

!! Catatan: Diharapkan seluruh penggerjaan Assignment tidak sepenuhnya mengandalkan penggunaan AI !!

"Proses belajar ibarat menanam pohon. Jika hanya mengandalkan AI tanpa memahami esensinya, yang berkembang bukan kompetensimu, melainkan ketergantungan yang melemahkan."

- Learning Design Dibimbang

Assignment Guidance: Case Study

Regression Python

Data Science dan Data Analyst Bootcamp

Periode Pembelajaran

Simple & Multiple Regression Models (Supervised)

Regularized Regression in Python

Objectives

1. Student mampu membagi dataset menjadi set pelatihan, validasi, dan pengujian.
2. Student mampu menggambar correlation plot dan melakukan seleksi fitur yang redundan.
3. Student mampu membangun model regresi Ridge.
4. Student mampu membangun model regresi Lasso.
5. Student mampu melakukan pemilihan model terbaik menggunakan validation set.
6. Student mampu melakukan evaluasi model yang telah dibangun menggunakan metrik evaluasi yang relevan.
7. Student mampu mendokumentasikan dan menginterpretasikan hasil analisis regresi.

Deskripsi Assignment

Assignment ini bertujuan untuk membantu student dalam menerapkan regresi teratur (regularized regression) untuk prediksi harga rumah menggunakan Boston Housing Dataset.

Student akan membangun model Ridge dan Lasso Regression, memilih parameter optimal, dan mengevaluasi kinerja model menggunakan berbagai metrik evaluasi. Analisis ini akan membantu dalam memahami pengaruh variabel terhadap harga rumah serta memilih model terbaik.

Detail Assignment

1. Pengumpulan Data

- Unduh dataset *Boston Housing Data* dari sumber berikut: [Boston Housing Dataset](#).
- Identifikasi variabel dalam dataset yang dapat digunakan untuk memprediksi harga rumah.
- Jelaskan bagaimana data ini relevan dengan studi kasus prediksi harga rumah.

Goals: Memahami struktur dataset dan signifikansi masing-masing variabel.

2. Persiapan Data

- Bagi dataset menjadi set pelatihan, validasi, dan pengujian.
- Pastikan tidak ada nilai yang hilang dalam dataset.
- Lakukan analisis eksplorasi awal untuk memahami distribusi data.

Goals: Memastikan data siap digunakan untuk membangun model prediksi.

3. Analisis Korelasi dan Seleksi Fitur

- Buat correlation plot untuk memahami hubungan antar variabel.
- Identifikasi fitur yang memiliki korelasi tinggi untuk menghindari multicollinearity.
- Lakukan seleksi fitur berdasarkan analisis yang telah dilakukan.

Goals: Menghindari redundansi fitur yang dapat menyebabkan overfitting.

4. Pelatihan Model

- Gunakan regresi Ridge dan Lasso dengan berbagai nilai λ untuk melatih model.
- Lakukan validasi untuk memilih nilai λ terbaik menggunakan metrik RMSE.
- Bandingkan performa model pada set validasi.

Goals: Membangun model regresi dengan regularisasi yang optimal untuk meningkatkan prediksi.

5. Interpretasi Model

- Analisis koefisien dari model terbaik untuk memahami pengaruh fitur terhadap harga rumah.
- Bandingkan hasil model Ridge dan Lasso untuk melihat perbedaan dalam pemilihan fitur.

Goals: Memahami bagaimana setiap fitur berkontribusi dalam prediksi harga rumah.

6. Evaluasi Model

- Gunakan metrik evaluasi seperti MAE, MAPE, dan RMSE untuk menilai performa model.
- Bandingkan performa model Ridge dan Lasso berdasarkan hasil evaluasi.
- Berikan interpretasi terhadap hasil evaluasi model.

Goals: Mengevaluasi performa model dengan metrik yang sesuai dan memahami kekuatan serta kelemahan masing-masing model.

7. Dokumentasi dan Pengumpulan

- Dokumentasikan seluruh langkah yang telah dilakukan secara sistematis dalam Google Colab.
- Simpan hasil kerja dalam Google Colab dan kirimkan link dengan format: **HW_REGRESSION_<YOUR COMPLETE NAME>**.
- Pastikan semua kode dapat dijalankan dan menghasilkan output yang sesuai.

Goals: Mendokumentasikan seluruh proses analisis dan pemodelan secara rapi dan sistematis.

Tools

Google Colaboratory

Pengumpulan Assignment

Deadline :

Maksimal H+7 Kelas (Pukul 23.30 WIB)

Details :

Dikumpulkan dalam bentuk Link Google Colaboratory, secara INDIVIDU, di LMS

Indikator Penilaian

No.	Aspek Penilaian	Parameter	Bobot Maksimal
1	Pengumpulan Data	Memahami struktur dataset dan signifikansi variabel	10
2	Persiapan Data	Memastikan data siap untuk pemodelan	10
3	Analisis Korelasi dan Seleksi Fitur	Menghindari redundansi fitur yang dapat menyebabkan overfitting	10
4	Melatih model regresi Ridge	Membangun model regresi Ridge dengan parameter yang optimal	20
5	Melatih model regresi Lasso	Membangun model regresi Lasso dengan parameter yang optimal	20
6	Pemilihan model terbaik	Memilih model terbaik berdasarkan	15

		validation set	
7	Evaluasi Model	Mengevaluasi dan menginterpretasikan performa model	15

Referensi/Template

1. <https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression>

Sanksi Penggunaan AI:

Apabila student terdeteksi 100% menggunakan AI, maka hasil assignment akan diberikan skor 0

Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor