

Perbandingan Metode Kuadrat Terkecil, LASSO, Ridge, dan Elastic Net dalam Memprediksi Harga Mobil

Aulia Zahra Panjaitan¹, Ahmad Badawi², Risa Rozzaqi Rambe³, Riska Audina⁴

¹²³⁴Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Negeri Medan, Indonesia

Article Info

Article history:

Received: November 2025

Accepted: Desember 2025

Keywords:

Elastic-Net

Harga Mobil

LASSO

Metode Kuadrat Terkecil

Ridge Regression

ABSTRACT

Harga mobil dipengaruhi oleh berbagai karakteristik teknis kendaraan yang saling berkorelasi, sehingga berpotensi menimbulkan multikolinearitas pada pemodelan regresi. Penelitian ini bertujuan membandingkan kinerja metode *Ordinary Least Squares (OLS)*, *Ridge Regression*, LASSO, dan *Elastic-Net* dalam memprediksi harga mobil menggunakan *Car Price Dataset* dari Kaggle. Data dianalisis menggunakan perangkat lunak R melalui tahap pra-pemrosesan, standarisasi variabel, pengujian multikolinearitas menggunakan *Variance Inflation Factor (VIF)*, serta deteksi outlier dengan *boxplot* dan grafik diagnostik. Pemodelan Ridge, LASSO, dan *Elastic-Net* dilakukan menggunakan teknik 10-fold *cross-validation* dengan pemilihan parameter optimal berdasarkan nilai *Mean Squared Error (MSE)*. Hasil menunjukkan adanya multikolinearitas kuat pada variabel *citympg*, *highwaympg*, dan *curbweight*. Metode regularisasi mampu melakukan penyusutan koefisien dan seleksi variabel, khususnya LASSO dan *Elastic-Net* yang menghasilkan model lebih sederhana. Namun, evaluasi kinerja prediksi menunjukkan bahwa OLS memperoleh nilai MSE terkecil sebesar 12.489.548, sehingga menjadi model terbaik pada penelitian ini. Variabel *curbweight* dan *enginesize* berpengaruh positif signifikan terhadap harga mobil, sedangkan *citympg* berpengaruh negatif signifikan.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Aulia Zahra Panjaitan

Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Negeri Medan

20221 Medan, Sumatera Utara, Indonesia

Email: auliazahra.4243260038@mhs.unimed.ac.id

1. INTRODUCTION

Harga mobil merupakan salah satu indikator penting dalam sektor otomotif yang berperan besar terhadap dinamika ekonomi konsumen, industri manufaktur, dan pasar perdagangan kendaraan. Pergerakan harga mobil dipengaruhi oleh berbagai faktor, mulai dari karakteristik fisik kendaraan, performa mesin, kapasitas produksi, hingga faktor eksternal seperti kondisi pasar dan preferensi konsumen. Pemahaman yang baik terhadap faktor-faktor penentu harga mobil sangat diperlukan untuk membantu pelaku industri dalam pengambilan keputusan strategis, termasuk penetapan harga, perencanaan produksi, dan pengembangan produk. Di sisi lain, konsumen juga dapat memperoleh gambaran yang lebih objektif mengenai nilai kendaraan berdasarkan spesifikasinya.

Dalam era digital, ketersediaan data otomotif semakin melimpah, salah satunya melalui platform open data seperti Kaggle. Dataset *Car Price Prediction* menyediakan informasi detail mengenai harga mobil serta karakteristiknya, seperti tipe bahan bakar, ukuran mesin, kapasitas silinder, tenaga, konsumsi bahan bakar, hingga dimensi kendaraan [1]. Data ini sangat relevan untuk dianalisis menggunakan pendekatan statistik dan machine learning, khususnya metode regresi dengan regularisasi yang mampu menangani variabel yang saling berkorelasi serta kompleksitas data yang tinggi.

Metode *Ordinary Least Squares (OLS)* atau metode kuadrat terkecil merupakan pendekatan dasar yang banyak digunakan dalam pemodelan regresi karena kemampuannya memberikan estimasi koefisien yang efisien pada kondisi ideal. Namun, OLS memiliki kelemahan ketika terdapat multikolinearitas antarvariabel atau jumlah prediktor yang besar, yang menyebabkan varians koefisien menjadi tidak stabil dan mengurangi performa prediksi. Untuk mengatasi hal tersebut, berbagai metode regularisasi telah dikembangkan, seperti *Ridge Regression* yang diperkenalkan oleh Hoerl dan Kennard [2], LASSO oleh Tibshirani [3], serta *Elastic Net* oleh Zou dan Hastie [4]. *Ridge Regression* (Regresi Ridge) merupakan metode yang dikembangkan untuk menstabilkan nilai koefisien regresi karena adanya multikolinearitas, *Ridge Regression* bekerja dengan menambahkan penalti L_2 yang menyusutkan koefisien tanpa menghilangkan variabel [5]. Pada metode LASSO dengan menggunakan penalti norma L_1 , LASSO memilih fitur berdasarkan besar kecilnya koefisien regresi, bukan berdasarkan signifikansi statistiknya terhadap variabel keluaran [6]. Metode *elastic net* merupakan pendekatan dalam analisis regresi yang menggabungkan prinsip-prinsip dari regresi ridge dan regresi Lasso [7]. *Elastic net* menerapkan dua jenis penalti, yaitu L_1 dan L_2 , pada fungsi kerugian, sehingga metode ini mampu menangani masalah multikolinearitas sekaligus melakukan seleksi fitur yang relevan secara efektif.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk membandingkan performa metode Kuadrat Terkecil (OLS), Ridge, LASSO, dan *Elastic Net* dalam memprediksi harga mobil menggunakan dataset *Car Price Prediction* dari Kaggle. Analisis dilakukan untuk melihat metode mana yang memberikan estimasi koefisien paling stabil, seleksi variabel paling efektif, serta akurasi prediksi terbaik. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pemodelan harga kendaraan dan memberikan wawasan mengenai penerapan metode regularisasi pada data industri otomotif.

2. METHOD

Seluruh proses pengolahan dan analisis data dalam penelitian ini dilakukan menggunakan perangkat lunak R (*The R Foundation for Statistical Computing*). *Software R* digunakan untuk melakukan tahap pra-pemrosesan data, standardisasi variabel, pemodelan *Ordinary Least Squares (OLS)*, pengujian multikolinearitas menggunakan *Variance Inflation Factor (VIF)*, deteksi outlier melalui analisis residual dan *boxplot*, serta pemodelan regresi berpenalti meliputi *Ridge*, LASSO, dan *Elastic Net* dengan optimasi parameter menggunakan metode *10-fold cross-validation*. Visualisasi hasil analisis berupa grafik *cross-validation*, jalur koefisien (*coefficient path*), *boxplot*, dan plot diagnostik juga dihasilkan menggunakan software R.

2.1 Data Penelitian dan Variabel

Data yang digunakan pada penelitian ini merupakan data sekunder yang berasal dari *Car Price Dataset* yang tersedia pada platform Kaggle. Dataset ini memuat informasi harga mobil sebagai variabel dependen serta berbagai karakteristik teknis kendaraan sebagai variabel independen. Variabel respon yang digunakan adalah harga mobil (*price*), sedangkan variabel prediktor terdiri dari beberapa karakteristik kendaraan yang diduga berpengaruh terhadap penetapan harga mobil, yaitu *wheelbase* (jarak sumbu roda), *carlength* (panjang kendaraan), *curbweight* (berat kendaraan), *enginesize* (kapasitas mesin), *citympg* (konsumsi bahan bakar di dalam kota), dan *highwaympg* (konsumsi bahan bakar di jalan tol). Deskripsi variabel penelitian disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Deskripsi	Satuan
y	Harga mobil (<i>price</i>)	USD
x ₁	<i>wheelbase</i>	mm
x ₂	<i>carlength</i>	mm
x ₃	<i>curbweight</i>	kg
x ₄	<i>enginesize</i>	cc
x ₅	<i>citympg</i>	mpg
x ₆	<i>highwaympg</i>	mpg

2.2 Pra-pemrosesan Data

Sebelum dilakukan analisis, dilakukan *pre-processing* data berupa pemeriksaan nilai hilang dan pembersihan data yang mengandung nilai nol. Selain itu, karena setiap variabel memiliki satuan dan skala yang berbeda, proses standarisasi dilakukan agar seluruh variabel prediktor berada pada skala yang sama. Standarisasi ini penting untuk menjaga kualitas estimasi koefisien regresi, khususnya pada metode regularisasi yang sensitif terhadap perbedaan skala seperti Ridge, LASSO, dan *Elastic Net*.

Dalam proses standarisasi dan *cross-validation*, digunakan perintah `set.seed(123)` untuk memastikan bahwa pembagian data dalam proses validasi silang bersifat konsisten dan dapat direproduksi. Penggunaan *seed* ini menjamin bahwa setiap kali analisis dijalankan, hasil pemilihan parameter penalti (λ) dan struktur model yang terbentuk akan tetap sama sehingga meningkatkan reproduktibilitas penelitian. Setelah diperoleh estimasi parameter pada skala standar, nilai koefisien kemudian dikonversi kembali ke satuan aslinya (*inverse transformation*) agar interpretasi model dapat dilakukan dengan lebih mudah. Langkah ini memastikan bahwa hasil analisis tetap memiliki makna praktis sesuai konteks harga mobil dan karakteristik teknis kendaraan.

2.3 Metode Kuadrat Terkecil (*Ordinary Least Squares* / OLS)

Tahap awal analisis dilakukan dengan membangun model menggunakan metode Kuadrat Terkecil (*Ordinary Least Squares* / OLS). OLS merupakan pendekatan regresi linear klasik yang bertujuan meminimumkan jumlah kuadrat selisih antara nilai aktual dan nilai prediksi atau yang dikenal sebagai *Residual Sum of Squares* (RSS). Fungsi tujuan OLS dapat dituliskan pada persamaan (1):

$$\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^T (y_t - \mathbf{X}_t' \boldsymbol{\beta})^2 \quad (1)$$

dengan:

y_t : variabel respon (harga mobil),

\mathbf{X}_t : vektor variabel prediktor,

$\boldsymbol{\beta}$: vektor parameter regresi yang diestimasi,

T : jumlah observasi.

Model OLS digunakan sebagai model baseline untuk membandingkan performa metode regularisasi lain.

2.4 Uji Multikolinearitas dan Deteksi Outlier

Setelah model OLS terbentuk, dilakukan pengujian asumsi klasik terutama multikolinearitas menggunakan indikator *Variance Inflation Factor* (VIF). Jika nilai VIF >10 menunjukkan adanya korelasi tinggi antar prediktor yang dapat menyebabkan ketidakstabilan estimasi koefisien OLS serta memperbesar varians parameter.

Selain itu, dilakukan deteksi outlier dan observasi berpengaruh guna mengidentifikasi data ekstrem yang berpotensi mendistorsi hasil estimasi. Deteksi dilakukan menggunakan analisis *boxplot* serta indikator statistik berupa:

Cook's Distance untuk mengukur pengaruh observasi terhadap estimasi model,

Leverage sebagai ukuran pengaruh titik data terhadap posisi garis regresi,

Studentized residuals untuk mengidentifikasi observasi dengan galat ekstrem.

Temuan multikolinearitas dan keberadaan observasi berpengaruh menjadi dasar penggunaan metode regresi regularisasi sebagai alternatif yang lebih stabil dibanding OLS.

2.5 Metode Ridge

Metode Ridge diperkenalkan oleh Hoerl dan Kennard sebagai metode yang efektif untuk mengatasi masalah multikolinearitas. Ridge merupakan pengembangan dari OLS dengan menambahkan penalti terhadap besar koefisien regresi sehingga nilai koefisien mengalami penyusutan (*shrinkage*) dan varians estimasi dapat dikendalikan. Pada regresi Ridge digunakan penalti L_2 -norm, yaitu:

$$\|\beta\|_2^2 = \sum_{i=1}^p \beta_i^2$$

Sehingga fungsi tujuan Ridge yang meminimumkan *Penalized Residual Sum of Squares* (PRSS) dapat dituliskan pada persamaan (2):

$$\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^T (y_t - \mathbf{X}_t' \beta)^2 + \lambda \sum_{i=1}^p \beta_i^2 \quad (2)$$

Parameter $\lambda > 0$ berfungsi sebagai parameter penalti. Semakin besar nilai λ , maka nilai koefisien regresi akan semakin menyusut mendekati nol, meskipun tidak pernah tepat menjadi nol.

2.6 Metode Least Absolute Shrinkage and Selection Operator (LASSO)

Metode LASSO dikembangkan oleh Tibshirani sebagai pendekatan regresi yang tidak hanya melakukan penyusutan koefisien, tetapi juga seleksi variabel secara otomatis. Berbeda dengan Ridge, LASSO menggunakan penalti L_1 -norm, yaitu:

$$\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$$

Fungsi tujuan LASSO dituliskan pada persamaan (3):

$$\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^T (y_t - \mathbf{X}_t' \beta)^2 + \lambda \sum_{i=1}^p |\beta_i| \quad (3)$$

Parameter penalti λ mengontrol tingkat penyusutan. Semakin besar nilai λ , semakin banyak koefisien yang ditekan hingga tepat bernilai nol, sehingga hanya variabel yang paling signifikan yang dipertahankan dalam model.

2.7 Metode Elastic Net

Elastic Net pertama kali diperkenalkan oleh Zou dan Hastie sebagai metode yang mengombinasikan keunggulan Ridge dan LASSO. *Elastic Net* mampu melakukan penyusutan koefisien sekaligus seleksi variabel secara lebih stabil, khususnya pada kondisi multikolinearitas tinggi.

Fungsi tujuan *Elastic Net* dirumuskan pada persamaan (4):

$$\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^T (y_t - \mathbf{X}_t' \beta)^2 + \lambda [(1 - \gamma) \|\beta\|_2^2 + \gamma \|\beta\|_1] \quad (4)$$

atau dapat dijabarkan pada persamaan (5):

$$\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^T (y_t - \mathbf{X}_t' \beta)^2 + \lambda [(1 - \gamma) \sum_{i=1}^p \beta_i^2 + \gamma \sum_{i=1}^p |\beta_i|] \quad (5)$$

dengan:

$\lambda > 0$: parameter penalti,

$\gamma \in [0,1]$: parameter pencampur antara Ridge dan LASSO.

Jika $\gamma = 0$ maka metode menjadi Ridge sesuai dengan fungsi tujuan pada persamaan (2), sedangkan jika $\gamma = 1$ maka menjadi LASSO sesuai dengan fungsi tujuan pada persamaan (3).

2.8 Pemilihan Model Terbaik dan Evaluasi Kinerja

Pemilihan parameter regularisasi λ dan γ dilakukan menggunakan teknik *10-fold cross-validation* untuk meminimumkan kesalahan prediksi pada data validasi. Setiap metode (OLS, Ridge, LASSO, dan *Elastic Net*) dibandingkan berdasarkan nilai *Mean Squared Error* (MSE) pada data uji. Model dengan nilai MSE terkecil dipilih sebagai model terbaik dalam memprediksi harga mobil.

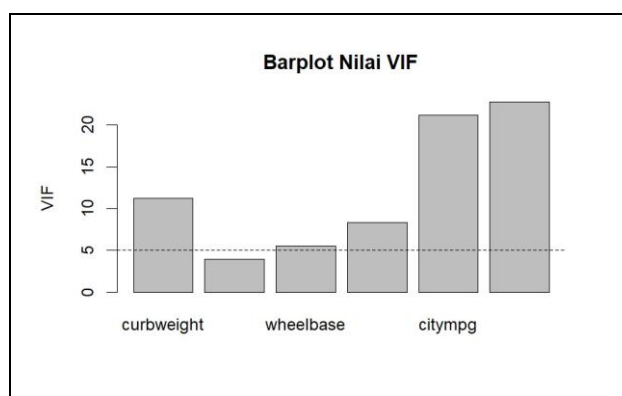
3. RESULTS AND DISCUSSION

Harga mobil merupakan fenomena yang dipengaruhi oleh berbagai karakteristik teknis kendaraan, seperti berat kendaraan, kapasitas mesin, dimensi bodi, serta efisiensi bahan bakar. Banyaknya variabel prediktor tersebut berpotensi menimbulkan masalah *multikolinearitas* pada model *Ordinary Least Squares* (OLS), yaitu kondisi ketika antar variabel saling berkorelasi tinggi sehingga dapat menghasilkan estimasi koefisien yang tidak stabil. Oleh karena itu, sebelum melakukan pemodelan lanjutan, dilakukan pengujian asumsi awal berupa pemeriksaan multikolinearitas dan deteksi outlier.

Hasil uji *Variance Inflation Factor* (VIF) menunjukkan bahwa beberapa variabel, seperti *citympg* dan *highwaympg*, memiliki nilai VIF yang sangat tinggi sehingga mengindikasikan korelasi kuat antar prediktor. Selain itu, pemeriksaan outlier menggunakan *boxplot* harga mobil serta grafik diagnostik OLS memperlihatkan adanya observasi ekstrem dan beberapa titik berpengaruh (*leverage* tinggi). Penelitian ini mengarah pada perlunya penerapan metode regularisasi seperti *Ridge*, *LASSO*, dan *Elastic-Net* untuk memperoleh parameter yang lebih stabil.

3.1 Uji Multikolinearitas

Pengujian multikolinearitas dilakukan untuk menilai apakah variabel-variabel prediktor dalam model memiliki korelasi yang tinggi satu sama lain. Pengujian menggunakan nilai *Variance Inflation Factor* (VIF), dengan batas umum $VIF > 10$ menandakan adanya multikolinearitas kuat yang dapat menyebabkan ketidakstabilan koefisien pada model *Ordinary Least Squares* (OLS).



Gambar 1. Barplot Nilai VIF untuk Variabel Prediktor

Selain melalui visualisasi, nilai VIF juga disajikan dalam tabel untuk menunjukkan besaran multikolinearitas pada masing-masing variabel prediktor sebagai berikut.

Tabel 2. Nilai VIF Variabel Prediktor

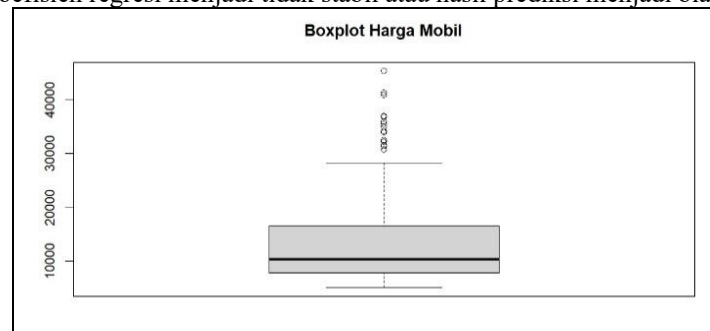
<i>curbweight</i>	<i>enginesize</i>	<i>wheelbase</i>	<i>carlength</i>	<i>citympg</i>	<i>highwaympg</i>	<i>price</i>
11.269	3.963	5.527	8.357	21.218	22.787	–

Keterangan: Variabel *price* merupakan variabel respon sehingga tidak memiliki nilai VIF.

Hasil tabel dan visualisasi menunjukkan bahwa variabel *citympg* dan *highwaympg* memiliki nilai VIF tertinggi, yaitu lebih dari 20, yang menandakan adanya multikolinearitas sangat kuat. Variabel *curbweight* juga memiliki nilai $VIF > 10$, sehingga turut berpotensi memengaruhi kestabilan estimasi parameter OLS. Sementara itu, variabel lain seperti *enginesize*, *wheelbase*, dan *carlength* memiliki $VIF < 10$ sehingga tidak menjadi sumber multikolinearitas utama. Berdasarkan temuan ini, diperlukan pemodelan berbasis regularisasi untuk menurunkan dampak multikolinearitas dan meningkatkan stabilitas koefisien regresi.

3.2 Deteksi Outlier

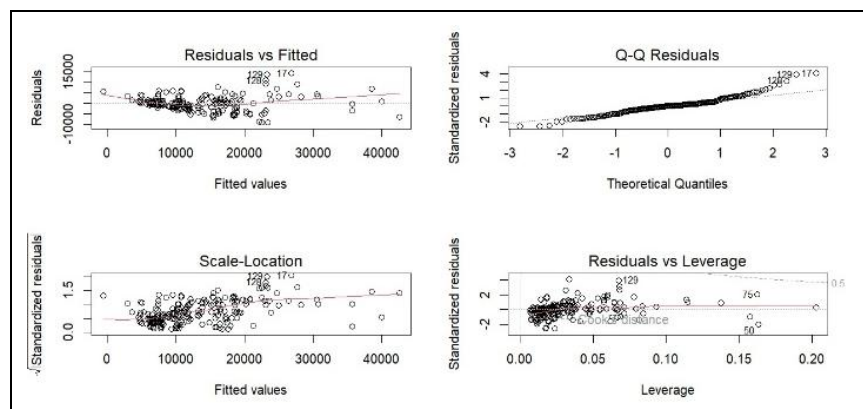
Deteksi outlier dilakukan untuk mengidentifikasi keberadaan observasi ekstrem yang berpotensi memengaruhi kestabilan model regresi. Pemeriksaan awal dilakukan menggunakan *boxplot* untuk melihat sebaran harga mobil dan titik-titik yang berada jauh dari mayoritas data. Outlier pada variabel respon dapat menyebabkan koefisien regresi menjadi tidak stabil atau hasil prediksi menjadi bias.



Gambar 2. *Boxplot* Harga Mobil

Hasil *boxplot* menunjukkan bahwa variabel *price* memiliki sejumlah observasi ekstrem yang berada jauh di atas rentang nilai utama. Titik-titik tersebut menggambarkan mobil dengan harga jauh lebih tinggi dibandingkan mayoritas data dalam sampel. Keberadaan outlier ini perlu diperhatikan karena dapat meningkatkan nilai residual secara drastis dan mempengaruhi kinerja model *Ordinary Least Squares (OLS)*.

Selain itu, pemeriksaan lanjutan dilakukan melalui grafik diagnostik OLS yang terdiri dari *Residuals vs Fitted*, *Normal Q-Q*, *Scale-Location*, dan *Residuals vs Leverage*. Grafik-grafik ini memberikan informasi mengenai kesesuaian asumsi model, penyimpangan residual, serta keberadaan titik-titik berpengaruh (*influential points*).



Gambar 3. Grafik Diagnostik Model OLS

Grafik *Residuals vs Fitted* memperlihatkan pola non-acak, yang mengindikasikan adanya ketidakhomogenan varians (heteroskedastisitas). *Normal Q-Q Plot* menunjukkan bahwa beberapa residual, terutama pada kuantil atas, menyimpang dari garis diagonal, menandakan bahwa residual tidak sepenuhnya mengikuti distribusi normal. *Plot Scale-Location* mengonfirmasi kembali variasi residual yang tidak konstan. Pada *plot Residuals vs Leverage*, terlihat beberapa titik observasi seperti nomor 129, 128, dan 170 yang memiliki leverage tinggi dan berada dekat dengan garis *Cook's Distance*, sehingga berpotensi memberikan pengaruh besar terhadap estimasi model. Temuan ini menunjukkan bahwa data mengandung observasi ekstrem dan titik berpengaruh yang dapat menurunkan keandalan model OLS.

3.3 Metode *Ordinary Least Squares (OLS)*

Metode OLS merupakan metode yang populer dan banyak digunakan untuk menemukan hasil estimasi model regresi dengan meminimumkan Residual Sum of Squares (RSS) [8]. Metode *Ordinary Least Squares (OLS)* digunakan sebagai pendekatan awal untuk menganalisis hubungan antara karakteristik teknis kendaraan dan harga mobil. Model ini melibatkan enam variabel prediktor, yaitu *curbweight*, *enginesize*, *wheelbase*, *carlength*, *citympg*, dan *highwaympg*. Berdasarkan hasil estimasi, diperoleh nilai *R-squared* sebesar 0,8033 dan *Adjusted R-squared* sebesar 0,7974, yang menunjukkan bahwa sekitar 80% variasi harga mobil dapat dijelaskan oleh variabel-variabel prediktor dalam model. Selain itu, hasil uji simultan menghasilkan nilai *F*-

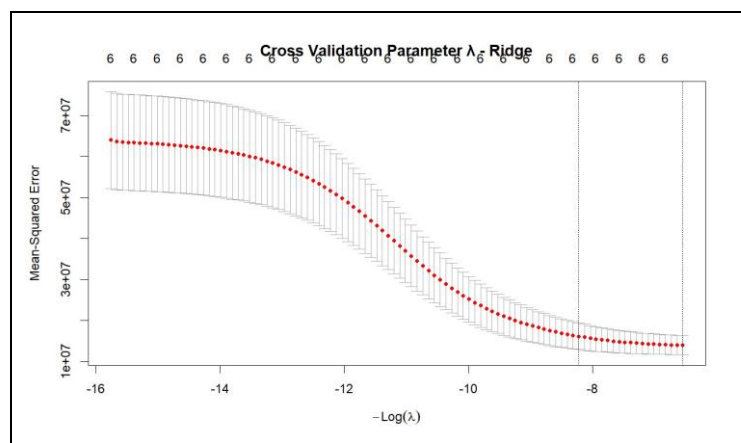
statistic sebesar 134,8 dengan $p\text{-value} < 2,2 \times 10^{-16}$, sehingga model secara keseluruhan signifikan dan layak digunakan untuk analisis.

Secara parsial, variabel *curbweight* ($\beta = 5,133$; $p = 0,00181$) dan *enginesize* ($\beta = 110,163$; $p < 2 \times 10^{-16}$) berpengaruh positif signifikan terhadap harga mobil. Hal ini menunjukkan bahwa kendaraan dengan bobot yang lebih berat serta kapasitas mesin yang lebih besar cenderung memiliki harga yang lebih tinggi. Variabel *citympg* memiliki koefisien negatif signifikan ($\beta = -403,287$; $p = 0,02398$), yang mengindikasikan bahwa semakin efisien konsumsi bahan bakar di dalam kota, maka harga mobil cenderung lebih rendah. Sementara itu, variabel *wheelbase* ($p = 0,36853$), *carlength* ($p = 0,18579$), dan *highwaympg* ($p = 0,12772$) tidak signifikan pada taraf kepercayaan 5%, sehingga kontribusinya terhadap variasi harga mobil relatif lebih kecil dalam model ini.

Model *OLS* menghasilkan nilai *Mean Squared Error (MSE)* sebesar 12.489.548 yang mencerminkan tingkat kesalahan prediksi antara nilai aktual dan nilai estimasi model. Meskipun kinerja prediksi *OLS* tergolong baik, hasil pengujian multikolinearitas dan deteksi outlier pada tahap sebelumnya menunjukkan adanya korelasi tinggi antarvariabel prediktor serta keberadaan observasi ekstrem. Kondisi tersebut berpotensi memengaruhi kestabilan estimasi koefisien *OLS*. Oleh karena itu, pada tahap berikutnya digunakan metode berbasis *regularization*, yaitu *Ridge Regression*, *LASSO*, dan *Elastic-Net*, sebagai pembandingan untuk memperoleh estimasi koefisien yang lebih stabil serta mengevaluasi potensi peningkatan kinerja model.

3.4 Metode Ridge

Metode Ridge digunakan untuk mengatasi permasalahan multikolinearitas pada model *OLS* dengan memberikan penalti *L2-norm* terhadap koefisien regresi sehingga estimasi menjadi lebih stabil. Pada penelitian ini digunakan parameter elastis $\gamma = 0$ yang merepresentasikan regresi Ridge murni. Penentuan nilai parameter penyusutan λ dilakukan dengan *10-fold Cross Validation* menggunakan fungsi *cv.glmnet* pada R dan pendekatan *one-standard-error* (λ_{1se}).

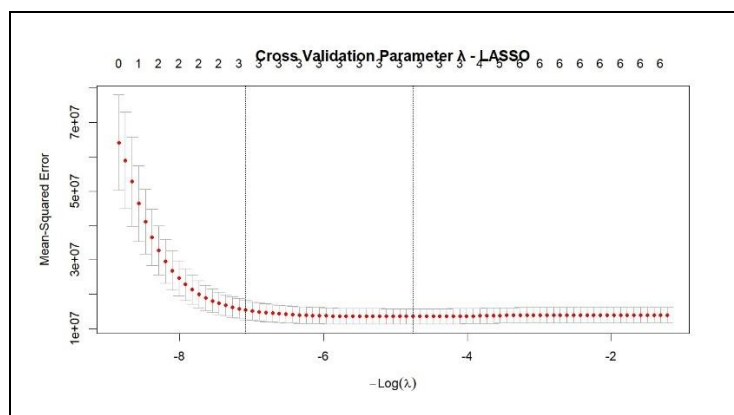


Gambar 4. Cross Validation Parameter λ pada Regresi Ridge

Hasil *cross validation* menunjukkan bahwa nilai λ optimum yang diperoleh sebesar 3717,737 dengan MSE minimum sebesar 13.976.758. Nilai tersebut menggambarkan tingkat kesalahan prediksi model terhadap harga mobil. Regresi Ridge tidak melakukan seleksi variabel sehingga seluruh prediktor tetap digunakan dalam model dengan koefisien yang telah disusutkan. Penyusutan ini bertujuan mengurangi varians akibat multikolinearitas sehingga menghasilkan model yang lebih stabil dibandingkan *OLS*, meskipun nilai MSE yang dihasilkan sedikit lebih besar.

3.5 Metode LASSO

Metode *Least Absolute Shrinkage and Selection Operator (LASSO)* digunakan sebagai pendekatan regularisasi dengan penalti *L1-norm* untuk melakukan penyusutan koefisien sekaligus seleksi variabel. Pada penelitian ini digunakan parameter elastis $\gamma = 1$, sehingga model yang diterapkan merupakan *LASSO* murni. Penentuan parameter penyusutan λ dilakukan dengan *10-fold Cross Validation* menggunakan fungsi *cv.glmnet* pada R serta pendekatan *one-standard-error* (λ_{1se}) untuk memperoleh model yang lebih stabil dan menghindari *overfitting*.

Gambar 5. Cross Validation Parameter λ pada Regresi LASSO

Berdasarkan hasil *cross validation*, diperoleh nilai λ optimum sebesar 1189,403 dengan MSE minimum sebesar 13.576.637. Nilai MSE ini lebih kecil dibandingkan model Ridge, menunjukkan bahwa metode LASSO memberikan performa prediksi yang lebih baik pada data yang dianalisis. Berbeda dengan regresi Ridge, metode LASSO memungkinkan beberapa koefisien menyusut tepat menjadi nol sehingga dilakukan seleksi variabel secara otomatis. Dengan mekanisme tersebut, model yang dihasilkan menjadi lebih sederhana dan hanya mempertahankan prediktor yang paling berkontribusi terhadap respon, sekaligus mengatasi multikolinearitas dan meningkatkan efisiensi interpretasi model.

3.6 Metode Elastic-Net

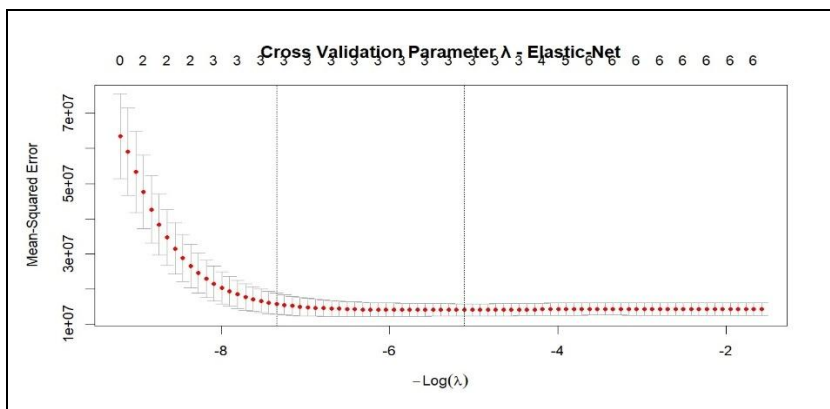
Pemodelan *Elastic-Net* dilakukan dengan mengoptimalkan pasangan parameter elastis γ dan penalti λ untuk setiap peningkatan nilai γ sebesar 0,1. Proses optimasi dilakukan menggunakan *metode 10-fold Cross Validation* (CV) pada setiap nilai γ , dan nilai λ dipilih menggunakan pendekatan *one-standard-error* (λ_{1se}). Evaluasi performa model dilakukan berdasarkan nilai *Mean Squared Error* (MSE) dan jumlah variabel prediktor yang tidak tereliminasi (p). Tabel berikut menyajikan ringkasan performa model untuk setiap nilai γ yang diuji pada rentang 0,1 hingga 0,9.

Tabel 3. Performa Model Berdasarkan Parameter Tuning *Elastic-Net*

α (γ)	λ_{1se}	MSE	p
0,1	3548,760	13.628.895	6
0,2	2345,626	13.782.876	6
0,3	2489,933	13.701.405	5
0,4	1867,450	13.704.275	4
0,5	2167,480	13.742.877	4
0,6	2175,615	13.602.582	3
0,7	1864,812	13.390.278	3
0,8	1486,754	13.557.840	3
0,9	1204,156	13.667.524	3

Tabel 3 merangkum performa model untuk berbagai nilai γ , di mana setiap peningkatan γ sebesar 0,1 menghasilkan pasangan parameter (γ , λ_{1se}) yang berbeda. Dari hasil tersebut terlihat bahwa *Elastic-Net* mulai melakukan seleksi variabel secara lebih agresif ketika γ mencapai nilai 0,4 ke atas, ditandai dengan berkurangnya jumlah variabel prediktor yang tidak tereliminasi (p). Nilai MSE juga memperlihatkan pola yang menurun pada rentang $\gamma = 0,6$ hingga $\gamma = 0,8$, menunjukkan bahwa kombinasi penalti L1 dan L2 memberikan kestabilan model sekaligus mempertahankan akurasi prediksi.

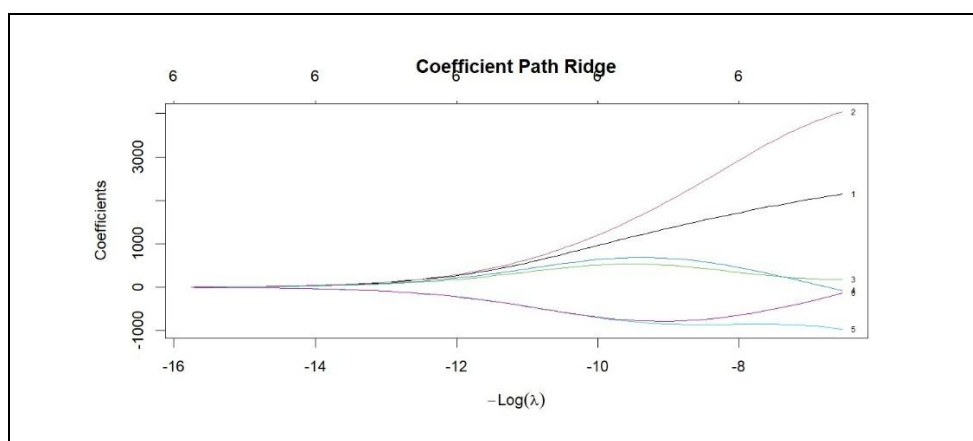
Pasangan parameter terbaik diperoleh pada $\gamma = 0,7$ dengan λ_{1se} sebesar 1864,812 dan nilai MSE terkecil, yaitu 13.390.278. Pada parameter ini, hanya tiga variabel prediktor yang dipertahankan dalam model, sehingga menghasilkan model yang lebih sederhana tanpa mengurangi kualitas prediksi. Kinerja ini lebih baik dibandingkan dengan metode Ridge yang menghasilkan MSE sebesar 13.976.758 dan LASSO dengan MSE sebesar 13.576.637. Dengan demikian, *Elastic-Net* terbukti memberikan performa prediksi terbaik dalam penelitian ini.

Gambar 6. Cross Validation Parameter λ pada Regresi *Elastic-Net*

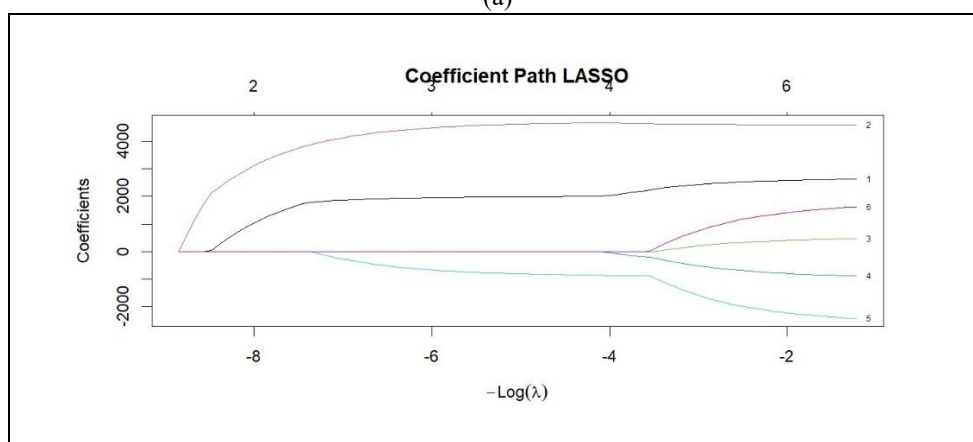
Gambar 6 memperlihatkan proses *cross validation* parameter λ pada *Elastic-Net* dengan $\gamma = 0,7$. Dua garis vertikal menunjukkan nilai $\log(\lambda_{\min})$ dan $\log(\lambda_{1se})$, di mana λ_{1se} dipilih sebagai parameter optimum karena lebih stabil dan cenderung menghasilkan model yang tidak overfitting. Berdasarkan hasil tersebut, metode *Elastic-Net* tidak hanya mampu menyeleksi variabel secara efektif, tetapi juga memberikan kestabilan koefisien akibat adanya penalti L2, sehingga cocok digunakan ketika data memiliki korelasi tinggi antarvariabel.

3.7 Koefisien Regresi dan Perbandingan Model

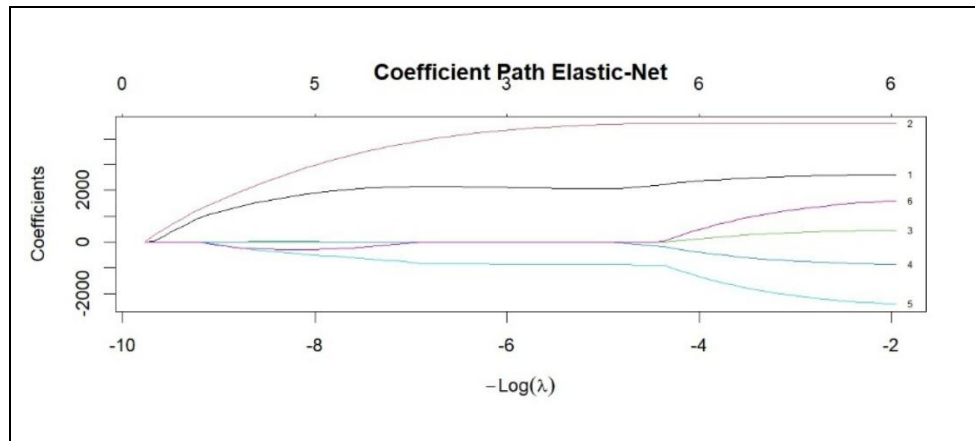
Nilai penyusutan λ yang berbeda memberikan nilai koefisien dan banyaknya variabel prediktor yang berbeda pula untuk setiap modelnya.



(a)



(b)



(c)

Gambar 7. Pergerakan koefisien untuk setiap parameter penyusutan λ pada (a) Regresi Ridge, (b) LASSO, dan (c) *Elastic-Net*

Gambar 7 tersebut menunjukkan jalur pergerakan estimasi koefisien untuk setiap nilai $\log(\lambda)$ pada ketiga metode regularisasi. Setiap garis pada grafik menggambarkan bagaimana koefisien suatu variabel berubah mengikuti besarnya penalti, sesuai dengan tanda positif maupun negatif dari koefisien tersebut. Seiring meningkatnya nilai λ , koefisien semakin disusutkan menuju nilai yang lebih kecil. Pada regresi Ridge, seluruh koefisien bergerak secara halus dan tidak ada yang menyusut hingga tepat bernilai nol, sehingga tidak terjadi eliminasi variabel. Kondisi ini berbeda dengan regresi LASSO di mana beberapa koefisien menyusut sepenuhnya hingga nol ketika penalti diperbesar, menunjukkan adanya proses seleksi variabel yang terjadi secara otomatis. Adapun pada *Elastic-Net*, pola pergerakan koefisien memperlihatkan kombinasi kedua sifat tersebut, yaitu sebagian koefisien tetap menyusut tanpa menghilang seperti pada Ridge, sementara sebagian lainnya tereduksi hingga nol sebagaimana pada LASSO. Dengan demikian, jalur koefisien ketiga model memberikan gambaran yang jelas mengenai pengaruh penalti terhadap struktur parameter serta kemampuan masing-masing metode dalam melakukan penyusutan dan seleksi variabel.

Tabel 4. Nilai Estimasi Koefisien Variabel Prediktor
(berdasarkan output Ridge $\gamma = 0$, LASSO $\gamma = 1$, *Elastic-Net* $\gamma = 0,4$)

Variabel	OLS	Regresi Ridge $\gamma = 0$	LASSO $\gamma = 1$	<i>Elastic-Net</i> $\gamma = 0,4$
Intercept	-6975,771	13276,711	13276,711	13276,711
curbweight	5,133	1641,511	1846,038	1945,488
enginesize	110,162	2717,527	4074,697	3077,196
wheelbase	88,590	382,656	—	—
carlength	-78,327	518,002	—	8,821
citympg	-403,286	-858,234	-224,378	-524,225
highwaympg	266,941	-692,858	—	-278,278

Keterangan: tanda (—) menunjukkan bahwa koefisien tereliminasi (disusutkan menjadi nol) oleh metode regularisasi.

Tabel 4 menunjukkan hasil estimasi koefisien variabel prediktor pada masing-masing metode regresi, yaitu *Ordinary Least Squares (OLS)*, *Ridge Regression* ($\gamma = 0$), LASSO ($\gamma = 1$), dan *Elastic-Net* ($\gamma = 0,4$). Model OLS mempertahankan seluruh variabel prediktor tanpa penyusutan koefisien. Metode Ridge juga mempertahankan seluruh variabel, namun melakukan penyusutan koefisien guna mengurangi pengaruh multikolinearitas. Sebaliknya, LASSO melakukan seleksi variabel sehingga hanya tiga variabel yang tetap digunakan dalam model, yaitu *curbweight*, *enginesize*, dan *citympg*, sementara variabel *wheelbase*, *carlength*, dan *highwaympg* tereliminasi. Metode *Elastic-Net* menghasilkan model dengan empat variabel terpilih, yaitu *curbweight*, *enginesize*, *carlength*, dan *citympg*, sedangkan *wheelbase* dan *highwaympg* disusutkan menjadi nol, mencerminkan keseimbangan antara sifat Ridge yang mempertahankan variabel dan kemampuan seleksi milik LASSO.

Secara umum, tanda koefisien relatif konsisten pada seluruh metode. Variabel *curbweight* dan *enginesize* menunjukkan pengaruh positif terhadap harga mobil, sehingga peningkatan bobot kendaraan dan

kapasitas mesin cenderung meningkatkan harga mobil. Variabel *citympg* memiliki koefisien negatif pada seluruh model, yang mengindikasikan bahwa semakin efisien konsumsi bahan bakar di dalam kota, harga mobil cenderung lebih rendah. Perbedaan utama antar metode terletak pada besaran koefisien serta jumlah variabel yang terpilih ke dalam model akibat penerapan proses regularisasi.

Tabel 4. Perbandingan Kinerja Model

Metode	MSE
OLS	12.489.548
Regresi Ridge	13.976.758
Regresi LASSO	13.834.780
<i>Elastic-Net</i>	13.310.670

Berdasarkan Tabel 4, diperoleh bahwa nilai *Mean Squared Error* (MSE) terkecil dihasilkan oleh model OLS sebesar 12.489.548, sehingga OLS terpilih sebagai model terbaik dalam penelitian ini. Meskipun metode regularisasi Ridge, LASSO, dan *Elastic-Net* dirancang untuk mengatasi multikolinearitas dan melakukan penyusutan koefisien, ketiganya menghasilkan nilai MSE yang lebih besar dibandingkan OLS pada data ini.

Meskipun metode regresi Ridge, LASSO, dan *Elastic-Net* diterapkan untuk mengatasi permasalahan multikolinearitas dan melakukan penyusutan koefisien, berdasarkan nilai *Mean Squared Error* (MSE) diperoleh bahwa model *Ordinary Least Squares* (OLS) terpilih sebagai model terbaik dengan MSE terkecil sebesar 12.489.548, dibandingkan dengan Ridge (13.976.758), LASSO (13.834.780), dan *Elastic-Net* (13.310.670). Hal ini menunjukkan bahwa pada data harga mobil yang digunakan dalam penelitian ini, pendekatan OLS mampu memberikan tingkat akurasi prediksi yang lebih baik dibandingkan ketiga metode regularisasi tersebut. Model regresi terbaik yang diperoleh menggunakan metode OLS dapat dinyatakan sebagai berikut:

$$\hat{y} = -6975,77 + 5,13(\text{curbweight}) + 110,16(\text{enginesize}) + 88,59(\text{wheelbase}) - 78,33(\text{carlength}) - 403,29(\text{citympg}) + 266,94(\text{highwaympg})$$

Berdasarkan persamaan tersebut, variabel *curbweight* dan *enginesize* memiliki koefisien positif signifikan yang menunjukkan bahwa peningkatan berat kendaraan dan kapasitas mesin cenderung meningkatkan harga mobil. Variabel *citympg* memiliki koefisien negatif signifikan yang mengindikasikan bahwa semakin efisien konsumsi bahan bakar di dalam kota, harga mobil cenderung lebih rendah. Sementara itu, variabel *wheelbase*, *carlength*, dan *highwaympg* tidak menunjukkan pengaruh signifikan pada taraf kepercayaan 5% sehingga kontribusinya terhadap harga mobil relatif lebih kecil dalam model ini.

Hasil ini menunjukkan bahwa meskipun terdapat indikasi multikolinearitas dan *outlier* pada data, pendekatan OLS tetap memberikan kinerja prediksi terbaik. Metode regularisasi Ridge, LASSO, dan *Elastic-Net* berperan sebagai pembanding yang membantu mengidentifikasi proses penyusutan koefisien dan seleksi variabel, namun tidak menghasilkan tingkat kesalahan prediksi yang lebih kecil dibandingkan OLS pada penelitian ini.

4. CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa pemodelan harga mobil menggunakan metode *Ordinary Least Squares* (OLS), *Ridge Regression*, LASSO, dan *Elastic-Net* menunjukkan adanya pengaruh multikolinearitas serta keberadaan outlier pada data prediktor, khususnya pada variabel *citympg*, *highwaympg*, dan *curbweight*. Metode regularisasi mampu mengatasi permasalahan tersebut melalui penyusutan koefisien dan seleksi variabel, di mana LASSO dan *Elastic-Net* menghasilkan model yang lebih sederhana dengan jumlah prediktor yang lebih sedikit, sementara Ridge mempertahankan seluruh variabel dengan koefisien yang lebih stabil. Namun, berdasarkan evaluasi kinerja model menggunakan nilai *Mean Squared Error* (MSE), metode OLS memberikan hasil prediksi terbaik dengan MSE terkecil sebesar 12.489.548 dibandingkan dengan Ridge, LASSO, dan *Elastic-Net*. Variabel yang terbukti berpengaruh signifikan terhadap harga mobil adalah *curbweight* dan *enginesize* yang berpengaruh positif, serta *citympg* yang berpengaruh negatif. Hasil penelitian ini menunjukkan bahwa meskipun metode regularisasi efektif dalam mengatasi multikolinearitas dan menyederhanakan model, pendekatan OLS tetap memberikan performa prediksi terbaik pada dataset yang digunakan. Ke depan, penelitian dapat dikembangkan dengan menambahkan variabel lain, menggunakan metode non-linear atau *machine learning*, serta melakukan validasi pada dataset yang lebih besar dan beragam guna meningkatkan generalisasi model prediksi harga mobil.

ACKNOWLEDGMENTS

Penulis mengucapkan terima kasih kepada Ibu Putri Maulidina Fadilah, S.Si., M.Si. selaku dosen pengampu mata kuliah Analisis Regresi atas bimbingan, arahan, serta masukan yang diberikan selama proses penyusunan penelitian ini. Dukungan dan pendampingan beliau sangat membantu dalam memahami penerapan metode regresi serta penyelesaian penelitian ini dengan baik. Selain itu, penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat terselesaikan.

FUNDING INFORMATION

Penelitian ini tidak memperoleh pendanaan dari lembaga, instansi, maupun pihak mana pun. Seluruh kegiatan penelitian dilakukan secara mandiri tanpa dukungan dana eksternal.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	R	D	O	E	Vi
Aulia Zahra Panjaitan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ahmad Badawi	✓							✓		
Risa Rozzaqi Rambe	✓					✓				
Riska Audina					✓	✓				

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**ding

Vi : **V**isualization

CONFLICT OF INTEREST STATEMENT

Penulis menyatakan bahwa tidak terdapat konflik kepentingan, baik dalam bentuk finansial maupun non-finansial, yang dapat memengaruhi secara langsung maupun tidak langsung terhadap hasil dan pelaporan penelitian ini. Penelitian ini dilakukan secara independen tanpa adanya afiliasi, hubungan pribadi, atau kepentingan tertentu yang berpotensi menimbulkan bias dalam proses penelitian.

INFORMED CONSENT

Penelitian ini tidak melibatkan subjek manusia secara langsung maupun penggunaan data identitas pribadi. Data yang digunakan merupakan data sekunder yang bersifat publik dan anonim sehingga tidak memerlukan persetujuan individu secara personal. Dengan demikian, *isu informed consent* tidak berlaku dalam penelitian ini.

ETHICAL APPROVAL

Penelitian ini tidak melibatkan subjek manusia maupun hewan secara langsung. Seluruh data yang digunakan merupakan data sekunder yang bersifat publik dan anonim sehingga tidak memerlukan persetujuan komite etik. Oleh karena itu, penelitian ini tidak memerlukan persetujuan dari *Institutional Review Board* (IRB) atau komite etik setara, dan seluruh proses penelitian dilakukan sesuai dengan ketentuan serta kaidah etika penelitian yang berlaku.

DATA AVAILABILITY

Data yang mendukung temuan dalam penelitian ini merupakan data sekunder yang diperoleh dari sumber publik (*Kaggle*) dan telah diolah oleh penulis. Dataset yang digunakan dapat diperoleh dari sumber aslinya atau tersedia dari penulis korespondensi atas permintaan yang wajar (*upon reasonable request*).

REFERENCES

- [1] M. Kumar, "Car Price Prediction," Kaggle Dataset. [Online]. Available: <https://www.kaggle.com/datasets/hellbuoy/car-price-prediction>
- [2] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970, doi: 10.1080/00401706.1970.10488634.
- [3] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
- [4] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net," *J. R. Statist. Soc.*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
- [5] R. Arisandi, D. Ruhiat, and E. Marlina, "Implementasi Ridge Regression untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data Time Series Klimatologi," *Jurnal Riset Matematika dan Sains Terapan*, vol. 1, no. 1, pp. 1–11, 2021.
- [6] L. Liu, J. Gao, G. Beasley, and S. Jung, "LASSO and Elastic Net Tend to Over-Select Features," *Mathematics*, vol. 11, no. 17, pp. 1–16, 2023, doi: 10.3390/math11173738.
- [7] A. R. Nur, A. K. Jaya, and S. Siswanto, "Comparative Analysis of Ridge, LASSO, and Elastic Net Regularization Approaches in Handling Multicollinearity for Infant Mortality Data in South Sulawesi," *Jurnal Matematika, Statistika, dan Komputasi*, vol. 20, no. 2, pp. 311–319, 2024, doi: 10.20956/j.v20i2.31632.
- [8] H. A. Khoirunissa, A. R. Wijaya, B. Isnaini, and K. Ferawati, "Analisis Faktor-Faktor Penyebab Inflasi di Indonesia Menggunakan Regresi Ridge, LASSO, dan Elastic-Net," *Indonesian Journal of Applied Statistics*, vol. 7, no. 2, pp. 121–130, 2025, doi: 10.13057/ijas.v7i2.96921.

