

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

Oleh:

**Amalia Azzahra<sup>1</sup>**

**Salwa Risda Pratiwi<sup>2</sup>**

**Asruri Mahfuzh Sitorus<sup>3</sup>**

**Aldiwongki Bartolomeus Turnip<sup>4</sup>**

Universitas Negeri Medan

Alamat: Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten  
Deli Serdang, Sumatera Utara (20221).

Korespondensi Penulis: [amaliaazzahra.4242260001@mhs.unimed.ac.id](mailto:amaliaazzahra.4242260001@mhs.unimed.ac.id),  
[salwarisda.4241260005@mhs.unimed.ac.id](mailto:salwarisda.4241260005@mhs.unimed.ac.id), [asruri4241260006@mhs.unimed.ac.id](mailto:asruri4241260006@mhs.unimed.ac.id),  
[aldi.4243260054@mhs.unimed.ac.id](mailto:aldi.4243260054@mhs.unimed.ac.id).

**Abstract.** *Hepatitis C is a chronic liver disease characterized by changes in blood biochemical biomarkers, particularly Alanine Aminotransferase (ALT). This study aims to assess and compare the performance of four regression methods, namely Multiple Linear Regression, LASSO, Ridge, and Elastic Net, in predicting Alanine Aminotransferase (ALT) levels in patients with Hepatitis C. The data used comes from the Hepatitis C Virus (HCV) Prediction dataset, which includes age and liver function biomarker variables consisting of 615 cases. The analysis was performed in RStudio through the stages of data cleaning, identification of important predictors, model building, and performance evaluation using the RMSE and  $R^2$  metrics. Regularization models were compared with linear regression to determine how effective penalties were in improving prediction accuracy and overcoming potential multicollinearity. Each model was compared based on prediction accuracy and the stability of the resulting coefficients. The results of the analysis showed that the regularization approach,*

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

*particularly Ridge Regression, was able to provide more stable and appropriate predictions for the analysis of liver biochemical data in patients with Hepatitis C.*

**Keywords:** *Multiple Linear Regression, LASSO, Ridge Regression, Hepatitis C Virus.*

**Abstrak.** Hepatitis C merupakan penyakit hati kronis yang ditandai oleh perubahan biomarker biokimia darah, khususnya *Alanine Aminotransferase* (ALT). Penelitian ini bertujuan untuk menilai dan membandingkan kinerja empat metode regresi, yaitu Regresi Linier Berganda, LASSO, Ridge, dan Elastic Net, dalam memprediksi kadar *Alanine Aminotransferase* (ALT) pada pasien yang menderita Hepatitis C. Data yang digunakan berasal dari dataset *Hepatitis C Virus (HCV) Prediction* yang mencakup variabel usia dan biomarker fungsi hati yang terdiri dari 615 responden. Analisis dilakukan di RStudio melalui tahapan pembersihan data, identifikasi prediktor penting, pembangunan model, dan evaluasi performa menggunakan metrik RMSE dan  $R^2$ . Model regularisasi dibandingkan dengan regresi linier untuk mengetahui seberapa efektif penalti dalam meningkatkan akurasi prediksi dan mengatasi potensi multikolinearitas. Setiap model dibandingkan berdasarkan ketepatan prediksi dan stabilitas koefisien yang dihasilkan. Hasil analisis menunjukkan bahwa pendekatan regularisasi, khususnya *Ridge Regression*, mampu memberikan prediksi yang lebih stabil dan sesuai untuk analisis data biokimia hati pada pasien Hepatitis C.

**Kata Kunci:** Regresi Linier Berganda, LASSO, *Ridge Regression*, Virus Hepatitis C.

## LATAR BELAKANG

Hati adalah organ padat terbesar dan mempunyai fungsi yang sangat penting untuk tubuh manusia. Salah satu penyakit yang dapat menyerang organ hati yaitu hepatitis. Hepatitis merupakan penyakit yang menyebabkan peradangan pada hati dan dapat mempengaruhi fungsi hati. Penyakit hepatitis umumnya dianggap sebagai infeksi dari beberapa virus yang berbeda. Virus tersebut adalah Virus Hepatitis A, Virus Hepatitis B, Virus Hepatitis C, Virus Hepatitis D, dan Virus Hepatitis E (Nurwananda & Sulaiman, 2022). Di antara berbagai jenis virus hepatitis tersebut, Hepatitis C merupakan salah satu yang patut mendapat perhatian khusus. Hepatitis C disebabkan oleh infeksi virus hepatitis C (HCV). Jenis ini dapat menular melalui proses berbagi pakai jarum suntik dan hubungan seksual. Seperti hepatitis B, jenis ini dapat ditularkan dari ibu yang terinfeksi

ke janinnya. Jenis ini sebagian besarnya bersifat kronis. Namun, tetap dianggap mengkhawatirkan karena belum ditemukan vaksin untuk jenis hepatitis ini (Balai Besar Laboratorium Kesehatan Makassar, 2024).

Sifat kronis dan belum adanya vaksin untuk Hepatitis C memperkuat gambaran tentang betapa seriusnya ancaman virus ini terhadap kesehatan hati. Ancaman ini semakin nyata mengingat cara penularannya. Virus yang menyebabkan penyakit ini berada dalam cairan tubuh manusia yang sewaktu-waktu bisa ditularkan ke orang lain. Memang sebagian orang yang terinfeksi virus ini bisa sembuh dengan sendirinya, namun demikian virus ini akan menetap dalam tubuh seumur hidup. Hepatitis digunakan untuk semua jenis peradangan pada sel-sel hati, yang bisa disebabkan oleh infeksi (virus, bakteri, parasite), obat-obatan (termasuk obat tradisional), konsumsi alkohol, lemak yang berlebih dan penyakit *autoimmune* (Novarina & Santoso, 2018). Kadar *Alanine Aminotransferase* (ALT) merupakan salah satu indikator penting kerusakan sel hati dan sering digunakan sebagai biomarker utama untuk menilai tingkat keparahan infeksi Hepatitis C. *Alanine Aminotransferase* (ALT) (juga disebut alanine transaminase) juga dikenal sebagai alanine transaminase, adalah enzim sitosolik yang mengkatalisis transfer grup alfa-amino dari alanin ke asam alfa-ketoglutarat (Loho & Hasan, 2021).

Dalam konteks ini, model regresi menjadi pendekatan statistik yang umum digunakan untuk memprediksi perubahan kadar ALT berdasarkan sejumlah variabel penjelas. Namun, regresi linier konvensional sering menghadapi kendala seperti multikolinearitas dan stabilitas koefisien. Karena model regresi linear berganda sederhana, itu sering digunakan. Hubungan kausalitas antara satu variabel dependen dan sejumlah variabel independen dimodelkan oleh model regresi linear berganda. Namun, model ini juga dapat digunakan untuk memprediksi variabel dependen. Namun, ada beberapa asumsi yang harus dipenuhi oleh model regresi linear berganda agar dapat diprediksi dan diinterpretasikan dengan baik (Fauzan et al., 2024).

Oleh karena itu, metode regresi modern seperti LASSO, Ridge, dan Elastic Net digunakan untuk meningkatkan akurasi prediksi melalui mekanisme regularisasi. Perbandingan berbagai metode regresi tersebut menjadi penting untuk menentukan model yang paling efektif dalam memprediksi kadar ALT pada pasien Hepatitis C. Dengan demikian, pemilihan metode regresi yang tepat menjadi langkah penting untuk memperoleh model prediksi kadar ALT yang lebih akurat dan stabil. Analisis ini tidak

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

hanya berkontribusi pada pengembangan metode pemodelan yang lebih baik, tetapi juga memberikan manfaat praktis dalam mendukung evaluasi klinis pasien Hepatitis C secara lebih tepat dan berbasis data.

## KAJIAN TEORITIS

### Regresi Linier Berganda

Regresi linier berganda adalah teknik statistik yang digunakan untuk mengukur dan memodelkan hubungan linier antara satu variabel respon (dependen) dengan dua atau lebih variabel prediktor (independen). Model ini banyak digunakan dalam penelitian ekonomi, sosial, dan studi pasar properti karena kemudahannya dalam interpretasi koefisien dan uji signifikansi tiap variabel (Siregar, et al., 2023).

Dalam model regresi linier berganda, variabel dependen dianggap sebagai kombinasi linier dari variabel independen dan digunakan untuk memprediksi hasil dari variabel depeden. Nilai koefisien model diperoleh melalui pengurangan galat kuadrat antara nilai variabel dependen eksperimental dan nilai yang dihasilkan dari model. Struktur dasar model ditunjukkan sebagai berikut (Lu, et al., 2025).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} + \varepsilon_i, i = 1, 2, \dots, n$$

dengan  $i$  mengacu pada individu ke- $i$  dalam populasi,  $y_i$  mewakili nilai-nilai yang diamati, yang sering disebut variabel terikat atau variabel respons,  $x_i$  disebut variabel bebas, variabel penjelas atau regressor,  $\beta_0$  mewakili suku intersep sedangkan  $\beta_1, \beta_2, \dots, \beta_p$  mengacu pada koefisien regresi dan  $\varepsilon_i$  disebut suku kesalahan yang mencakup semua faktor lain yang mempengaruhi variabel terikat selain dari regressor.

### Regresi Ridge

Regresi Ridge adalah salah satu metode yang dapat digunakan untuk mengatasi masalah multikolinieritas melalui modifikasi *Least Square Method* (LSM). Modifikasi ini dilakukan dengan menambahkan konstanta bias yang relatif kecil pada diagonal matriks lambda sehingga koefisien estimasi ridge dipengaruhi oleh besarnya konstanta bias. Regresi Ridge bertujuan untuk meminimalkan varians dan kesalahan standar estimator sehingga uji signifikansi estimator koefisien regresi cenderung menolak. Hal ini berarti estimator regresi secara signifikan berbeda dari nol, atau terdapat efek yang

signifikan antara variabel independen terhadap variabel dependen (Retno et al., 2025). Estimasi parameter dalam Ridge didefinisikan sebagai:

$$\hat{\beta}^{ridge} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (y - \hat{y}_i)^2 + \lambda \sum_{j=1}^k (\hat{\beta}_j)^2$$

### **Regresi *Least Absolute Shrinkage and Selection Operator* (LASSO)**

*Least Absolute Shrinkage and Selection Operator* (LASSO) adalah sebuah metode regresi yang bertujuan untuk mengecilkan koefisien regresi ( $\beta_j$ ) yang memiliki korelasi, hingga mencapai nol atau mendekatinya. Metode ini tidak hanya mengestimasi parameter tetapi juga mengeliminasi variabel yang tidak signifikan, sehingga menghasilkan model yang lebih efisien. Penerapan metode LASSO memerlukan pemahaman masalah yang menjadi fokus dalam penelitian, yaitu faktor apa saja yang berpengaruh terhadap masalah penelitian (Kamila et al., 2025). Estimasi parameter dalam LASSO didefinisikan sebagai:

$$\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (y - \hat{y}_i)^2 + \lambda \sum_{j=1}^k |\hat{\beta}_j|$$

### **Regresi Elastic Net**

Regresi Elastic Net adalah metode regresi linier yang menggabungkan penalti dari *Ridge Regression* (L2) dan *Lasso Regression* (L1). Ini menangani masalah multikolinearitas, yang merupakan variabel prediktor berkorelasi tinggi, dan *overfitting* pada data berdimensi besar. Dengan melakukan seleksi fitur sekaligus menyusutkan koefisien, metode ini memberikan hasil yang lebih akurat dan stabil.

Elastic-Net memberikan penyusutan koefisien dan melakukan seleksi variabel sehingga dapat mengatasi kekurangan dari metode regresi Ridge dan LASSO. Pada Elastic-Net, diberikan kendala dengan parameter penalti *L1*-norm dan *L2*-norm yang besarnya dikontrol oleh parameter yang elastis, yaitu parameter  $\gamma$  (Khoirunissa et al., 2024). Estimasi parameter dalam Elastic Net didefinisikan sebagai:

$$\hat{\beta}^{elastic\ net} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (y - \hat{y}_i)^2 + \lambda \left( \alpha \sum_{j=1}^k |\hat{\beta}_j| + (1 - \alpha) \sum_{j=1}^k (\hat{\beta}_j)^2 \right)$$

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

## METODE PENELITIAN

### Identifikasi Dataset

Penelitian ini merupakan studi komparatif metode kuantitatif menggunakan data sekunder. Data diperoleh dari dataset publik pada Kaggle yang berisi informasi klinis dan biokimia pasien. Dataset ini dapat diunduh di <https://www.kaggle.com/code/anapedralpez/hepatitis-c-virus-hcv-prediction-liver/log>.

Sampel yang dianalisis berjumlah 615 responden dan analisis dilakukan menggunakan RStudio. Variabel yang digunakan pada penelitian ini diberikan pada Tabel 1.

**Tabel 1.** Penjelasan Variabel Dataset

No	Variabel	Keterangan	Tipe
y	ALT	<i>Alanine Aminotransferase</i> , enzim hati yang meningkat jika terjadi kerusakan sel hati.	Riil
X1	Age	Umur	Integer
X2	Sex	Jenis Kelamin (Laki-laki dan Perempuan)	Kategori
X3	ALB	Albumin, protein utama dalam darah yang dapat menunjukkan gangguan fungsi hati.	Riil
X4	ALP	<i>Alkaline Phosphatase</i> , enzim yang meningkat pada gangguan empedu atau kerusakan hati.	Riil
X5	AST	<i>Aspartat Aminotransferase</i> , enzim lain yang mencerminkan kondisi sel hati dan otot naik saat terjadi kerusakan.	Riil
X6	BIL	Bilirubin, Pigmen empedu; kadar tinggi menunjukkan gangguan hati atau saluran empedu.	Riil
X7	CHE	<i>Cholinesterase</i> , enzim yang berkurang pada kerusakan hati kronis atau gagal hati.	Riil
X8	CHOL	Kolestrol, kadar kolesterol total dalam darah hati berperan dalam metabolisme lemak.	Riil
X9	CREA	<i>Creatinine</i> , penanda fungsi ginjal, tetapi juga berkaitan dengan kondisi metabolik pasien.	Riil
X10	GGT	<i>Gamma-Glutamyl Transferase</i> , enzim yang berkaitan dengan kerusakan hati dan konsumsi alkohol.	Riil
X11	PROT	Jumlah total protein dalam darah.	Riil

### Tahapan Pra-proses dan Pemodelan Data

Untuk memperoleh hasil analisis yang sistematis, penelitian ini dilakukan melalui beberapa tahapan utama yang mencakup proses pengolahan data, pemodelan, serta evaluasi performa model. Prosedur analisis data dalam penelitian ini disusun melalui beberapa tahapan utama sebagai berikut:

## 1. Pembersihan *Missing Value*

Pada tahap ini dilakukan pemeriksaan terhadap nilai hilang pada dataset. Tujuannya adalah mencegah bias akibat kehilangan data dan memastikan dataset siap untuk proses pemodelan.

## 2. Konversi Variabel *Dummy*

Variabel kategorikal seperti jenis kelamin dikonversi menjadi variabel dummy sehingga dapat diproses oleh model regresi.

## 3. Pembagian Data (*Train-Test Split*)

Dataset dibagi menjadi dua bagian, yaitu data training sebesar 70% dan data testing sebesar 30%. Data training digunakan untuk melatih model, sedangkan data testing digunakan untuk mengevaluasi kemampuan model dalam memprediksi kadar ALT pada data baru.

## 4. Regresi Linier Berganda Klasik (MKT)

Model ini digunakan sebagai *baseline* untuk membandingkan kinerja metode regularized regression. Model dibangun dengan menggunakan semua variabel prediktor yang relevan.

## 5. Regularized Regression

- 1) *Ridge Regression (L2 Regularization)*, *Ridge Regression* menambahkan penalty berupa jumlah kuadrat koefisien (*L2 norm*) ke fungsi *loss*. Nilai alpha diatur ke 0.
- 2) *Lasso Regression (L1 Regularization)*, *Lasso regression* menambahkan penalty berupa nilai absolut koefisien (*L1 norm*) yang dapat menghasilkan koefisien nol untuk beberapa variabel (*feature selection*). Nilai alpha diatur ke 1
- 3) *Elastic Net Regression*, *Elastic Net* menggabungkan keunggulan *Ridge* dan *Lasso* dengan menggunakan kombinasi *L1* dan *L2* penalties. Nilai alpha diatur ke 0.5 sebagai titik tengah antara *Ridge* dan *Lasso*.

## 6. Validasi Model Menggunakan *10-Fold Cross-Validation*

*Cross-validation* digunakan untuk memilih hyperparameter terbaik (*lambda* dan *alpha*), sekaligus menilai stabilitas model pada data yang berbeda. Teknik ini membagi data training menjadi 10 bagian, melatih model pada 9 bagian, dan menguji pada 1 bagian secara bergantian

## 7. Evaluasi Model dievaluasi menggunakan data testing dengan metrik:

- 1) *RMSE (Root Mean Square Error)*

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

2)  $R^2$  (Koefisien Determinasi)

Model terbaik ditentukan berdasarkan RMSE terkecil serta  $R^2$  terbesar.

8. Perbandingan dan Interpretasi Hasil

Kinerja keempat model dibandingkan untuk menentukan model paling akurat dalam memprediksi kadar ALT. Hasil dibandingkan secara deskriptif serta dijelaskan perbedaan performanya.

## HASIL DAN PEMBAHASAN

### Uji Asumsi Model

Sebelum dilakukan analisis regresi linier berganda, perlu dipastikan bahwa model yang dibangun telah memenuhi asumsi-asumsi dasar regresi. Salah satu asumsi yang diperhatikan dalam penelitian ini adalah independensi antar variabel bebas, yang diuji menggunakan *Variance Inflation Factor* (VIF) sebagai indikator tingkat multikolinearitas antar variabel bebas dalam model regresi. Berikut adalah tabel hasil uji multikolinearitas.

**Tabel 2.** Nilai *Variance Inflation Factor* (VIF) Variabel Independen

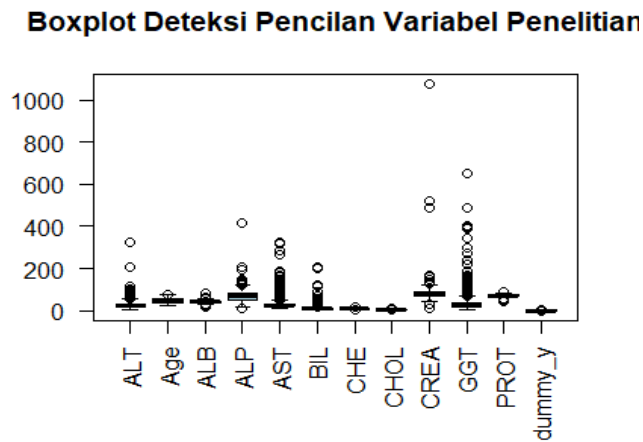
Variabel	VIF
Age	1.11
ALB	1.68
ALP	1.39
AST	1.54
BIL	1.21
CHE	1.47
CHOL	1.37
CREA	1.04
GGT	1.74
PROT	1.61

Seluruh variabel memiliki nilai  $VIF < 5$ , yang menunjukkan tidak terdapat multikolinearitas yang serius dalam model. Dengan demikian, seluruh biomarker hati dapat digunakan secara simultan untuk memodelkan kadar ALT pada pasien Hepatitis C.

### Deteksi Pencilan

Deteksi pencilan dilakukan untuk mengidentifikasi observasi yang berpotensi memberikan pengaruh berlebih terhadap hasil estimasi model regresi. Pada tahap awal, pencilan diamati menggunakan boxplot untuk masing-masing variabel penelitian.

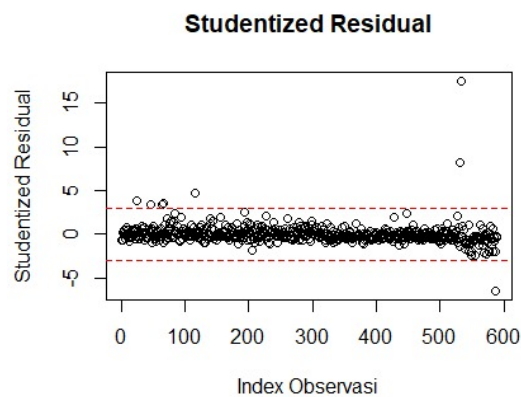
**Gambar 1.** Boxplot Deteksi Pencilan Variabel Penelitian



Berdasarkan boxplot gabungan, terlihat bahwa beberapa variabel biokimia seperti AST, ALP, CREA, dan GGT memiliki sejumlah nilai ekstrem di luar batas *whisker*. Hal ini mencerminkan adanya variasi klinis yang cukup besar antar pasien, khususnya pada indikator fungsi hati dan metabolisme. Meskipun terdapat pencilan, observasi tersebut tidak dihapus, karena merepresentasikan kondisi klinis nyata dan penting dalam analisis penyakit hati.

Selanjutnya, deteksi pencilan diperkuat menggunakan *studentized residual*, yang dihitung dari model regresi linier berganda. Gambar 2 menampilkan sebaran *studentized residual* terhadap indeks observasi, dengan batas ambang  $\pm 3$  yang ditandai oleh garis putus-putus.

**Gambar 2.** Grafik Studentized Residual



# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

## Regresi Linier Berganda

Hubungan antara kadar *Alanine Aminotransferase* (ALT) dan berbagai biomarker fungsi hati dianalisis untuk mengidentifikasi faktor-faktor yang berpengaruh pada pasien Hepatiitis C secara statistik. Pendekatan yang digunakan dalam analisis tersebut adalah regresi linier berganda, dengan ALT sebagai variabel respons dan Age, ALB, ALP, AST, BIL, CHE, CHOL, CREA, GGT, serta PROT sebagai variabel prediktor. Hasil estimasi parameter regresi linier berganda disajikan pada Tabel 3.

**Tabel 3.** Hasil Regresi Linier Berganda

Variabel	Koefisien ( $\beta$ )	Std. Error	p-value
Konstanta	22.62	12.17	0.064
Age	-0.22	0.08	0.008 **
ALB	0.21	0.18	0.236
ALP	0.15	0.04	2.87e-05 ***
AST	0.17	0.03	1.82e-08 ***
BIL	-0.15	0.05	0.00261 **
CHE	1.95	0.43	7.80e-06 ***
CHOL	1.76	0.81	0.02953 *
CREA	-0.03	0.02	0.08289 .
GGT	0.03	0.02	0.10665
PROT	-0.46	0.18	0.01371 *

*Multiple R-squared: 0.1918, Adjusted R-squared: 0.1778*

Secara keseluruhan, model menunjukkan bahwa model memiliki signifikansi secara keseluruhan, dengan nilai *Adjusted R<sup>2</sup>* sebesar 0,178, yang mengindikasikan bahwa variabel-variabel prediktor mampu menjelaskan sekitar 17.8% variasi kadar ALT.

## Evaluasi dan Perbandingan Metode Regresi

Evaluasi kinerja model dilakukan untuk menilai kemampuan prediksi dan generalisasi dari masing-masing metode regresi. Data dibagi menjadi data training dan data testing dengan proporsi 70:30, kemudian kinerja model dibandingkan menggunakan *Root Mean Squared Error (RMSE)* dan *R-Squared (R<sup>2</sup>)* pada data training dan data testing sebagai ukuran tingkat kesalahan prediksi.

**Tabel 4.** RMSE dan Koefisien Determinasi Metode Regresi

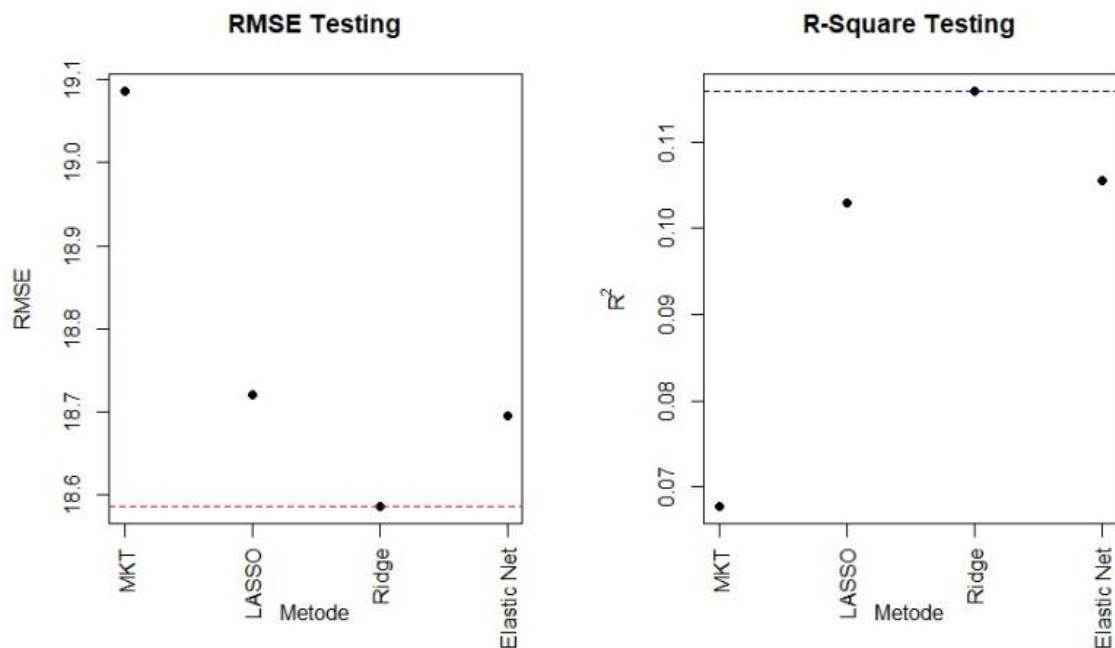
No	Model	RMSE		R <sup>2</sup>	
		<i>Data Training</i>	<i>Data Testing</i>	<i>Data Training</i>	<i>Data Testing</i>

1	Regresi Berganda	Linier	18.912	19.086	0.210	0.068
2	Regresi LASSO		19.827	18.721	0.132	0.103
3	Regresi Ridge		19.807	18.586	0.134	0.116
4	Regresi Elastic Net		19.827	18.695	0.132	0.105

Berdasarkan hasil perbandingan RMSE dan  $R^2$ , *Ridge Regression* menghasilkan kesalahan prediksi terendah dan nilai  $R^2$  tertinggi pada data testing. Hal ini menunjukkan bahwa *Ridge Regression* memiliki akurasi dan kemampuan generalisasi yang lebih baik dibandingkan regresi linier klasik, LASSO, dan Elastic Net dalam memodelkan kadar ALT pada pasien Hepatitis C.

Hal ini juga ditunjukkan secara visual pada Gambar 3, yang memperlihatkan perbandingan nilai RMSE dan *R-Squared* antar metode regresi dengan data testing.

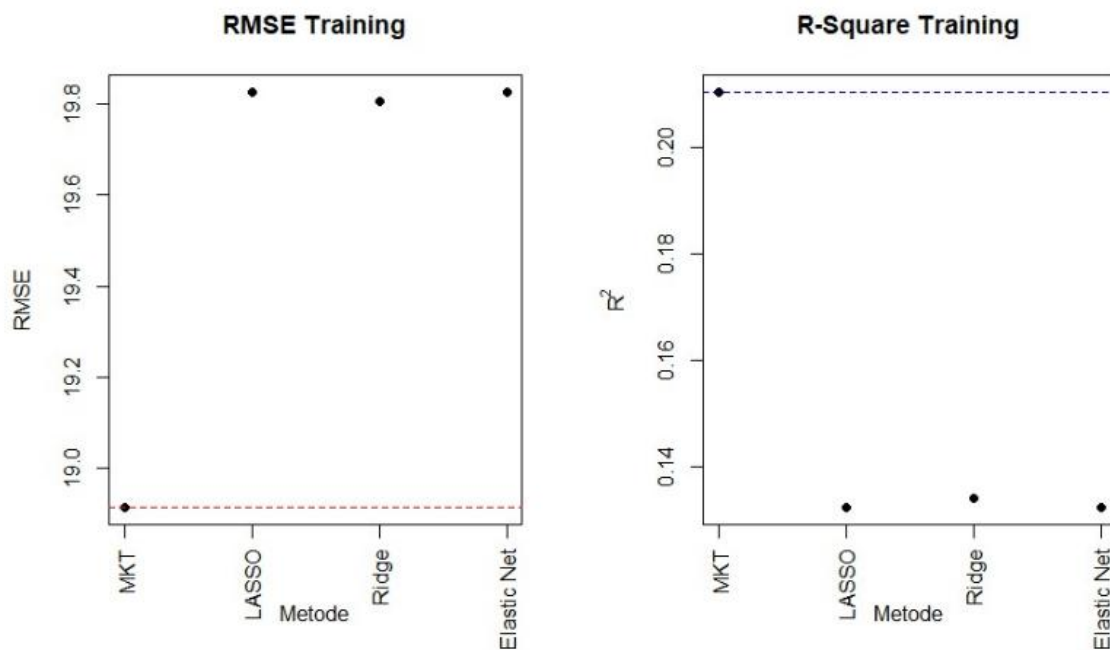
**Gambar 3.** Perbandingan RMSE dan R-square data Testing



# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

Untuk visual berikutnya pada Gambar 4, yang memperlihatkan perbandingan nilai RMSE dan *R-Squared* antar metode regresi dengan data training.

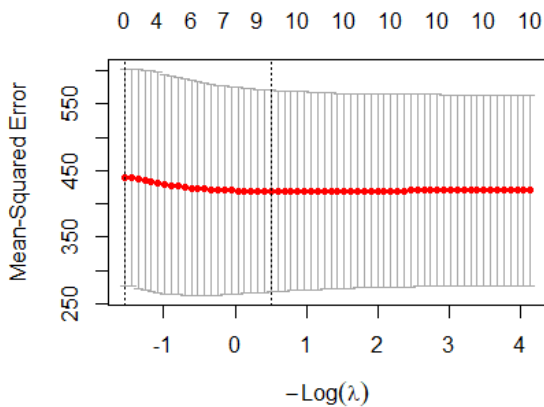
**Gambar 4.** Perbandingan RMSE dan *R-Squared* Data Training



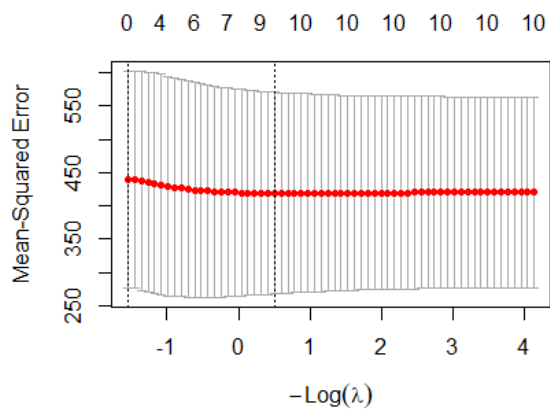
Hasil evaluasi menunjukkan bahwa regresi linier klasik memiliki kinerja terbaik pada data training dengan nilai RMSE terendah dan  $R^2$  tertinggi, namun performanya menurun pada data testing. Sebaliknya, model dengan regularisasi, khususnya *Ridge Regression*, memberikan kinerja paling stabil pada data testing dengan RMSE terendah dan  $R^2$  tertinggi dibandingkan metode lain. Temuan ini mengindikasikan bahwa regularisasi berperan penting dalam meningkatkan kemampuan generalisasi model pada data HCV yang memiliki korelasi antarvariabel, meskipun secara umum nilai  $R^2$  yang diperoleh masih relatif rendah sehingga kemampuan penjelasan model terhadap variasi data terbatas.

Untuk memperjelas proses pemilihan parameter penalti pada metode regularisasi, berikutnya Gambar 5, 6, dan 7 menampilkan hasil *cross-validation* untuk LASSO, Ridge, dan Elastic Net.

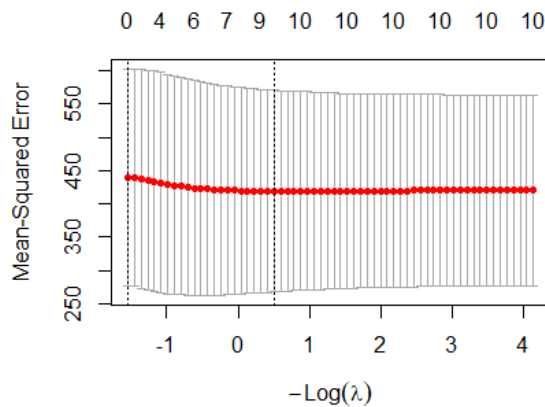
**Gambar 5.** Grafik Cross-Validation LASSO



**Gambar 6.** Grafik Cross-Validation Ridge



**Gambar 7.** Grafik Cross-Validation Net



Grafik *cross-validation* pada metode LASSO, Ridge, dan Elastic Net menunjukkan bahwa pemilihan parameter penalti  $\lambda$  berperan penting dalam mengendalikan kesalahan prediksi. *Ridge Regression* menghasilkan kurva kesalahan yang lebih stabil di sekitar nilai minimum, yang menunjukkan konsistensi performa model pada data uji. Hal ini mendukung hasil perbandingan RMSE yang menunjukkan bahwa *Ridge Regression* memiliki kinerja prediksi terbaik.

# PERBANDINGAN KINERJA REGRESI LINIER BERGANDA, LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI KADAR ALT PADA PASIEN HEPATITIS C

## Pembahasan dalam Konteks Hepatitis C

Hasil penelitian menunjukkan bahwa beberapa biomarker fungsi hati, khususnya AST dan ALP, memiliki pengaruh yang signifikan terhadap variasi kadar ALT pada pasien Hepatitis C. Temuan ini mencerminkan adanya proses kerusakan hepatoseluler yang terjadi akibat infeksi virus secara kronis. Selain itu, variabel CHE dan CHOL juga berperan penting dalam model, yang mengindikasikan terganggunya fungsi sintesis hati sebagai konsekuensi dari penurunan kapasitas metabolik hati pada kondisi Hepatitis C.

Berdasarkan pendekatan analisis regresi, *Ridge Regression* menunjukkan bahwa pendekatan regularisasi efektif dalam menangani korelasi moderat antar biomarker hati yang sering ditemukan dalam data klinis. Pendekatan ini mampu menghasilkan estimasi koefisien yang lebih stabil tanpa menghilangkan variabel yang memiliki makna klinis. Akibatnya, *Ridge Regression* dapat menjadi metode prediktor yang andal untuk menganalisis data laboratorium pasien yang menderita Hepatitis C, terutama untuk pemodelan kadar ALT yang didasarkan pada biomarker hati.

## KESIMPULAN

Berdasarkan hasil analisis terhadap dataset *Hepatitis C Virus (HCV) Prediction*, dapat disimpulkan bahwa biomarker fungsi hati memiliki peran penting dalam memodelkan kadar *Alanine Aminotransferase* (ALT) pada pasien Hepatitis C. Hasil regresi linier berganda menunjukkan bahwa variabel AST, ALP, CHE, CHOL, dan PROT berpengaruh signifikan terhadap variasi ALT, yang mencerminkan proses kerusakan hepatoseluler dan gangguan fungsi sintesis hati akibat infeksi kronis. Evaluasi kinerja model menunjukkan bahwa metode regresi dengan pendekatan regularisasi memberikan hasil prediksi yang lebih baik dibandingkan regresi linier klasik. Di antara metode yang dibandingkan, *Ridge Regression* menghasilkan nilai RMSE terendah dan  $R^2$  terbesar pada data testing, yang mengindikasikan kinerja prediksi yang paling stabil. Hal ini menunjukkan bahwa *Ridge Regression* efektif dalam menangani korelasi antar biomarker hati tanpa menghilangkan variabel yang memiliki relevansi klinis. Dengan demikian, *Ridge Regression* berpotensi digunakan sebagai pendekatan prediktif yang andal dalam analisis data laboratorium pasien Hepatitis C.

## DAFTAR REFERENSI

- Balai Besar Laboratorium Kesehatan Makassar. (2024). *Mengenal Hepatitis*. <https://bblabkesmasmakassar.go.id/mengenal-hepatitis/>
- Fauzan, M. D., Najib, M. K., Nurdiati, S., Maulia, S. D., Roro, R., Triwulandari, C., & Aziz, M. F. (2024). Prediksi Angka Harapan Hidup Menggunakan Regresi Linear Berganda , Lasso , Ridge , Elastic Net , dan Kuantil Lasso. *Jurnal Sains Matematika Dan Statistika*, 10(2), 192–206.
- Kamila, D. R., Imro, N., & Sulistianingsih, E. (2025). *METODE LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) UNTUK PENDUGAAN PARAMETER REGRESI LOGISTIK BINER (Studi Kasus : Faktor-faktor Tingkat Kemiskinan di Indonesia Tahun 2021 )*. 14(1), 57–66.
- 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 Husna. *Indonesian Journal of Applied Statistics*, 7(2), 121–130. <https://doi.org/10.13057/ijas.v7i2.96921>
- Loho, I. M., & Hasan, I. (2021). *Drug-Induced Liver Injury: Tantangan dalam diagnosis*. Laporan ilmiah/Prosiding. <https://media.neliti.com/media/publications/398608-drug-induced-liver-injury-tantangan-dala-fec527b4.pdf>
- Lu, X., Teh, S. Y., Tay, C. J., Kassim, N. F. A., Fam, P. S., & Soewono, E. (2025). Application of multiple linear regression model and long short-term memory with compartmental model to forecast dengue cases in Selangor, Malaysia based on climate variables. *Infectious Disease Modelling*, 10(1), 240-256.
- Novarina, A. T., & Santoso, E. (2018). Sistem Pakar Diagnosis Penyakit Hepatitis Menggunakan Metode Dempster Shafer. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(6), 2252–2258.
- Nurwananda, S. S., & Sulaiman, R. (2022). APLIKASI HIMPUNAN FUZZY INTUISIONISTIK DALAM DIAGNOSA PENYAKIT HEPATITIS MENGGUNAKAN EXTENDED HAUSDORFF DISTANCE. *Jurnal Ilmiah Matematika*, 10(01), 41–49.
- Retno, D., Saputro, S., Wahyu, N. L., & Widyaningsih, Y. (2025). *Performance of Ridge Regression , Least Absolute Shrinkage and Selection Operator , and Elastic Net in Overcoming Multicollinearity*. 5(2), 370–382.

**PERBANDINGAN KINERJA REGRESI LINIER BERGANDA,  
LASSO, RIDGE, DAN ELASTIC NET DALAM MEMPREDIKSI  
KADAR ALT PADA PASIEN HEPATITIS C**

Siregar, R. N. T., Sitorus, V., & Ananta, W. P. (2023). Analisis Prediksi Harga Rumah Di Bandung Menggunakan Regresi Linear Berganda. *Journal of Creative Student Research*, 1(6), 396-404.