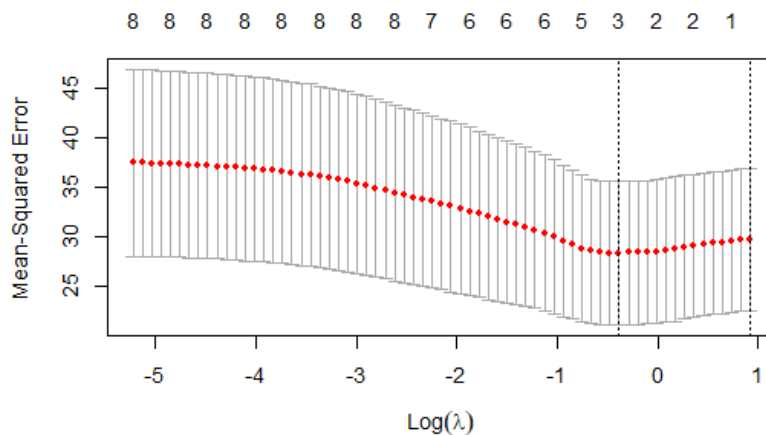


Gambar 2: Hasil pemilihan parameter lambda (regresi Gulud).

Nilai dugaan lambda (tetapan bias) terbaik yang digunakan dalam regresi LASSO adalah 8.599. Pemilihan nilai lambda tersebut didasarkan pada nilai *Mean-Squared Error* (MSE) terkecil menggunakan *software* R berdasarkan grafik pada gambar 2. *R-square* yang dihasilkan dari regresi *Ridge* adalah 25,887%. Model regresi yang terbentuk dari metode regresi gulud adalah sebagai berikut:

$$\hat{Y} = 23.18 - 0.0009895X_1 + 0.334792X_2 - 0.05930123X_3 + 0.000003906609X_4 - 0.0000005532127X_5 - 0.000003252158X_6 - 0.007847244X_7 - 0.05248241X_8$$

3.3.2 Regresi LASSO



Gambar 3: Hasil pemilihan parameter lambda (regresi LASSO).

Nilai dugaan lambda (tetapan bias) terbaik yang digunakan dalam regresi LASSO adalah 0.682. Pemilihan nilai lambda tersebut didasarkan pada nilai *Mean-Squared Error* (MSE) terkecil menggunakan *software* R berdasarkan grafik pada gambar 3. *R-square* yang dihasilkan dari regresi LASSO adalah 31,363%. Model regresi yang terbentuk dari metode regresi LASSO adalah sebagai berikut:

$$\hat{Y} = 17,410 + 0,698X_2 - 0,000005X_6 - 0,109X_8$$

3.4 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan membandingkan *akaike information criterion* (AIC) dan *prediction residual sum of square* (PRESS) dari regresi Gulud (*Ridge*) dan regresi LASSO.

Tabel 3: Perbandingan AIC dan PRESS dari regresi Gulud dan LASSO.

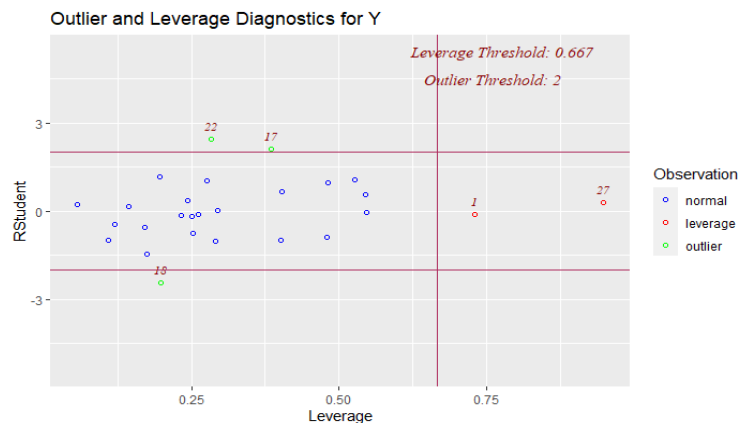
Regresi Gulud (<i>Ridge</i>)		Regresi LASSO	
AIC	PRESS	AIC	PRESS
164.9339	971.6494	164.9316	635.8383

Model terbaik dipilih berdasarkan nilai AIC dan PRESS terkecil. Oleh karena itu, model terbaik yang dipilih adalah model dari analisis regresi LASSO.

3.5 Pemeriksaan Pencilan, Titik *Leverage*, dan Amatan Berpengaruh

Tabel 4: Hasil analisis diagnostik pencilan dan titik *leverage*.

Amatan ke-	h_{ii}	r_i	D_i	Keputusan
1	0,729	-0,106	0,003	Titik <i>leverage</i>
18	0,198	2,155	0,203	Pencilan
22	0,282	2,155	0,203	Pencilan
27	0,949	0,293	0,1778	Titik <i>leverage</i>

Gambar 4: Grafik diagnostik pencilan dan titik *leverage*.

Tabel 4 menunjukkan bahwa terdapat amatan pencilan dan titik *leverage* pada data yang berpotensi menjadi amatan berpengaruh. Nilai $|r_{18}|$ dan $|r_{22}| > 2$, maka amatan ke-18 (Kabupaten Pangandaran) dan amatan ke-22 (Kota Cirebon) merupakan amatan pencilan. Lalu amatan ke-1 (Kabupaten Bogor) dan amatan ke-27 (Kota Banjar) memiliki nilai $h_{ii} > \frac{2(4)}{27} = 0,296$, maka kedua amatan tersebut adalah titik *leverage*. Sementara itu, keempat amatan dalam tabel 4 memiliki nilai $D_i < F_{(4;27-4;1-0,05)} = 2,456$, maka amatan-amatan tersebut bukan amatan berpengaruh. Amatan-amatan dalam tabel 4 tidak memberikan pengaruh terhadap model regresi.

3.6 Penanganan Pencilan, Titik *Leverage*, dan Amatan Berpengaruh.

Keberadaan pencilan dapat ditangani menggunakan regresi *Robust*. Nilai koefisien (k) yang digunakan dalam regresi *Robust* dengan metode pembobot *Huber* adalah 1,345. Langkah pertama adalah menghitung nilai residu (e_i) awal dari pemodelan LASSO sehingga pembobot (w_i) dapat dihitung. Model regresi yang sudah kekar terhadap metode pembobot *Huber* adalah iterasi ke-26.

Tabel 5: Pendugaan koefisien regresi *Robust* Estimasi-M dengan Pembobot *Huber*.

Peubah	Koefisien	p-value
Intersep	17,590	0,000
X_2	1,211	0,088.
X_6	- 0,0000003	0,651
X_6	- 0,152	0,006*
<i>R-square</i>		0.422
<i>p-value</i>		0.014*

Keterangan: *) = signifikan pada taraf nyata 5%, .) = signifikan pada taraf nyata 10%

Berdasarkan tabel 5 diperoleh model regresi yang sudah kekar dari pencilan dengan metode Estimasi-M pembobot *Huber* sebagai berikut:

$$\hat{Y} = 17,590 + 1,211X_2 - 0,0000003X_6 - 0,152X_8$$

3.7 Uji Asumsi

Tabel 6: Hasil uji asumsi Gauss-Markov.

Asumsi	Uji	p-value	Keputusan
$E(\epsilon) = 0$	<i>t-test</i>	1	Tak Tolak H_0
Non-autokorelasi	<i>runs test</i>	0,823	Tak Tolak H_0
Homoskedastisitas	<i>NCV test</i>	0.980	Tak Tolak H_0
<i>Normality</i>	<i>Kolmogorov-Smirnov test</i>	0.999	Tak Tolak H_0

Berdasarkan tabel 6, hasil uji asumsi menunjukkan bahwa semua p-value lebih dari taraf nyata 5%, maka tak tolak H_0 . Hal tersebut menandakan bahwa semua asumsi yaitu sisaan sama dengan nol, tidak ada autokorelasi, sisaan homogen, dan sisaan menyebar normal terpenuhi.

3.8 Uji Signifikansi Parameter

Berdasarkan hasil pada tabel 5, diketahui bahwa p-value $< \alpha = 0.05$, maka tolak H_0 . Kesimpulan yang dapat diambil yaitu semua peubah prediktor signifikan secara keseluruhan terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021. Sementara itu, uji parsial menunjukkan bahwa berat badan lahir rendah (X_2) dan tempat pengelolaan makanan memenuhi syarat Kesehatan (X_8) signifikan terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021. Sementara itu, rumah tangga berperilaku bersih dan sehat (X_6) tidak signifikan terhadap persentase bayi

gizi buruk di Jawa Barat pada tahun 2021.

4. Simpulan dan Saran

4.1 Simpulan

Berdasarkan hasil penelitian, indikasi adanya multikolinearitas yang terdeteksi melalui plot korelasi antar peubah prediktor dapat dikendalikan oleh regresi LASSO. Adapun amatan pencilan dapat dikendalikan oleh regresi *Robust* Estimasi-M dengan pembobot *Huber*. Model akhir yang terbentuk setelah menghilangkan multikolinearitas dan pencilan memiliki *R-square* 42,20%, artinya peubah prediktor dalam penelitian ini dapat menjelaskan variabilitas presentase bayi gizi buruk di Jawa Barat pada tahun 2021 sebesar 42,20% dan sebesar 57,8% dijelaskan oleh peubah prediktor lain yang tidak dimasukkan ke dalam model regresi.

Uji signifikansi parameter dengan uji t dan uji F menyatakan bahwa berat badan lahir rendah (BBLR), perilaku hidup bersih dan sehat pada rumah tangga, serta tempat pengelolaan makanan (TPM) memenuhi syarat kesehatan berpengaruh signifikan secara keseluruhan terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021, sedangkan berat badan lahir rendah (BBLR) dan tempat pengelolaan makanan (TPM) memenuhi syarat kesehatan berpengaruh signifikan secara parsial terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021. Berat badan lahir rendah (BBLR) berpengaruh linear positif terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021, sedangkan perilaku hidup bersih dan sehat pada rumah tangga serta tempat pengelolaan makanan (TPM) memenuhi syarat kesehatan berpengaruh linear negatif terhadap persentase bayi gizi buruk di Jawa Barat pada tahun 2021.

4.2 Saran

Saran yang dapat disampaikan oleh peneliti untuk penelitian lebih lanjut adalah pemilihan dugaan peubah prediktor yang lebih relevan dan berhubungan. Selain itu, pemakaian analisis model regresi lain untuk dibandingkan menjadi model terbaik perlu dilakukan dalam menangani multikolinearitas dan pencilan.

Daftar Pustaka

- Arda D, Lalla N, Suprpto S, Agustang A. 2023. Analysis of the Effect of Malnutrition Status on Toddlers. *Jurnal Ilmiah Kesehatan Sandi Husada*. 12(1): 111-116. doi:10.35816/jiskh.v12i1.910.
- Anggraini BSK, Dzikrullah AA. 2024. Implementasi Analisis Clustering K-Medoids dalam Pengelompokan Bayi Lahir, Gizi Buruk, dan BBLR Berdasarkan Kecamatan di Kabupaten Sleman Tahun 2020. *Emerging Statistics and Data Science Journal*. 2(1): 30-40.

- Ariesthi KD, Esem O, Fitri HN. 2020. Pengaruh Sumber Air Minum dan Sanitasi Lingkungan terhadap Kejadian Gizi Kurang pada Balita di Kabupaten Kupang. *CHM-K Applied Scientifics Journal*. 3(3): 76-80.
- Atamia NA, Susanti Y, Haandajani SS. Perbandingan Analisis Regresi Robust Estimasi-S dan Estimasi-M dengan Pembobot Huber dalam Mengatasi Outlier. Di dalam:[editor tidak diketahui]. Prosiding Seminar Nasional Matematika. 4(2021):673-679.
- Azizah RJ, Wachidah L. 2022. Regresi Robust Estimasi-M dengan Pembobot Huber dan Tukey Bisquare pada Data Tingkat Pengangguran di Indonesia Menurut Provinsi Tahun 2020. *Bandung Conference Series: Statistics*. 2(2): 18-26. doi: 10.293/bess.v2i2.2648.
- Diagama W, Amir Y, Hasneli Y. 2019. Hubungan Jumlah Kunjungan Posyandu dengan Status Gizi Balita (1-5 Tahun). *Jurnal Ners Indonesia*. 9(2): 97-108.
- Dinas Kesehatan [Diskes] Provinsi Jawa Barat. 2023. *Rakor Evaluasi Terpadu, Salah Satu Upaya Percepatan Penurunan Stunting*. [diakses 21 Mei 2024]. <https://diskes.jabarprov.go.id/informasipublik>.
- Edriani TS, Rahmadani A, Noor DMM. 2021. Analisa Hubungan Kepadatan Penduduk dengan Pola Penyebaran COVID-19 Provinsi DKI Jakarta menggunakan Regresi Robust. *Indonesian Journal of Applied Mathematics*. 1(2):51-60.
- Habibah A, Farid FM, Annisa S. 2023. Metode Regresi Gulud untuk Mengatasi Masalah Multikolinearitas pada Kasus Indeks Kualitas Lingkungan Hidup di Indonesia Tahun 2021. *Journal of Statistics and Its Application*. 2(1): 1-12.
- Hawari IF, Najib MK, Nurdianti S, Marpaung YFY, Kusumawati N, Nurfadila M, Sijabat KR, Hernawan BF. 2024. Pengaruh Teknik *Oversampling* pada Algoritma *Machine Learning* dalam Klasifikasi *Body Mass Index* (BMI). *Jurnal Riset dan Aplikasi Matematika (JRAM)*. 8(1):51-68.
- Husain A dan Jamaluddin SRW. 2023. Pemodelan Data Angka Kematian Bayi Menggunakan Regresi Robust. *SAINTEK: Jurnal Sains, Teknologi & Komputer*. 1(1) 1-7.
- Irmah NF, Nurislamyati, Febriyanti M. 2021. Analisis Situasi Gizi Buruk pada Bayi dan Balita di Kecamatan Woha Kabupaten Bima Nusa Tenggara Barat. *Bima Nursing Journal*. 2(2): 154-162.
- Kementerian Kesehatan [Kemenkes]. 2023. *Gizi Buruk*. [diakses 20 Mei 2024]. <https://ayosehat.kemkes.go.id/topik-penyakit/defisiensi-nutrisi/gizi-buruk>.
- Nurbaroqah A, Pratikno B, Supriyanto. 2022. Pendekatan Regresi Robust dengan Fungsi Pembobot Bisquare Tukey pada Estimasi-M dan Estimasi-S. *Jurnal Ilmiah Matematika dan Pendidikan Matematika (JMP)*. 14(1):19-30.
- Nurmaliza, Herlina S. 2019. Hubungan Pengetahuan Pendidikan Ibu terhadap Status Gizi Balita. *Jurnal Kesmas Asclepius*. 1(2): 106-115.
- Sehangunaung GA, Mandey SL, Roring F. 2023. Analisis Pengaruh Harga, Promosi dan Kualitas Pelayanan terhadap Kepuasan Konsumen Pengguna Aplikasi

- Lazada di Kota Manado. *Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*. 11(3):1-11.
- Siregar AL, Sari RF, Widyasari R. 2024. Penerapan Analisis Lasso dalam Menentukan Penyebab Utama Faktor Kematian Pasien Komorbid yang Terinfeksi Covid-19. *Jurnal Penelitian Matematika dan Pendidikan Matematika*. 7(1): 290-296.
- Triveni, Maulani RG, Andolina N. 2023. Hygiene dan Sanitasi terhadap Kejadian Wasting pada Bayi Usia 0-59 Bulan. *Pro Health Jurnal Ilmiah Kesehatan*. 5(1): 320-323.
- Ummusalma C dan Syafridah A. 2019. Hubungan Perilaku Hidup Bersih dan Sehat dengan Status Gizi Balita di Puskesmas Baktiya Barat Tahun 2019. *Jurnal Ilmiah Sains, Teknologi, Ekonomi, Sosial dan Budaya*. 6(2): 53-53.
- Usada NK, Wanodya KS, Trisna N. 2021. Analisis Spasial Gizi Kurang Balita di Kota Tangerang Tahun 2019. *Jurnal Biostatistik, Kependudukan, dan Infromatika Kesehatan*. 2(1): 1-15.