

Perbandingan Regresi Ridge dan LASSO dalam Pemilihan Variabel dan Prediksi Keluarga Berisiko Stunting di Sumatera Barat

Berliana Nofriadi¹, Syafriandi², Yenni Kurniawati³

^{1,2,3}Universitas Negeri Padang

¹berliananoфриadi@gmail.com

ABSTRAK

Stunting merupakan isu kesehatan masyarakat kronis yang timbul dari kekurangan gizi jangka panjang, terutama pada anak-anak, yang berdampak serius pada pertumbuhan fisik dan perkembangan kognitif. Penelitian ini bertujuan menganalisis secara komprehensif faktor-faktor yang memengaruhi Jumlah Keluarga Berisiko Stunting (KRS) di Provinsi Sumatera Barat tahun 2024 serta membandingkan kinerja model Regresi Ridge dan Regresi LASSO. Digunakan data sekunder (count data) dari Sistem Informasi Keluarga (SIGA) BKKBN yang dikelompokkan berdasarkan wilayah administratif. Analisis awal menggunakan Regresi Linier Ordinary Least Square (OLS) mengonfirmasi adanya masalah multikolinearitas tinggi, di mana empat variabel prediktor melampaui ambang batas $VIF > 10$. Untuk mengatasi ketidakstabilan ini, diterapkan metode Regularisasi. Optimasi parameter penyusutan (λ) dilakukan melalui *k-fold Cross Validation*, dengan akurasi model dievaluasi menggunakan Mean Squared Error (MSE). Hasil menunjukkan bahwa Regresi LASSO adalah model yang superior dengan MSE terendah ($1,988 \times 10^{-5}$) dibandingkan Regresi Ridge ($0,0001243$). LASSO berhasil menstabilkan estimasi dan melakukan seleksi fitur, mengeliminasi variabel PUS terlalu muda dan PUS terlalu banyak. Model ini mengidentifikasi lima faktor signifikan pada KRS: akses air minum, jamban layak, PUS terlalu tua, jarak kelahiran terlalu dekat, dan bukan peserta KB modern. Temuan ini memberikan dasar berbasis bukti yang efisien bagi kebijakan intervensi stunting yang terfokus, menunjukkan kapabilitas metode regularisasi dalam mengelola data berdimensi kompleks di bidang kependudukan.

Kata Kunci: Regresi Regularisasi; Regresi LASSO; Multikolinearitas; Data Count; Keluarga Berisiko Stunting.

ABSTRACT

Stunting is a chronic public health issue arising from long-term nutritional deficiency, especially in children, which has a serious impact on physical growth and cognitive development. This study aims to comprehensively analyze the factors influencing the Number of Stunting Risk Families (KRS) in West Sumatra Province in 2024 and to compare the performance of Ridge and LASSO Regression models. Secondary data (count data) from the BKKBN Family Information System (SIGA) categorized by administrative region was utilized. Initial analysis using Ordinary Least Square (OLS) Regression confirmed high multicollinearity, with four predictor variables exceeding the $VIF > 10$ threshold. To solve this statistical issue, the Regularization method was implemented. The optimization of the shrinkage parameter (λ) was performed through *k-fold Cross Validation*, and model accuracy was evaluated using Mean Squared Error (MSE). The results indicate that LASSO Regression is the superior model with the lowest MSE (1.988×10^{-5}) compared to Ridge Regression (0.0001243). LASSO successfully stabilized the estimates and performed feature selection, eliminating the variables PUS too young and PUS too many. The model identified five significant factors for KRS: access to drinking water, decent latrines, PUS too old, birth spacing too close, and non-modern KB participants. This finding provides an efficient, evidence-based foundation for targeted stunting intervention policies, demonstrating the capability of regularization methods in managing complex demographic data.

Keywords: Regularized Regression; LASSO Regression; Multicollinearity; Count Data; Stunting Risk Families.

PENDAHULUAN

Stunting merupakan isu kesehatan masyarakat kronis yang timbul dari kekurangan gizi jangka panjang, terutama pada anak-anak, yang berdampak serius pada pertumbuhan fisik dan perkembangan kognitif (Mengetahui Stunting Dan Gizi Buruk, 2018). Permasalahan stunting masih menjadi tantangan di Indonesia. Berdasarkan data Survei Status Gizi Indonesia (SSGI) tahun 2024, prevalensi stunting nasional berada di angka 19,8%, sementara target penurunan yang ditetapkan dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) hingga tahun 2029 adalah 14,2%. Di Provinsi Sumatera Barat, meskipun terjadi penurunan dari 25,2% pada tahun 2022, angka prevalensi stunting pada tahun 2024 tercatat masih tinggi, yakni 23,6%, sehingga membutuhkan intervensi serius untuk mencapai target nasional 14% (SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%, 2025).

Dalam publikasi TVRI Sumatera Barat (2024), Kepala Perwakilan BKKBN Provinsi Sumatera Barat, Fatmawati mengungkapkan bahwa angka prevalensi stunting di Sumatera Barat mencapai 23,6%, menurun dari 25,2% pada tahun 2022. Meskipun terjadi penurunan, angka ini masih tergolong tinggi dan memerlukan upaya serius untuk mencapai target nasional sebesar 14% pada 2024. Data mengenai Keluarga Berisiko Stunting (KRS) menunjukkan bahwa berbagai faktor, seperti ketersediaan air bersih dan sanitasi, usia ibu saat melahirkan, jarak kelahiran, dan jumlah anak, diduga kuat berkontribusi terhadap tingginya kasus stunting di wilayah tersebut. Untuk menganalisis dan memodelkan pengaruh faktor-faktor risiko ini, analisis regresi digunakan, yang merupakan metode statistik untuk memetakan hubungan variabel prediktor dengan variabel respons.

Model regresi linier standar menggambarkan variabel respons Y sebagai fungsi linier dari beberapa variabel prediktor X_i :

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

Dalam model ini, β mewakili koefisien regresi, dan ε adalah komponen *error* yang tidak dapat dijelaskan oleh prediktor. Pendekatan umum untuk mengestimasi parameter model ini adalah menggunakan metode Kuadrat Terkecil Biasa (*Ordinary Least Square* atau OLS). Estimasi OLS $\hat{\beta}$ diperoleh dengan meminimumkan jumlah kuadrat *error* regresi:

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

Estimasi parameter yang paling umum digunakan dalam analisis regresi adalah metode kuadrat terkecil (*Ordinary Least Square*). *Ordinary Least Square* (OLS) merupakan salah satu cara untuk mengestimasi parameter dengan meminimumkan jumlah dari kuadrat error persamaan regresinya. Rencher dan Schaalje (2008) menyatakan estimasi parameter dari persamaan (1) dengan menggunakan OLS ialah:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (3)$$

Gujarati and Porter (2008) bahwa meskipun model OLS adalah fondasi analisis regresi, ia bergantung pada asumsi kunci, salah satunya adalah tidak adanya multikolinearitas antar variabel prediktor. Lebih lanjut, Rencher dan Schaalje (2008) menyatakan tingginya multikolinearitas mengindikasikan variabel-variabel prediktor berkorelasi erat, suatu kondisi yang berisiko memperbesar variansi parameter dan menurunkan akurasi hasil estimasi. Untuk mengatasi ketidakstabilan ini, pendekatan regularisasi atau *shrinkage* diaplikasikan, yang bekerja dengan menyusutkan koefisien regresi mendekati nol relatif terhadap estimasi OLS (James et al., 2013).

Dua teknik regularisasi yang menonjol adalah Regresi Ridge dan Regresi LASSO. Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO), yang diperkenalkan oleh Tibshirani (1996), meminimumkan jumlah kuadrat *error* dengan menambahkan penalti berdasarkan jumlah nilai mutlak koefisien ($\sum |\beta_j|$):

$$\hat{\beta}(\text{lasso}) = \sum_{i=1}^n \|y_i - \sum_{j=1}^k x_{ij}\beta_j\|^2 + \lambda \sum_{j=1}^k |\beta_j| \quad (4)$$

Sementara itu, Regresi Ridge menggunakan metode estimasi serupa, namun menerapkan penalti berdasarkan jumlah kuadrat koefisien ($\sum \beta_j^2$). Secara matematis estimasi parameter regresi ridge menurut James (2013) dapat dinyatakan:

$$\hat{\beta}(\text{ridge}) = \sum_{i=1}^n (y_i - \sum_{j=1}^k x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^k \beta_j^2 \quad (5)$$

Perbedaan krusial terletak pada efek penalti: Ridge hanya menyusutkan koefisien mendekati nol, sedangkan LASSO memiliki kemampuan unik untuk menyusutkan koefisien secara tepat menjadi nol, yang secara efektif menjalankan fungsi pemilihan variabel (*feature selection*).

Dengan demikian, pemodelan faktor-faktor yang memengaruhi Keluarga Berisiko Stunting di Sumatera Barat memerlukan solusi metodologis yang canggih. Oleh karena itu, studi ini mengaplikasikan metode Regresi Regularisasi, khususnya Ridge dan LASSO, untuk menanggulangi isu multikolinearitas yang terdeteksi dan secara simultan melakukan pemilihan variabel. Tujuan utama penelitian ini adalah mengevaluasi dan membandingkan kinerja prediktif dari Regresi Ridge dan Regresi LASSO dalam mengidentifikasi faktor-faktor penentu yang paling efisien dan signifikan terhadap Jumlah Keluarga Berisiko Stunting (KRS) di Sumatera Barat, guna menyediakan model statistik yang stabil dan dapat ditindaklanjuti untuk kebijakan intervensi stunting.

METODE PENELITIAN

Penelitian ini merupakan penelitian terapan dengan pendekatan kuantitatif analitik. Tujuan utama studi ini adalah menganalisis dan memodelkan faktor-faktor yang memengaruhi Jumlah Keluarga Berisiko Stunting (KRS) di Provinsi Sumatera Barat dengan membandingkan efektivitas Regresi Ridge dan Regresi LASSO. Metode ini dipilih karena kemampuannya mengatasi masalah multikolinearitas yang tinggi pada data dan melakukan seleksi fitur secara otomatis, sehingga dapat menghasilkan model prediksi yang stabil dan efisien.

Jenis data yang dimanfaatkan dalam studi kasus ini adalah data sekunder. Data ini bersumber dari publikasi Sistem Informasi Keluarga (SIGA) BKKBN mengenai jumlah KRS pada tahun 2024, dikelompokkan berdasarkan wilayah administratif (kabupaten/kota) di Provinsi Sumatera Barat. Populasi penelitian adalah seluruh keluarga di Provinsi Sumatera Barat sebanyak 1.327.164 keluarga, dan sampel yang digunakan adalah 222.502 keluarga yang tergolong berisiko stunting (BKKBN, 2024), dikelompokkan berdasarkan kabupaten/kota untuk analisis.

Studi ini menggunakan satu variabel dependen, yaitu Jumlah Keluarga Berisiko Stunting Y, dan tujuh variabel independen (X_1 sampai X_7) yang merupakan indikator penapisan KRS. Tabel 1. menampilkan variabel yang digunakan dalam studi kasus.

Tabel 1. Variabel Studi Kasus

Variabel	Variabel	Satuan	Skala
Y	Keluarga berisiko stunting	Jumlah Keluarga	Rasio
X_1	Keluarga tidak mempunyai sumber air minum utama yang layak	Jumlah Keluarga	Rasio
X_2	Keluarga tidak mempunyai jamban yang layak	Jumlah Keluarga	Rasio
X_3	PUS terlalu muda (umur istri di Bawah 20 tahun)	Jumlah Keluarga	Rasio
X_4	PUS terlalu tua (umur istri 35 sampai 40 tahun)	Jumlah Keluarga	Rasio
X_5	PUS terlalu dekat (kurang 2 tahun jarak kelahiran)	Jumlah Keluarga	Rasio
X_6	PUS terlalu banyak (lebih sama 3 anak)	Jumlah Keluarga	Rasio
X_7	Bukan peserta KB modern	Jumlah Keluarga	Rasio

Teknik analisis yang diterapkan adalah Regresi Regularisasi (Ridge dan LASSO) menggunakan perangkat lunak R Studio. Tahapan analisis data difokuskan pada pengujian asumsi, penentuan parameter regularisasi (λ), dan perbandingan kinerja model, sebagai berikut:

1. Membentuk model regresi linier berganda menggunakan *Ordinary Least Square* (OLS) untuk menetapkan model dasar dan mengidentifikasi masalah multikolinearitas menggunakan persamaan (2)
2. Melakukan pengujian multikolinearitas antar variabel independen. Untuk mendeteksi multikolinearitas dalam data, digunakan nilai VIF (*Variance Inflation Factor*)¹. VIF dihitung menggunakan formula berikut:

$$VIF_i = \frac{1}{1-R_j^2} \quad (6)$$

Di mana R_j^2 merepresentasikan nilai koefisien determinasi. Berdasarkan literatur, multikolinearitas diidentifikasi apabila nilai VIF suatu variabel prediktor melebihi 10 (Andana et al., 2017).

3. Menentukan model regresi ridge dan LASSO dengan:
 - a. Estimasi parameter menggunakan persamaan (4) untuk model ridge dan persamaan (3) untuk model LASSO
 - b. Menentukan λ terbaik. Penentuan parameter penyusutan optimal (λ) untuk kedua model Ridge dan LASSO, dilakukan melalui metode Cross Validation (CV). Salah satu bentuk CV yang digunakan adalah k -fold Cross Validation (Efron & Tibshirani, 1993). Dalam prosedur ini, data dibagi menjadi k subset (*fold*). Model dilatih menggunakan $k-1$ fold data dan divalidasi pada fold sisanya, proses ini diulang hingga setiap fold digunakan sebagai data validasi. Hasilnya dihitung sebagai rata-rata dari MSE setiap fold:

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

Proses ini bertujuan menemukan nilai λ yang menghasilkan MSE terendah, yang kemudian digunakan untuk membentuk model Regresi Ridge dan LASSO terbaik.

- c. Membentuk model regresi ridge dan regresi LASSO final berdasarkan λ terbaik
4. Menentukan model terbaik dengan membandingkan nilai Mean Square Error (MSE) model regresi ridge dan regresi LASSO.

Akurasi model diukur menggunakan MSE (*Mean Square Error*), sebuah metrik yang mengukur deviasi antara nilai prediksi dengan nilai yang sebenarnya. Menurut Pham (2019) MSE dihitung dengan rumus:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (8)$$

HASIL DAN PEMBAHASAN

Hasil

Uji Multikolinearitas

Pemodelan awal menggunakan regresi linier *Ordinary Least Square* (OLS) menghasilkan estimasi parameter yang termuat pada Tabel 2.

Tabel 2. Estimasi Parameter Regresi Linier

Variabel	Estimasi Parameter
<i>Intercept</i>	-0,001846
X_1	0,557565
X_2	0,817495
X_3	2,553991
X_4	-0,207418
X_5	2,170386
X_6	0,020715
X_7	0,476882

Untuk mengevaluasi asumsi regresi klasik, dilakukan perhitungan nilai VIF (*Variance Inflation Factor*) guna mendeteksi multikolinearitas yang termuat pada Tabel 3.

Tabel 3. Nilai VIF Tiap Variabel Independen

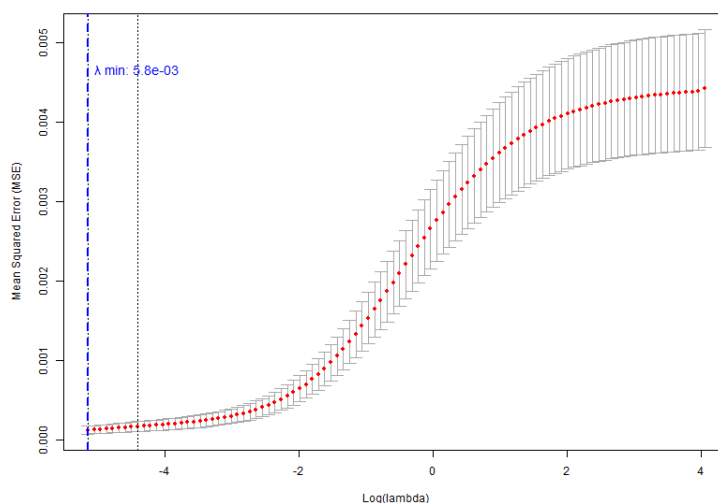
Variabel	Nilai VIF
X_1	3,072981
X_2	6,205869
X_3	7,038267
X_4	35,802176
X_5	28,994987
X_6	147,602681
X_7	17,858912

Hasil pada Tabel 3 menunjukkan bahwa terdapat empat variabel independen yang melampaui batas kriteria multikolinearitas ($VIF > 10$). Variabel tersebut adalah X_4 (PUS terlalu tua), X_5 (PUS terlalu dekat), X_6 (PUS terlalu banyak), dan X_7 (Bukan peserta KB modern). Deteksi multikolinearitas yang signifikan ini menegaskan ketidaklayakan model regresi klasik dan memerlukan adopsi metode yang lebih robust, yaitu regresi Ridge dan regresi LASSO.

Model Regresi Ridge

Langkah krusial dalam Regresi Ridge adalah memilih nilai λ optimal. Melalui proses *k-fold cross validation*, yang ditampilkan pada Gambar 1. Data diaplikasikan pada regresi ridge membangkitkan λ acak sebanyak 100 index. Dari hasil *k-fold cross validation* diperoleh λ

terbaik sebesar 0,005804, yang menghasilkan *Mean Squared Error* (MSE) minimum sebesar 0,0001243.



Gambar 1. Cross Validation pada Regresi Ridge

Koefisien Regresi Ridge pada Tabel 4 memperlihatkan bahwa semua variabel tetap dipertahankan dalam model. Namun, terjadi penyusutan drastis koefisien dibandingkan hasil OLS. Sebagai contoh, koefisien X_3 yang awalnya 2,553991 pada OLS menyusut menjadi hanya 0,00449025 pada model Ridge.

Tabel 4. Koefisien Regresi Ridge

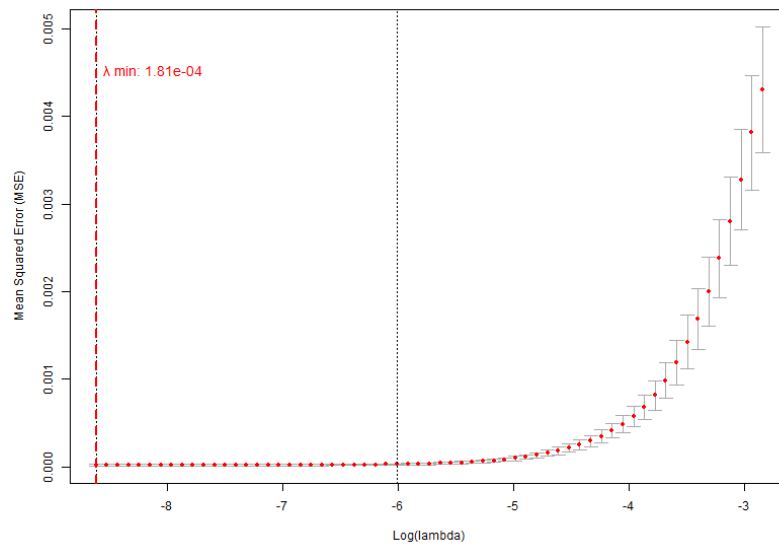
Variabel	Koefesien Regresi Ridge
<i>Intercept</i>	0,091477875
X_1	0,01210078
X_2	0,025267528
X_3	0,00449025
X_4	0,00058933
X_5	0,004475559
X_6	0,008251456
X_7	0,017629079

Model Regresi Ridge yang terbentuk dengan λ optimal adalah:

$$Y = 0,09148 + 0,0121X_1 + 0,02527X_2 + 0,00449X_3 + 0,00059X_4 + 0,00448X_5 + 0,00825X_6 + 0,01763X_7$$

Model Regresi LASSO

Serupa dengan Ridge, penentuan λ optimal sangat penting untuk Regresi LASSO. Data diaplikasikan pada regresi LASSO membangkitkan λ acak sebanyak 63 index. Hasil *k-fold cross validation* pada Gambar 2 menunjukkan λ terbaik sebesar 0,0001814, dengan nilai MSE minimum yang jauh lebih rendah, yaitu 0,00001988.



Gambar 2. Cross Validation pada Regresi LASSO

Koefisien Regresi LASSO pada Tabel 5 menunjukkan kemampuan metode ini dalam seleksi variabel otomatis. Dengan λ terbaik, LASSO hanya mempertahankan lima peubah bebas yaitu X_1, X_2, X_4, X_5, X_7 , sementara variabel X_3 (PUS terlalu muda) dan X_6 (PUS terlalu banyak) dieliminasi dari model dengan koefisien diset menjadi nol karena kontribusinya terhadap variabel dependen dianggap tidak signifikan.

Tabel 5. Koefisien Regresi LASSO pada Variabel

Variabel	Koefisien Regresi LASSO
<i>Intercept</i>	0,091477874
X_1	0,007597453
X_2	0,035716976
X_3	0
X_4	-0,003950041
X_5	0,001999843
X_6	0
X_7	0,031280416

Model LASSO yang dihasilkan adalah:

$$Y = 0,09148 + 0,00759X_1 + 0,03572X_2 - 0,00395X_4 + 0,00199X_5 + 0,03128X_7$$

Pemilihan Model Terbaik

Perbandingan akurasi kedua model didasarkan pada nilai MSE minimum yang dicapai, yang dapat dilihat pada Tabel 6.

Tabel 6. Nilai MSE Tiap Model Regresi

Model	Nilai MSE
Regresi Ridge	$1,243 \times 10^{-4}$
Regresi LASSO	$1,988 \times 10^{-5}$

Tabel 6 menunjukkan bahwa Regresi LASSO memiliki nilai MSE yang lebih kecil $1,988 \times 10^{-5}$ dibandingkan Regresi Ridge $1,243 \times 10^{-4}$. Kriteria pemilihan model terbaik

adalah MSE terkecil, sehingga Regresi LASSO ditetapkan sebagai model yang paling unggul karena secara efektif mengatasi multikolinearitas dan memberikan kemampuan prediksi yang superior.

Pembahasan

Regresi linier berganda (OLS) digunakan sebagai metode klasik, tetapi kelemahannya—yakni ketidakstabilan model akibat multikolinearitas—terdeteksi secara jelas. Hasil uji VIF mengonfirmasi multikolinearitas pada empat variabel yaitu X_4 , X_5 , X_6 , X_7 , yang menjadikan pemodelan OLS tidak valid untuk interpretasi lebih lanjut.

Regresi Ridge berhasil menanggulangi multikolinearitas dengan menyusutkan koefisien regresi, didukung oleh λ terbaik 0,005804. Meskipun terjadi penyusutan koefisien, Ridge mempertahankan semua variabel dalam model. Sebaliknya, Regresi LASSO menunjukkan keunggulan ganda: tidak hanya menyusutkan koefisien, tetapi juga melakukan seleksi variabel. LASSO, dengan λ terbaik 0,0001814, secara spesifik mengeliminasi variabel X_3 (PUS terlalu muda) dan X_6 (PUS terlalu banyak), mengindikasikan bahwa kontribusi kedua faktor tersebut terhadap risiko stunting dinilai tidak signifikan oleh model regularisasi. Dengan demikian, LASSO menghasilkan model yang lebih sederhana dan efisien.

Model Regresi LASSO ditetapkan sebagai model yang paling unggul dan efisien untuk memodelkan faktor-faktor yang memengaruhi Jumlah Keluarga Berisiko Stunting (KRS) di Sumatera Barat. Keunggulan ini ditegaskan oleh nilai Mean Squared Error (MSE) LASSO yang jauh lebih kecil ($1,988 \times 10^{-5}$) dibandingkan MSE Regresi Ridge ($1,243 \times 10^{-4}$). Proses *k-fold cross validation* menghasilkan parameter penyusutan optimal (λ) sebesar 0,0001814 untuk LASSO. Selain menstabilkan estimasi yang sebelumnya terganggu oleh multikolinearitas tinggi pada empat variabel prediktor, metode LASSO berhasil melakukan seleksi variabel otomatis.

Model LASSO final hanya mempertahankan lima dari tujuh variabel bebas. Secara spesifik, variabel X_3 (PUS terlalu muda) dan X_6 (PUS terlalu banyak) dieliminasi dari model dengan koefisien disetel menjadi nol. Hal ini mengindikasikan bahwa kontribusi kedua faktor tersebut terhadap risiko stunting dinilai tidak signifikan oleh model regularisasi, sehingga menghasilkan model yang lebih sederhana dan efisien.

Kelima variabel yang dipertahankan dan dianggap signifikan adalah: Keluarga tidak mempunyai jamban yang layak (X_2), Jarak kelahiran terlalu dekat (X_5), Keluarga tidak mempunyai sumber air minum utama yang layak (X_1), Bukan peserta KB modern (X_7), dan PUS terlalu tua (X_4). Dari koefisien yang diperoleh, faktor Keluarga tidak mempunyai jamban yang layak (X_2) memiliki dampak peningkatan terbesar terhadap Jumlah KRS (koefisien +0,03572), diikuti oleh Bukan peserta KB modern (X_7) (koefisien +0,03128). Hal ini menyoroti pentingnya perbaikan sanitasi lingkungan dan pengaturan jarak kehamilan dalam upaya pencegahan stunting. Variabel PUS terlalu tua (X_4) menjadi satu-satunya variabel signifikan yang memiliki koefisien negatif (−0,00395), menunjukkan pengaruh yang sedikit berlawanan atau kompleks setelah disusutkan oleh regularisasi.

PENUTUP

Analisis perbandingan model regularisasi menunjukkan bahwa Regresi LASSO merupakan model terbaik dan paling efisien untuk memodelkan faktor-faktor yang memengaruhi Jumlah Keluarga Berisiko Stunting (KRS) di Sumatera Barat. Keunggulan ini ditegaskan oleh nilai *Mean Squared Error* (MSE) LASSO sebesar $1,988 \times 10^{-5}$, yang secara signifikan lebih kecil daripada MSE Regresi Ridge $1,243 \times 10^{-4}$. Regresi LASSO berhasil melakukan pemilihan variabel otomatis, mengidentifikasi lima faktor penapisan yang

signifikan: Keluarga tidak mempunyai sumber air minum utama yang layak (X_1), Keluarga tidak mempunyai jamban yang layak (X_2), PUS terlalu tua (X_4), PUS terlalu dekat (X_5), dan Bukan peserta KB modern (X_7). Model ini secara efektif mengeliminasi dua variabel X_3 dan X_6 —yang kontribusinya dianggap tidak signifikan.

Berdasarkan temuan ini, direkomendasikan bahwa program intervensi percepatan penurunan stunting harus difokuskan pada lima faktor signifikan yang dipertahankan oleh model LASSO, khususnya perbaikan sanitasi lingkungan dan peningkatan partisipasi KB modern. Untuk pengembangan studi lebih lanjut, disarankan agar penelitian berikutnya mempertimbangkan perbandingan dengan Regresi Elastic Net atau mengaplikasikan metode regularisasi ini dalam kerangka *Generalized Linear Models* (GLM) untuk menguji konsistensi dan efektivitas metodologi secara lebih luas.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Departemen Statistika, Universitas Negeri Padang atas dukungan fasilitas, bantuan akademik, dan lingkungan ilmiah yang mendukung kelancaran pelaksanaan penelitian ini.

REFERENSI

- Andana, A. P., Safitri, D., & Rusgiyono, A. (2017). Model Regresi Menggunakan Least Absolute Shrinkage And Selection Operator (Lasso) Pada Data Banyaknya Gizi Buruk Kabupaten/Kota Di Jawa Tengah. *Jurnal Gaussian*, 6(1), 22–30. <https://doi.org/-10.14710/j.gauss.6.1.21-30>
- Angka Prevalensi Stunting Di Sumbar Turun Jadi 23,6 Persen (2024). BKKBN. (2024). *Sistem Informasi Keluarga (SIGA)*. BKKBN. <https://siga.bkkbn.go.id/>
- Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. In *Chapman & Hall/CRC* (1st ed.). https://doi.org/10.1007/978-0-387-32833-1_125
- Gujarati, N., D., & Porter, D. C. (2008). *Basic Econometrics* (5th ed.). McGraw-Hill School Education Group.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *Springer Texts in Statistics An Introduction to Statistical Learning - with Applications in R*.
- Mengenal Stunting Dan Gizi Buruk (2018). <https://promkes.kemkes.go.id/?p=8486#:~:text=Permasalahan stunting terjadi mulai dari,obesitas dan kematian akibat infeksi.&text=Situs Adoption Nutrition menyebutkan%2C stunting,di awal kehidupan seorang anak>.
- Pham, H. (2019). A new criterion for model selection. *Mathematics*, 7(12), 1–12. <https://doi.org/10.3390/MATH7121215>
- Rencher, C., A., Schaalje, & Bruce, G. (2008). *Linear Models in Statistics*. Wiley-Interscience.
- SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8% (2025). <https://www.badankebijakan.kemkes.go.id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198/>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/https://id.welib.org/md5/c5f8a4bd73e650d0f573423cc6f1ef83#>