ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 2, Tahun 2026, Halaman 301 - 311
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

ANALISIS KOMPARATIF DAN EVALUASI KINERJA METODE BRIDGE, LASSO, DAN RIDGE REGRESSION DALAM PREDIKSI PERSENTASE LEMAK TUBUH BERBASIS ANTROPOMETRI

Chesa Nabilla Azzahra[1], M. Syarlan[1], A'inun Fauziyah[1], Zahra Hana Andrea[1], Etis Sunandi[1*], Idhia Sriliana[1], Susi Wijuniamurti[1]
[1] Program Studi S1 Statistika, Universitas Bengkulu, Jl. WR. Supratman, Kelurahan Kandang Limun, Kecamatan Muara Bangka Hulu, Kota Bengkulu 38371, Indonesia.
***e-mail**: esunandi@unib.ac.id

DOI: 10.14710/j.gauss.15.2.301-311

Article Info:
Received: 2026-03-16
Accepted: 2026-08-10
Available Online: 2026-08-25

Keywords:
Body Fat Percentage;Bridge Regression;LASSO Regression; Ridge Regression;Penalized Regression

Abstract: This study comparatively evaluates the performance of Bridge, LASSO, and Ridge regression methods in predicting body fat percentage based on anthropometric measurements. Exploratory analysis indicated strong relationships between body fat percentage and abdominal and hip circumferences, with evidence of multicollinearity among the predictors. The ordinary least squares (OLS) model identified significant predictors but violated several classical assumptions, particularly multicollinearity, heteroskedasticity, and autocorrelation. Therefore, penalized regression methods were applied to obtain more stable parameter estimates and improve predictive performance. Model performance was assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), adjusted R^2, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Variance Inflation Factor (VIF). The results show that Bridge regression with $\gamma = 0.5$ achieved the best predictive performance, producing the lowest MSE and AIC values. LASSO regression provided the simplest model by eliminating irrelevant predictors and achieved the lowest BIC and VIF values, while Ridge regression produced stable estimates but did not outperform the other methods. Overall, Bridge regression is recommended for prediction, whereas LASSO is preferable when model simplicity and interpretability are prioritized.

1. PENDAHULUAN

Persentase lemak tubuh merupakan indikator penting dalam penilaian status gizi dan risiko penyakit tidak menular, seperti obesitas dan diabetes mellitus, yang prevalensinya terus meningkat di Indonesia maupun dunia. Pengukuran secara langsung menggunakan Dual-Energy X-Ray Absorptiometry (DXA) atau Bioelectrical Impedance Analysis (BIA) memiliki akurasi tinggi, tetapi memerlukan peralatan khusus dan biaya yang relatif mahal sehingga kurang praktis untuk penelitian berskala besar maupun kegiatan skrining. Oleh karena itu, pendekatan berbasis variabel antropometri, seperti berat badan, tinggi badan, lingkar leher, lingkar perut, dan lingkar pinggul, menjadi alternatif yang lebih sederhana dan ekonomis (Kyle dkk., 2004; Nevill dkk., 2006; Khairani & Sudiarti, 2020).

Regresi linier berganda banyak digunakan untuk memodelkan persentase lemak tubuh berdasarkan variabel antropometri. Khairani dan Sudiarti (2020) menunjukkan bahwa variabel antropometri berhubungan signifikan dengan persentase lemak tubuh dan mampu menghasilkan model prediksi yang baik. Namun, tingginya korelasi antarvariabel antropometri sering menyebabkan multikolinearitas sehingga koefisien regresi menjadi tidak stabil dan kemampuan prediksi model menurun (Kutner dkk., 2005; Montgomery dkk., 2021).

Salah satu pendekatan untuk mengatasi multikolinearitas adalah regresi regularisasi (*penalized regression*), seperti Ridge Regression, LASSO Regression, dan Bridge

Regression. Ketiga metode tersebut menggunakan fungsi penalti untuk menghasilkan estimasi parameter yang lebih stabil pada data yang mengalami multikolinearitas (Hoerl & Kennard, 1970; Tibshirani, 1996; Frank & Friedman, 1993; Hastie dkk., 2009; James dkk., 2021; Bühlmann & van de Geer, 2011). Penerapan metode regularisasi menjadi penting pada data antropometri karena hubungan yang kuat antarvariabel pengukuran tubuh berpotensi menimbulkan multikolinearitas, sehingga dapat menurunkan kestabilan model regresi dan akurasi prediksi.

Meskipun berbagai penelitian telah menerapkan metode regresi regularisasi, penelitian mengenai prediksi persentase lemak tubuh berbasis variabel antropometri masih memiliki keterbatasan. Khairani dan Sudiarti (2020) memodelkan persentase lemak tubuh menggunakan regresi linier berganda berbasis variabel antropometri tanpa menerapkan metode regularisasi sehingga potensi multikolinearitas belum dievaluasi secara khusus. Di sisi lain, Pratiwi dan Sari (2017) serta Enwere dkk. (2023) membandingkan beberapa metode regularisasi untuk mengatasi multikolinearitas, namun tidak diterapkan pada kasus prediksi persentase lemak tubuh berbasis data antropometri. Sementara itu, Andari dkk. (2023) menggunakan algoritma Bayesian Ridge untuk prediksi body fat, tetapi belum membandingkan performanya dengan metode regularisasi lain dalam kerangka evaluasi yang sama.

Berdasarkan kondisi tersebut, kebaruan penelitian ini terletak pada penerapan dan perbandingan langsung Ridge Regression, LASSO Regression, dan Bridge Regression pada data antropometri untuk prediksi persentase lemak tubuh. Selain itu, seluruh model dievaluasi menggunakan prosedur cross-validation yang sama sehingga perbandingan kinerja dapat dilakukan secara lebih konsisten dan objektif. Dengan demikian, penelitian ini tidak hanya mengevaluasi kemampuan prediksi masing-masing metode, tetapi juga membandingkan efektivitasnya dalam menangani multikolinearitas pada data antropometri. Penelitian ini bertujuan membandingkan kinerja Ridge Regression, LASSO Regression, dan Bridge Regression dalam memprediksi persentase lemak tubuh berbasis variabel antropometri. Kinerja model dievaluasi berdasarkan nilai MSE, RMSE, R^2, Adjusted R^2, AIC, dan BIC untuk memperoleh metode yang memberikan performa prediksi terbaik (Hastie dkk., 2009; James dkk., 2021; Montgomery dkk., 2021).

2. TINJAUAN PUSTAKA

Regresi linier berganda merupakan metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel respon dengan beberapa variabel prediktor. Model regresi linier berganda secara umum dinyatakan pada Persamaan (1) sebagai

$$\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon \tag{1}$$

dengan $\boldsymbol{y}$ merupakan vektor respon berukuran $n \times 1$, $\boldsymbol{X}$ adalah matriks variabel prediktor berukuran $n \times p$, $\boldsymbol{\beta}$ merupakan vektor parameter regresi, dan $\boldsymbol{\varepsilon}$ adalah vektor galat acak. Estimasi parameter dilakukan menggunakan metode *Ordinary Least Squares* (OLS) dengan meminimumkan jumlah kuadrat galat sehingga diperoleh estimator pada Persamaan (2) (Nasution dkk., 2024).

$$\hat{\beta}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2}$$

Penduga OLS memberikan hasil yang optimal apabila asumsi klasik terpenuhi. Namun, apabila terjadi korelasi yang tinggi antarvariabel prediktor maka matriks $\boldsymbol{X}^T\boldsymbol{X}$ menjadi mendekati singular sehingga koefisien regresi menjadi tidak stabil dan memiliki varians yang besar. Kondisi tersebut dikenal sebagai multikolinearitas. Salah satu ukuran yang umum digunakan untuk mendeteksi multikolinearitas adalah *Variance Inflation Factor* (VIF) yang dirumuskan pada Persamaan (3), dengan R_j^2 merupakan koefisien determinasi hasil regresi variabel prediktor ke-(j) terhadap variabel prediktor lainnya. Nilai VIF yang

semakin besar menunjukkan tingkat multikolinearitas yang semakin tinggi sehingga diperlukan metode estimasi alternatif (Nisa & Hastuti, 2024).

$$VIF_j = \frac{1}{1-R_j^2} \tag{3}$$

Salah satu pendekatan yang banyak digunakan untuk mengatasi multikolinearitas adalah regresi regularisasi dengan menambahkan fungsi penalti pada proses estimasi parameter. Regresi Ridge menggunakan penalti kuadrat (*L2 penalty*) sehingga fungsi objektifnya dinyatakan pada Persamaan (4) sebagai

$$min_\beta\{\sum_{i=1}^n (y_i - x_i^T\beta)^2 + \lambda \sum_{j=1}^p \beta_j^2\} \tag{4}$$

dengan $\lambda \geq 0$ merupakan parameter penalti yang mengendalikan besar penyusutan koefisien regresi. Semakin besar nilai λ, maka koefisien regresi akan semakin mendekati nol tanpa menghilangkan variabel dari model.

Berbeda dengan Ridge, regresi Lasso menggunakan penalti absolut ($L1$ *penalty*) sehingga fungsi objektifnya dirumuskan pada Persamaan (5) sebagai

$$min_\beta\{\sum_{i=1}^n (y_i - x_i^T\beta)^2 + \lambda \sum_{j=1}^p |\beta_j|\} \tag{5}$$

Penalti tersebut memungkinkan sebagian koefisien bernilai nol sehingga selain mengurangi varians estimasi juga mampu melakukan seleksi variabel secara otomatis (Khoirunissa, dkk., 2024).

Regresi Bridge merupakan pengembangan dari Ridge dan Lasso dengan menggunakan fungsi penalti berpangkat γ, sehingga fungsi objektifnya dinyatakan pada Persamaan (6) sebagai

$$min_\beta\{\sum_{i=1}^n (y_i - x_i^T\beta)^2 + \lambda \sum_{j=1}^p |\beta_j|^\gamma\}, 0 < \gamma \leq 2 \tag{6}$$

Nilai $\gamma = 2$ menghasilkan regresi Ridge, sedangkan $\gamma = 1$ menghasilkan regresi Lasso. Apabila $0 < \gamma < 1$, metode Bridge mampu menghasilkan penyusutan koefisien yang lebih kuat sehingga memberikan fleksibilitas dalam proses estimasi parameter dan seleksi variabel sesuai karakteristik data. Estimasi parameter dilakukan melalui optimasi numerik dengan menentukan nilai parameter penalti optimum λ menggunakan *K-fold Cross Validation*. Dalam proses ini, sejumlah kandidat nilai λ dievaluasi secara sistematis melalui validasi silang, kemudian dihitung rata-rata galat prediksi pada setiap nilai λ. Nilai λ yang menghasilkan rata-rata galat prediksi terkecil dipilih sebagai parameter penalti optimum untuk membentuk model akhir (Enwere, Nduka, & Ogoke, 2023).

Kinerja model dievaluasi menggunakan beberapa ukuran ketepatan prediksi, yaitu *Mean Squared Error* (MSE) yang dirumuskan pada Persamaan (7).

$$MSE = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{7}$$

sedangkan *Root Mean Squared Error* (RMSE) diperoleh dari akar kuadrat nilai MSE. Selain itu, kemampuan model dalam menjelaskan keragaman data diukur menggunakan koefisien determinasi (R^2)dan *Adjusted* R^2 yang masing-masing dirumuskan pada Persamaan (8) dan Persamaan (9).

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

$$R_{adj}^2 = 1 - (\frac{n-1}{n-p-1})(1 - R^2) \tag{9}$$

Kompleksitas model selanjutnya dibandingkan menggunakan *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC), yang masing-masing dirumuskan pada Persamaan (10) dan Persamaan (11) sebagai

$$AIC = 2k - 2\ln(L) \tag{10}$$

$$BIC = k\ ln(n) - 2ln(L) \tag{11}$$

Model terbaik ditentukan berdasarkan nilai MSE, RMSE, AIC, dan BIC yang minimum serta nilai R^2 dan Adjusted R^2 yang maksimum (Jannah, Rusyana, & Ramadhani, 2024). MSE pada Persamaan (7) mengukur rata-rata kuadrat kesalahan prediksi sehingga lebih sensitif terhadap kesalahan yang besar, sedangkan RMSE merupakan akar dari MSE sehingga memiliki satuan yang sama dengan data dan lebih mudah diinterpretasikan. Sementara itu, R^2 pada Persamaan (8) menunjukkan proporsi keragaman variabel dependen yang dapat dijelaskan oleh model, sedangkan Adjusted R^2 pada Persamaan (9) merupakan pengembangan R^2 yang telah memperhitungkan jumlah variabel prediktor sehingga lebih tepat digunakan untuk membandingkan model dengan jumlah variabel yang berbeda. Di sisi lain, AIC dan BIC pada Persamaan (10) dan Persamaan (11) digunakan untuk memilih model terbaik dengan mempertimbangkan kecocokan model dan kompleksitasnya. AIC memberikan penalti yang lebih ringan terhadap jumlah parameter sehingga lebih berorientasi pada kemampuan prediksi, sedangkan BIC memberikan penalti yang lebih besar sehingga cenderung memilih model yang lebih sederhana (parsimonious) (Burnham & Anderson, 2004).

3. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan pemodelan statistik yang bertujuan membandingkan kinerja regresi linier klasik dan beberapa metode *penalized regression* dalam memprediksi persentase lemak tubuh (*Body Fat*). Data yang dianalisis adalah data sekunder yang berasal dari *platform* Kaggle berupa *dataset Body Fat* tahun 2021. Data yang digunakan terdiri atas 436 observasi, yang masing-masing merepresentasikan satu individu.

Variabel yang digunakan dalam penelitian ini terdiri atas satu variabel respon dan enam variabel prediktor. Variabel respon yang digunakan adalah persentase lemak tubuh (*Body Fat*) yang dinotasikan dengan Y. Adapun variabel prediktor meliputi umur (*Age*) (X_1), berat badan (*Weight*) (X_2), tinggi badan (*Height*) (X_3), lingkar leher (*Neck*) (X_4), lingkar perut (*Abdomen*) (X_5), dan lingkar pinggul (*Hip*) (X_6). Selanjutnya, cuplikan data yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Data Penelitian

Body Fat %	*Age (tahun)*	*Weight (kg)*	*Height (m)*	*Neck (cm)*	*Abdomen (cm)*	*Hip (cm)*
12,3	23	69,97	1,72	36,2	85,2	94,5
6,1	22	78,59	1,84	38,5	83	98,7
25,3	22	69,85	1,68	34	87,9	99,2
10,4	26	83,80	1,84	37,40	86,4	101,2
28,7	24	83,58	1,81	34,40	100	101,9
20,9	24	95,37	1,90	39	94,4	107,8
19,2	26	82,10	1,77	36,40	90,7	100,3
12,4	25	79,83	1,84	37,80	88,5	97,1
4,1	25	86,64	1,88	38,10	82,5	99,9
⋮	⋮	⋮	⋮	⋮	⋮	⋮
17,49	24	47,63	1,60	26	63	86,6

Analisis data dilakukan dengan bantuan *software* R-Studio. Tahapan analisis dalam penelitian ini diuraikan secara sistematis sebagai berikut:

1. Mengumpulkan data sekunder *Body Fat* tahun 2021 beserta faktor-faktor yang mempengaruhinya, dan menetapkan *Body Fat* sebagai variabel respon dan keenam variabel antropometri sebagai variabel prediktor.

2. Melakukan pemeriksaan struktur dan ringkasan data, pengecekan *missing values*, *outlier*, serta statistik deskriptif. Melakukan visualisasi data menggunakan histogram, *boxplot*, *scatter plot*, dan *correlation plot*.
3. Melakukan standardisasi terhadap seluruh variabel prediktor agar memiliki rataan nol dan varians satu, agar tidak terdapat perbedaan skala antarvariabel.
4. Mengestimasi model regresi linier berganda menggunakan metode *Ordinary Least Squares* (OLS) sebagai model dasar menggunakan estimator pada Persamaan (2).
5. Menghitung nilai *Variance Inflation Factor* (VIF) menggunakan Persamaan (3) pada model OLS untuk mengidentifikasi multikolinearitas antarvariabel prediktor.
6. Mendefinisikan fungsi Bridge Regression berdasarkan fungsi objektif pada Persamaan (6) yang meminimalkan jumlah kuadrat galat dengan tambahan penalti $\lambda \sum_{j=1}^{p} |\beta_j|^{\gamma}$. Parameter penalti λ ditentukan menggunakan 5-*fold cross-validation*, sedangkan parameter γ dievaluasi pada empat kandidat nilai, yaitu 0,5; 1; 1,5; dan 2. Pada proses ini, data dibagi secara acak menjadi lima kelompok (*fold*) yang berukuran hampir sama. Pada setiap iterasi, empat *fold* digunakan sebagai data pelatihan (*training*) dan satu *fold* sebagai data validasi (*validation*). Selanjutnya, dilakukan evaluasi terhadap setiap kandidat nilai λ pada masing-masing nilai γ berdasarkan rata-rata nilai *Mean Squared Error* (MSE) pada Persamaan (7) pada data validasi. Kombinasi parameter yang menghasilkan rata-rata MSE terkecil dipilih sebagai parameter optimum, kemudian digunakan untuk mengestimasi model Bridge Regression menggunakan metode optimisasi numerik.
7. Mengestimasi model Ridge Regression menggunakan fungsi penalti pada Persamaan (4). Nilai parameter penalti optimum λ ditentukan menggunakan 5-*fold cross-validation*, dengan memilih nilai λ yang menghasilkan rata-rata MSE pada Persamaan (7) validasi terkecil.
8. Mengestimasi model LASSO Regression menggunakan fungsi penalti pada Persamaan (5). Nilai parameter penalti optimum λ ditentukan menggunakan 5-*fold cross-validation*, dengan memilih nilai λ yang menghasilkan rata-rata MSE pada Persamaan (7) validasi terkecil.
9. Mengestimasi model akhir menggunakan parameter penalti optimum yang diperoleh dari proses *cross-validation*, kemudian menghitung ukuran kinerja setiap model yang meliputi *Mean Squared Error (MSE)* menggunakan Persamaan (7), R^2 menggunakan Persamaan (8), *Adjusted* R^2 menggunakan Persamaan (9), AIC menggunakan Persamaan (10), dan BIC menggunakan Persamaan (11).
10. Menghitung kembali nilai VIF menggunakan Persamaan (3) pada variabel yang tetap aktif pada model *penalized* untuk melihat efektivitas regularisasi dalam mengurangi multikolinearitas.
11. Melakukan evaluasi visual dengan *scatter plot Actual vs Predicted* untuk setiap model guna menilai kedekatan prediksi terhadap garis identitas ($y = x$).
12. Membandingkan performa model OLS, Bridge, Ridge, dan LASSO berdasarkan hasil evaluasi yang diperoleh, kemudian menentukan model terbaik dalam memprediksi persentase lemak tubuh serta menginterpretasikan koefisien yang dihasilkan.
13. Menarik kesimpulan.

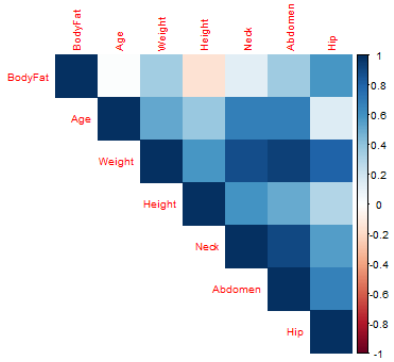
4. HASIL DAN PEMBAHASAN

Analisis diawali dengan eksplorasi data secara numerik dan visual untuk mengidentifikasi karakteristik data, distribusi, variasi, serta hubungan antarvariabel sebagai dasar dalam penyusunan model regresi. Ringkasan statistik deskriptif disajikan pada Tabel 2 berikut:

Tabel 2. Statistika Deskriptif Data

Statistika Deskriptif	*BodyFat*	*Age*	*Weight*	*Height*	*Neck*	*Abdomen*	*Hip*
Min	0	1	42,18	0,75	26	58	82,5
Q1	15,0225	21	59,87	1,66	31,5	70	94
Median	20,5	28,5	70,81	1,73	35,45	82,85	98,3
Mean	20,24	34,54	72,28	1,73	35,25	82,81	98,65
Q3	25,41	46	81,56	1,8	38,3	93,22	102,5
Max	47,5	81	164,72	1,97	51,2	148,1	147,7

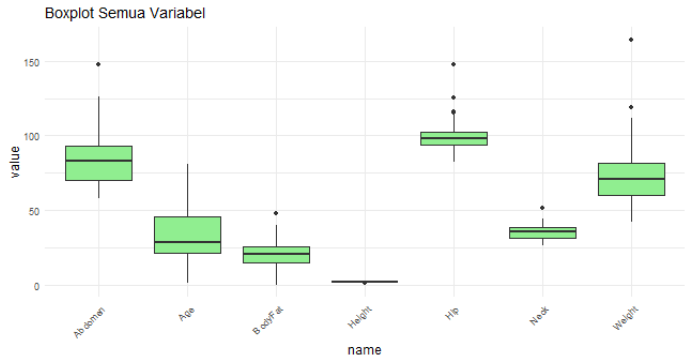
Berdasarkan tabel statistik deskriptif, variabel *Body Fat* memiliki rata-rata sebesar 20,25 dengan median 20,5, sehingga distribusinya relatif simetris. Variabel *Age* memiliki rata-rata sebesar 34,55 tahun dengan rentang usia 1 hingga 81 tahun. Nilai minimum usia sebesar 1 tahun tampak tidak lazim untuk studi antropometri dewasa, sehingga perlu dilakukan verifikasi terhadap kemungkinan kesalahan pencatatan data, keberadaan *outlier* ekstrem, atau penetapan kembali kriteria inklusi responden sebelum dilakukan pemodelan. Pada variabel *Weight*, rata-rata sebesar 72,28 dengan median 70,82 dan nilai maksimum mencapai 164,72 menunjukkan adanya individu dengan berat badan yang jauh lebih tinggi dibandingkan sebagian besar responden, sehingga berpotensi menjadi pencilan yang dapat memengaruhi hasil analisis. Variabel *Height* memiliki rata-rata sebesar 1,734 meter dengan rentang 0,75 hingga 1,97 meter. Nilai minimum tinggi badan sebesar 0,75 meter juga perlu dikonfirmasi karena tidak sesuai dengan karakteristik umum data antropometri dewasa dan dapat mengindikasikan adanya kesalahan input atau observasi yang perlu ditinjau kembali. Pada ukuran lingkar tubuh, *Neck* memiliki rata-rata sebesar 35,25, *Abdomen* sebesar 82,82, dan *Hip* sebesar 98,65, dengan pola penyebaran yang relatif moderat berdasarkan nilai kuartilnya. Secara umum, variabel penelitian menunjukkan variasi nilai yang cukup besar, terutama pada *Weight, Abdomen, Age,* dan *Height*. Secara umum, variabel penelitian menunjukkan variasi nilai yang cukup besar, terutama pada *Weight, Abdomen, Age*, dan *Height*. Temuan ini mengindikasikan perlunya perhatian terhadap keberadaan nilai ekstrem dalam proses pemodelan karena berpotensi memengaruhi hasil estimasi model regresi.



Gambar 1. *Plot* Korelasi Data

Heatmap korelasi menunjukkan kekuatan dan arah hubungan linear antarvariabel, dengan warna biru menandakan korelasi positif dan warna merah menunjukkan korelasi negatif. *Body Fat* memiliki korelasi positif yang cukup kuat dengan *Abdomen* dan *Hip*, sedangkan *Weight* berkorelasi kuat dengan *Neck*, *Abdomen*, dan *Hip*. *Height* hanya menunjukkan hubungan sedang dengan *Weight* dan relatif lemah dengan *Body Fat*, sedangkan *Age* memiliki hubungan yang tidak sekuat variabel ukuran tubuh. Secara keseluruhan, *Abdomen* merupakan variabel yang paling erat hubungannya dengan *Body Fat*

serta terdapat indikasi korelasi tinggi antarvariabel ukuran tubuh yang berpotensi menimbulkan multikolinearitas.



Gambar 2. *Box Plot* Data

Boxplot menunjukkan bahwa *Weight, Abdomen,* dan *Hip* memiliki variasi data yang relatif lebih besar dibandingkan variabel lainnya, sedangkan *Height* memiliki variasi paling kecil. Selain itu, terdapat beberapa *outlier* pada *Age, Body Fat, Weight, Abdomen,* dan *Hip* yang mengindikasikan adanya pengamatan ekstrem dan berpotensi memengaruhi hasil analisis regresi.



Gambar 3. *Scatter Plot* Data

Berdasarkan *scatterplot*, variabel *Abdomen, Hip,* dan *Weight* menunjukkan hubungan positif yang relatif kuat dengan *Body Fat*, yang mengindikasikan bahwa peningkatan ukuran lingkar tubuh dan berat badan cenderung diikuti oleh peningkatan persentase lemak tubuh. Variabel *Neck* juga menunjukkan hubungan positif, namun dengan kekuatan yang lebih rendah. Sebaliknya, *Height* memiliki hubungan negatif, sedangkan *Age* menunjukkan hubungan yang relatif lemah. Secara keseluruhan, *Abdomen* dan *Hip* merupakan variabel yang paling kuat hubungannya dengan persentase lemak tubuh dibandingkan variabel prediktor lainnya.

Selanjutnya dilakukan pembentukan model Regresi Linier Berganda menggunakan metode Ordinary Least Squares (OLS). Hasil estimasi parameter disajikan pada Tabel 3.

Tabel 3. Model Regresi OLS

Variabel	*Estimate*	*Std. Error*	*t value*	$Pr(>\|t\|)$
(*Intercept*)	−5,14342	11,0387	−0,47	0,6415
Age	−0,01774	0,02764	−0,64	0,5214
Weight	−0,18917	0,08181	−2,31	0,0212*
Height	−17,0617	3,5096	−4,86	$1{,}64 \times 10^{-06}$*
Neck	−1,1927	0,18903	−6,31	$6{,}97 \times 10^{-10}$*
Abdomen	0,49547	0,0645	7,682	$1{,}08 \times 10^{-13}$*
Hip	0,71236	0,08407	8,474	$3{,}8 \times 10^{-16}$*

Catatan: * adalah koefisien regresi yang signifikan tidak sama dengan nol pada *p-value*$<$ 0,05

$$\widehat{BodyFat} = -5{,}14342 - 0{,}01774\,Age - 0{,}18917\,Weight - 17{,}0617\,Height - 1{,}1927\,Neck + 0{,}49547 Abdomen + 0{,}71236\,Hip \quad (12)$$

Berdasarkan hasil Regresi Linier Berganda pada Persamaan (12), variabel *Weight, Height, Neck, Abdomen,* dan *Hip* berpengaruh signifikan terhadap *Body Fat* pada taraf signifikansi 5%, sedangkan *Age* tidak menunjukkan pengaruh yang signifikan. Koefisien regresi menunjukkan bahwa *Weight, Height,* dan *Neck* memiliki hubungan negatif dengan *Body Fat*, yang berarti peningkatan pada variabel tersebut cenderung diikuti penurunan persentase lemak tubuh dengan asumsi variabel lain konstan. Sebaliknya, *Abdomen* dan *Hip* memiliki hubungan positif, sehingga peningkatan lingkar perut dan lingkar pinggul cenderung meningkatkan persentase lemak tubuh. Di antara seluruh variabel prediktor, *Abdomen* dan *Hip* menunjukkan pengaruh yang paling dominan berdasarkan nilai statistik uji dan tingkat signifikansinya, sehingga keduanya dapat dianggap sebagai prediktor utama dalam model.

Sebelum analisis lanjutan, dilakukan pengujian asumsi klasik yang meliputi normalitas, multikolinearitas, heteroskedastisitas, dan autokorelasi. Hasil pengujian disajikan pada Tabel 4.

Tabel 4. Uji Asumsi Model Regresi OLS

Uji Asumsi	Nilai	*p-value*
Normalitas *Residual*	$A = 0{,}415$	0,333
Multikolinearitas	Age: 14,533; Weight: 27,496 Height: 2,042; Neck: 9,019 Abdomen: 15,306; Hip: 5,758	–
Heteroskedastisitas	$BP = 85{,}874, df = 6$	$< 2{,}2 \times 10^{-16}$
Autokorelasi	$DW = 1{,}6023$	$1{,}03 \times 10^{-05}$

Berdasarkan Tabel 4, residual model OLS berdistribusi normal, namun masih terdapat pelanggaran asumsi klasik berupa multikolinearitas pada beberapa variabel predictor terutama pada variabel *Weight* dan *Abdomen*, serta heteroskedastisitas dan autokorelasi residual. Temuan ini menunjukkan bahwa model OLS belum sepenuhnya memenuhi asumsi yang diperlukan sehingga penggunaan metode regularisasi menjadi relevan.

Analisis selanjutnya dilakukan menggunakan metode *penalized regression* yaitu Bridge, Ridge, dan LASSO, untuk mengatasi multikolinearitas serta meningkatkan stabilitas dan kemampuan prediksi model. Kinerja masing-masing metode dievaluasi berdasarkan MSE, RMSE, R^2, *Adjusted* R^2, AIC, BIC, dan VIF sebagaimana disajikan pada Tabel 5.

Tabel 5. Perbandingan Kinerja Model *Penalized* Regression

Kriteria	Bridge $\gamma = 0{,}5$	Bridge $\gamma = 1$	Bridge $\gamma = 1{,}5$	Bridge $\gamma = 2$	Ridge	LASSO
Lambda	33,340	91,668	16,675	8,343	0,441	0,097
MSE	25,008	25,525	25,168	25,278	26,151	25,495
RMSE	5,001	5,052	5,017	5,028	5,114	5,049
R^2	0,555	0,545	0,552	0,550	0,534	0,546
R^2 *adjusted*	0,548	0,539	0,546	0,544	0,528	0,542
AIC	1417,569	1426,492	1420,356	1422,258	1437,056	1421,977
BIC	1446,113	1455,036	1448,900	1450,801	1465,600	1442,365
VIF	27,430	27,430	27,430	27,430	27,430	7,466

Berdasarkan Tabel 4, model Bridge dengan $\gamma = 0{,}5$ menunjukkan performa prediksi terbaik di antara seluruh model *penalized*. Model ini menghasilkan MSE sebesar 25,008 dan RMSE sebesar 5,001, dengan nilai R^2 sebesar 0,555 dan R^2 adjusted sebesar 0,548 menunjukkan bahwa model ini mampu menjelaskan sekitar 55% variasi persentase lemak tubuh. Nilai AIC pada model ini juga merupakan yang paling kecil di antara seluruh model *penalized*, yaitu 1417,569. Namun demikian, nilai BIC terendah justru diperoleh oleh model LASSO sebesar 1442,365, sedangkan Bridge $\gamma = 0{,}5$ memiliki BIC sebesar 1446,113. Hal ini menunjukkan bahwa meskipun Bridge $\gamma = 0{,}5$ unggul dari sisi kesalahan prediksi dan AIC, tetapi dalam mempertimbangkan penalti kompleksitas model yang lebih ketat (BIC), LASSO lebih unggul dalam parsimoni model.

Model Ridge menunjukkan kinerja yang relatif lebih rendah dibandingkan metode lainnya. Nilai MSE sebesar 26,151 dan RMSE sebesar 5,114, merupakan nilai terbesar di antara seluruh model *penalized*. Nilai R^2 sebesar 0,534 dan R^2 adjusted sebesar 0,528 juga lebih rendah dibanding Bridge dan LASSO. Selain itu, Ridge memiliki nilai AIC dan BIC terbesar, sehingga secara keseluruhan kurang optimal berdasarkan indikator evaluasi yang digunakan.

Dari sisi multikolinearitas, nilai VIF maksimum pada model Ridge dan Bridge tetap sebesar 27,43, sama seperti pada model OLS. Hal ini bukan berarti regularisasi gagal, melainkan karena Ridge dan Bridge bekerja melalui mekanisme *shrinkage* terhadap koefisien tanpa mengubah hubungan linear antar prediktor dalam matriks desain. Dengan demikian, VIF berbasis variabel asli tetap tinggi karena struktur korelasi antar prediktor tidak berubah oleh penalti. Oleh karena itu, pada model regularisasi VIF tidak dijadikan ukuran utama keberhasilan metode, melainkan hanya sebagai informasi mengenai kondisi awal multikolinearitas.

Sebagai gantinya, keberhasilan Ridge dan Bridge lebih tepat dinilai dari kestabilan koefisien dan kemampuan prediksi. Hasil estimasi menunjukkan bahwa koefisien pada model penalized menjadi lebih stabil dibandingkan OLS, terutama pada variabel yang sebelumnya sangat dipengaruhi multikolinearitas. Hal ini menunjukkan bahwa regularisasi mampu menurunkan varians estimasi sehingga menghasilkan model yang lebih stabil. Oleh karena itu, evaluasi utama pada penelitian ini difokuskan pada konsistensi koefisien, nilai MSE, RMSE, R^2, adjusted R^2, AIC, BIC, serta plot *actual versus predicted*.

Variasi parameter γ pada metode Bridge menunjukkan bahwa perubahan bentuk penalti memengaruhi nilai λ optimal, tetapi hanya memberikan perbedaan kinerja yang relatif kecil. Hal ini terlihat dari nilai MSE yang berada pada kisaran 25,008–25,525 dan nilai R^2 sebesar 0,545–0,555. Dengan demikian, perubahan nilai γ tidak memberikan pengaruh yang berarti terhadap performa model, sehingga metode Bridge dapat dikatakan cukup stabil terhadap variasi parameter penalti.

Selanjutnya, dilakukan evaluasi koefisien regresi dan nilai VIF untuk menilai pengaruh penalti terhadap model. Perbandingan hasil disajikan pada Tabel 6.

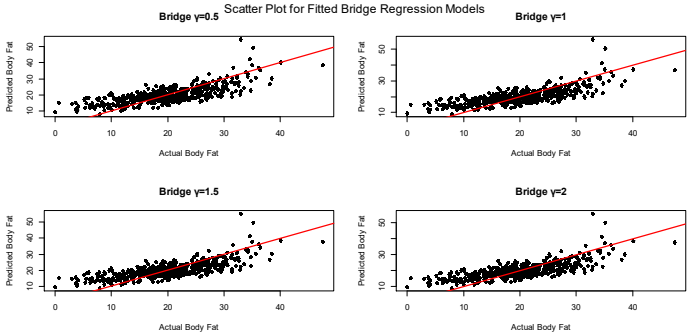
Tabel 6. Koefisien Setiap Model

Kriteria	OLS	Bridge $\gamma = 0{,}5$	Bridge $\gamma = 1$	Bridge $\gamma = 1{,}5$	Bridge $\gamma = 2$	Ridge	LASSO
Intercept	−5,143	20,250	20,250	20,250	20,250	20,250	20,250
Age	−0,018	0,000	0,000	−0,075	−0,044	0,021	0,000
Weight	−0,189	−2,389	−0,001	−1,272	−1,110	−0,054	0,000
Height	−17,062	−1,733	−2,059	−1,926	−1,982	−2,151	-2,060
Neck	−1,193	−4,693	−4,299	−4,348	−4,092	−3,192	-4,374
Abdomen	0,496	6,722	4,747	5,721	5,322	3,689	4,822
Hip	0,712	4,876	4,071	4,478	4,465	4,108	4,069

Berdasarkan Tabel 6, terlihat bahwa seluruh metode *penalized* menunjukkan adanya proses *shrinkage* atau pengecilan koefisien dibandingkan model OLS. Perubahan paling mencolok terjadi pada variabel *Height*, di mana nilai koefisien menurun drastis dari -17.062 pada OLS menjadi sekitar -2 pada seluruh model *penalized*. Penurunan ini menunjukkan bahwa estimasi OLS sebelumnya mengalami inflasi akibat multikolinearitas yang tinggi, sehingga penalti membantu menghasilkan estimasi yang lebih stabil. Model LASSO menunjukkan karakteristik sparsity dengan mengeliminasi dua variabel, yaitu *Age* dan *Weight*, melalui koefisien bernilai nol. Hal ini konsisten dengan penurunan nilai VIF yang signifikan pada LASSO (7,47), menandakan bahwa seleksi variabel yang dilakukan berhasil mengurangi korelasi antar prediktor. Sebaliknya, model Bridge dan Ridge mempertahankan seluruh variabel tetapi dengan magnitude koefisien yang lebih terkendali, sehingga tetap menjaga kontribusi prediktor utama seperti *Abdomen*, *Hip*, dan *Neck*.

Secara umum, model Bridge $\gamma = 0{,}5$ memberikan akurasi prediksi terbaik berdasarkan MSE dan AIC, sedangkan model LASSO lebih lebih unggul dalam aspek parsimoni dan reduksi multikolinearitas berdasarkan BIC dan nilai VIF. Ridge tidak menunjukkan keunggulan dominan dalam konteks data ini, meskipun tetap mampu menghasilkan estimasi koefisien yang lebih stabil dibandingkan OLS. Dengan kata lain, Bridge lebih sesuai untuk tujuan prediktif, sedangkan LASSO lebih tepat digunakan ketika fokus analisis berada pada kesederhanaan dan stabilitas struktur model.

Setelah evaluasi numerik dan struktural, analisis dilanjutkan dengan evaluasi visual antara nilai aktual (*Actual Body Fat*) dan nilai prediksi (*Predicted Body Fat*) melalui scatter plot. Gambar berikut menampilkan scatter plot dari setiap model *penalized*, di mana garis merah ($y = x$) merepresentasikan garis identitas. Semakin dekat titik observasi terhadap garis tersebut, semakin tinggi tingkat akurasi model.

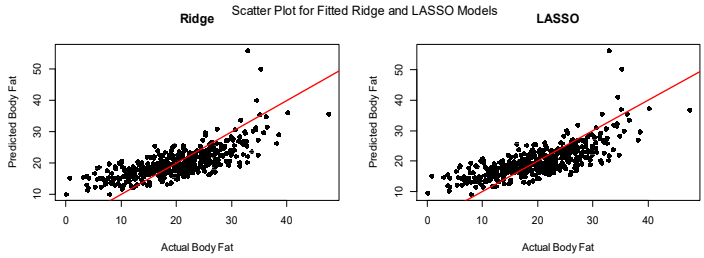


Gambar 4. *Scatter Plot Actual vs Predicted* pada Bridge Regression

Berdasarkan Gambar 4, seluruh variasi Bridge menunjukkan hubungan linear positif antara nilai aktual dan nilai prediksi. Variasi $\gamma = 0{,}5$ memperlihatkan kedekatan titik observasi yang paling konsisten terhadap garis identitas, terutama pada rentang nilai lemak tubuh menengah, sehingga menunjukkan akurasi prediksi yang lebih baik dibandingkan variasi γ lainnya. Meskipun pola hubungan tetap stabil pada seluruh nilai γ, penyimpangan prediksi masih terlihat pada observasi dengan persentase lemak tubuh yang tinggi (di atas $\pm 40\%$). Temuan ini menunjukkan bahwa model Bridge mampu menangkap pola data dengan baik, namun masih menghadapi keterbatasan dalam memprediksi nilai ekstrem. Secara visual, hasil ini sejalan dengan evaluasi numerik yang menempatkan $\gamma = 0{,}5$ sebagai model terbaik.

Berdasarkan Gambar 5, model Ridge dan LASSO sama-sama menunjukkan hubungan linear positif antara nilai aktual dan prediksi, dengan sebagian besar titik berada di sekitar garis identitas pada rentang nilai menengah. Namun, pada nilai lemak tubuh yang tinggi

terlihat kecenderungan prediksi berada di bawah garis identitas, terutama pada Ridge, yang mengindikasikan adanya underestimation. Secara visual, perbedaan antara Ridge dan LASSO tidak terlalu mencolok, meskipun keduanya menunjukkan kedekatan terhadap garis identitas yang sedikit kurang konsisten dibandingkan model Bridge $\gamma = 0{,}5$. Temuan ini sejalan dengan hasil evaluasi numerik, di mana Ridge tidak menunjukkan keunggulan dominan, sedangkan LASSO lebih unggul dari sisi efisiensi model dibandingkan akurasi prediksi maksimum



Gambar 5. *Scatter Plot Actual vs Predicted* Pada Ridge Dan LASSO Regression

5. KESIMPULAN

Berdasarkan hasil penelitian, variabel antropometri yang memiliki hubungan paling kuat dengan persentase lemak tubuh adalah Abdomen dan Hip. Model regresi OLS menunjukkan adanya multikolinearitas yang cukup tinggi, terutama pada variabel Weight dan Abdomen, sehingga diperlukan metode regularisasi untuk memperoleh estimasi parameter yang lebih stabil dan meningkatkan kemampuan prediksi model.

Hasil perbandingan menunjukkan bahwa Bridge Regression dengan parameter $\gamma = 0{,}5$ memberikan performa terbaik berdasarkan nilai MSE, RMSE, R^2, dan AIC. Sementara itu, LASSO menghasilkan model yang lebih sederhana melalui seleksi variabel serta memiliki nilai BIC terendah. Oleh karena itu, Bridge Regression $\gamma = 0{,}5$ direkomendasikan untuk tujuan prediksi persentase lemak tubuh, sedangkan LASSO lebih sesuai digunakan apabila kesederhanaan dan interpretabilitas model menjadi prioritas.

DAFTAR PUSTAKA

Andari, R. Y., Pradipta, R. A., & Radianto, D. O. (2023). Using Bayesian Ridge Algorithm to Predict Effectiveness of Body Fat Measurement. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 43–49.

Burnham, K. P., Anderson, D. R. (2004). *Multimodel inference: Understanding AIC and BIC in model selection*. Sociological Methods & Research, 33(2), 261–304.

Bühlmann, P. & van de Geer, S., 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Berlin: Springer.

Enwere, K., Nduka, E., & Ogoke, U. (2023). Comparative analysis of Ridge, Bridge and Lasso regression models in the presence of multicollinearity. *IPS Intelligentsia Multidisciplinary Journal*, 3(1), 1–8. https://doi.org/10.54117/iimj.v3i1.5

Frank, I.E. & Friedman, J.H., 1993. *A statistical view of some chemometrics regression tools*. *Technometrics*, 35(2), pp.109–135. https://doi.org/10.1080/00401706.1993.10485033

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning*. 2nd ed. New York: Springer.

Hoerl, A.E. & Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), pp.55–67.

James, G., Witten, D., Hastie, T. & Tibshirani, R., 2021. *An Introduction to Statistical Learning*. 2nd ed. New York: Springer.

Jannah, S. R., Rusyana, A., & Ramadhani, E. (2025). Perbandingan metode regresi OLS, Ridge, dan Lasso dalam memodelkan kemiskinan di Indonesia tahun 2024. *Indonesian Journal of Applied Statistics 7(2)*. https://jurnal.uns.ac.id/ijas/article/view/96921

Khairani, N. & Sudiarti, T., 2020. Model Prediksi Persen Lemak Tubuh Remaja Putri: Studi Cross Sectional. *Nutri-Sains: Jurnal Gizi, Pangan dan Aplikasinya*, 4(1), pp.51–66.

Khoirunissa, H. A., Wijaya, A. R., Isnaini, B., & Ferawati, K. (2024). Analisis faktor-faktor penyebab inflasi di Indonesia menggunakan regresi Ridge, LASSO, dan Elastic-Net. *Indonesian Journal of Applied Statistics 7(2).* https://jurnal.uns.ac.id/ijas/article/view/96921

Kutner, M.H., Nachtsheim, C.J., Neter, J. & Li, W., 2005. *Applied Linear Statistical Models*. 5th ed. New York: McGraw-Hill.

Kyle, U.G., Bosaeus, I., De Lorenzo, A.D., Deurenberg, P. dkk., 2004. Bioelectrical impedance analysis—part I: review of principles and methods. *Clinical Nutrition*, *23(5),* pp.1226–1243. https://doi.org/10.1016/j.clnu.2004.06.004

Montgomery, D.C., Peck, E.A. & Vining, G.G., 2021. *Introduction to Linear Regression Analysis*. 6th ed. Hoboken: Wiley.

Nasution, M. R., Sutarman, Darnius, O., & Rosmaini, E. (2024). Regularisasi regresi linier berganda pada data berdimensi tinggi untuk mengatasi efek multikolinearitas. *MES (Journal of Mathematics Education and Science)* **10**(1): 43–51. https://jurnal.uisu.ac.id/index.php/mesuisu/article/download/9469/pdf DOI: 10.30743/mes.v10i1.9469.

Nevill, A.M., Stewart, A.D., Olds, T. & Holder, R., 2006. Relationship between adiposity and body size reveals limitations of BMI. *American Journal of Physical Anthropology*, 129(1), pp.151–156.

Nisa, C., & Hastuti, S. H. (2024). Kajian simulasi perbandingan metode Ridge Regression dan Adjusted Ridge Regression untuk penanganan multikolinearitas. *Jurnal Gaussian* 12(3): 330–339. https://ejournal3.undip.ac.id/index.php/gaussian/article/view/37678 DOI: 10.14710/j.gauss.12.3.330-339

Pratiwi, H., & Sari, K. (2017). Perbandingan Metode Ridge, LASSO, dan Bridge Regression dalam Mengatasi Multikolinearitas. *Jurnal Statistika Industri dan Komputasi*, *2*(1), 15–24.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), pp.267–288.