

**LAPORAN PRAKTIKUM**  
**MATA KULIAH DATA SCIENCE**  
**PERTEMUAN 10**  
**REGRESI BERGANDA & REGULARISASI**  
**(Ridge, Lasso, Elastic Net)**  
**(Ukuran Pemusatan, Penyebaran, dan Korelasi)**

**Nama :** Peno  
**NIM :** 221220095  
**Kelas :** 29 (DS B)  
**Tanggal Praktikum :** 27 November 2025  
**Tanggal Pengumpulan :** 27 November 2025



**PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNIK  
DAN ILMU KOMPUTER UNIVERSITAS MUHAMMADIYAH  
PONTIANAK**

**2025**

## 1. Pengantar Praktikum

### 1.1 Tujuan Praktikum

1. Mengimplementasikan Ridge, Lasso, dan Elastic Net menggunakan Scikit-Learn.
2. Memahami perbedaan output dan karakteristik dari setiap metode regularisasi.
3. Melakukan *hyperparameter tuning* dengan teknik *Cross-Validation*.
4. Membandingkan performa model OLS (Ordinary Least Squares) dengan model berregularisasi pada dataset nyata.

### 1.2 Dataset yang Digunakan

Dataset yang digunakan adalah **California Housing Dataset** (bawaan Scikit-Learn).

- **Jumlah Sampel:** 20,640 baris.
- **Fitur (8):** MedInc, HouseAge, AveRooms, AveBedrms, Population, AveOccup, Latitude, Longitude.
- **Target:** Median House Price.
- **Karakteristik Data:** Terdapat multikolinearitas antar fitur (misalnya Latitude dan Longitude, atau AveRooms dan AveBedrms).

### 1.3 Teori Singkat

- **OLS (Baseline):** Meminimalkan *error* kuadrat terkecil tanpa penalti. Rentan *overfitting*.
- **Ridge (L2):** Menambahkan penalti kuadrat koefisien. Mencegah koefisien menjadi terlalu besar (shrinkage), cocok untuk mengatasi multikolinearitas.
- **Lasso (L1):** Menambahkan penalti nilai mutlak koefisien. Dapat membuat koefisien menjadi nol (0), sehingga berfungsi sebagai *feature selection* otomatis.
- **Elastic Net:** Kombinasi linier antara Ridge dan Lasso. Menggabungkan keunggulan keduanya.

## 2. Pelaksanaan Praktikum

### 2.1 Persiapan Data (Setup)

Dataset dibagi dengan rasio 80:20 dan dilakukan **Standardisasi** (*StandardScaler*) agar penalti regularisasi bekerja secara adil pada semua fitur.

- **Training set size:** 16,512 sampel
- **Test set size:** 4,128 sampel

### 2.2 OLS Baseline

Model regresi linier standar digunakan sebagai patokan performa.

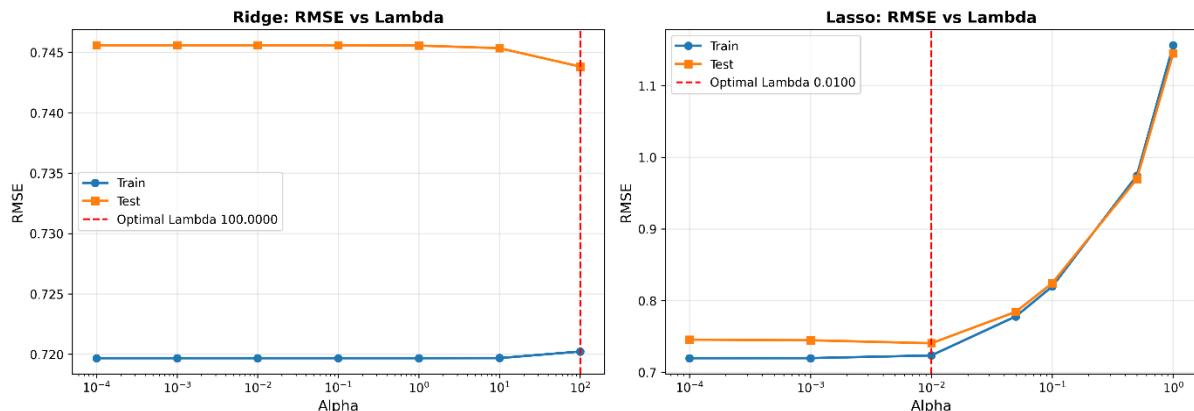
- **Test RMSE:** 0.7456
- **Test R<sup>2</sup>:** 0.5758

- **Analisis:** Model ini memiliki RMSE tertinggi dibandingkan model lainnya, menunjukkan adanya ruang untuk perbaikan melalui regularisasi.

### 2.3 Ridge Regression

Dilakukan tuning nilai alpha (kekuatan regularisasi).

- **Optimal Alpha (RidgeCV):** 0.0100
- **Test RMSE:** 0.7456 (Sangat mirip dengan OLS)
- **Analisis:** Karena alpha optimal sangat kecil (0.01), efek *shrinkage* pada koefisien sangat minim.



Gambar 1. Grafik RMSE vs Lambda untuk Ridge dan Lasso

Keterangan: Grafik kiri menunjukkan kurva Ridge yang landai, sedangkan grafik kanan menunjukkan kurva Lasso yang memiliki titik minimum (lembah) yang jelas.

### 2.4 Lasso Regression

Lasso memiliki kemampuan untuk menihilkan koefisien fitur yang tidak penting.

- **Optimal Alpha (LassoCV):** 0.0006
- **Test RMSE:** 0.7450 (Lebih baik dari OLS & Ridge)
- **Non-zero coefficients:** 8/8 (Semua fitur dipertahankan)
- **Analisis:** Meskipun Lasso bisa melakukan seleksi fitur, pada kasus ini alpha optimalnya sangat kecil (0.0006) sehingga model memilih untuk mempertahankan semua fitur demi akurasi terbaik. Namun, pada percobaan manual dengan alpha=0.1, Lasso berhasil mengeliminasi 5 fitur.

### 2.5 Elastic Net

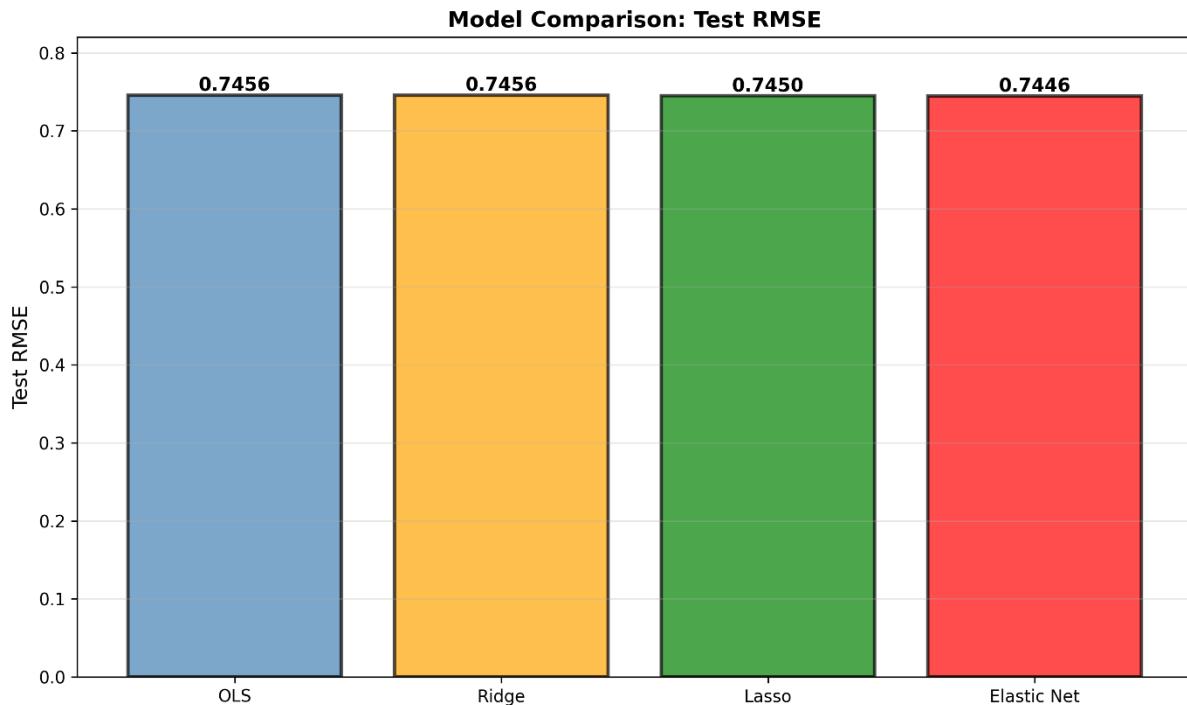
Mencari kombinasi terbaik antara L1 (Lasso) dan L2 (Ridge) menggunakan l1\_ratio.

- **Best L1 Ratio:** 0.99
- **Test RMSE:** 0.7446
- **Analisis:** Rasio 0.99 menunjukkan bahwa model terbaik berperilaku hampir murni seperti Lasso, namun dengan sedikit sentuhan Ridge untuk stabilitas.

### 3. Hasil dan Pembahasan

#### 3.1 Perbandingan Model

Berikut adalah peringkat model berdasarkan error terendah (RMSE) pada data uji:



Gambar 2. Perbandingan Test RMSE Semua Model

| Peringkat | Model       | Test RMSE | Keterangan               |
|-----------|-------------|-----------|--------------------------|
| 1         | Elastic Net | 0.7446    | <b>Performa Terbaik</b>  |
| 2         | Lasso       | 0.7450    | Sangat mendekati terbaik |
| 3         | Ridge       | 0.7456    | Sama dengan OLS          |
| 4         | OLS         | 0.7456    | Baseline                 |

#### 3.2 Analisis & Interpretasi (Menjawab Pertanyaan Modul)

1. Manakah model yang memberikan Test RMSE terbaik? Mengapa?

Model Elastic Net memberikan hasil terbaik dengan RMSE 0.7446.

- **Signifikansi:** Perbedaannya sangat tipis (hanya beda di angka desimal ke-4) dibandingkan OLS (0.7456).
- **Alasan:** Elastic Net menang karena mampu menyeimbangkan bias dan variance dengan sangat presisi, menggunakan kombinasi 99% sifat Lasso dan 1% sifat Ridge. Ini sedikit lebih fleksibel daripada Lasso murni.

2. Bagaimana Ridge mengatasi masalah Multicollinearity?

Ridge mengatasi multikolinearitas dengan melakukan shrinkage (penyusutan) pada koefisien.

- **Bukti:** Nilai *Norm Koefisien* turun dari **1.5842** (OLS) menjadi **1.5841** (Ridge). Penurunan ini kecil karena alpha optimalnya kecil, tetapi secara teori Ridge membagi bobot di antara fitur-fitur yang berkorelasi (seperti Latitude dan Longitude), sehingga model lebih stabil terhadap perubahan data baru.

### 3. Apa yang terjadi dengan Lasso? Fitur mana yang penting?

- **Eliminasi Fitur:** Pada model optimal ( $\alpha=0.0006$ ), Lasso **tidak** mengeliminasi fitur (8/8 fitur aktif). Namun, saat tuning manual dengan alpha lebih besar (misal 0.5), hanya tersisa 1 fitur.
- **Fitur Penting:** Berdasarkan besaran koefisiennya, fitur yang paling berpengaruh terhadap harga rumah adalah **MedInc** (Pendapatan Median) dengan koefisien positif 0.85, serta **Latitude** dan **Longitude** dengan koefisien negatif yang besar (lokasi sangat menentukan harga).

### 4. Kapan Elastic Net lebih baik daripada Pure Ridge atau Pure Lasso?

Elastic Net terbukti lebih baik ketika data memiliki karakteristik campuran yang kompleks. Hasil `best_11_ratio = 0.99` menunjukkan bahwa karakteristik dataset ini sangat cocok dengan pendekatan Lasso (L1), namun penambahan sedikit regularisasi Ridge (L2) membantu menjaga stabilitas model, menghasilkan RMSE yang sedikit lebih rendah daripada Lasso murni.

### 4. Kesimpulan

Berdasarkan praktikum yang telah dilakukan, dapat disimpulkan bahwa:

1. **Regularisasi Berdampak Positif:** Penerapan Lasso dan Elastic Net berhasil menurunkan *error* (RMSE) dibandingkan model OLS biasa, meskipun perbedaannya tipis pada dataset ini.
2. **Pentingnya Scaling:** Regularisasi sangat sensitif terhadap skala data, sehingga proses *StandardScaler* yang dilakukan di awal sangat krusial.
3. **Trade-off Bias-Variance:** Melalui grafik *Lambda vs RMSE*, terlihat pola "U-shape" (terutama pada Lasso) yang menunjukkan bahwa *hyperparameter* ( $\alpha$ ) yang terlalu kecil menyebabkan *overfitting* (mirip OLS), sedangkan  $\alpha$  terlalu besar menyebabkan *underfitting* (error tinggi). Titik optimal berada di antaranya.
4. **Rekomendasi:** Untuk dataset California Housing ini, model **Elastic Net** direkomendasikan karena memberikan akurasi prediksi tertinggi.