



Latihan Soal LKS AI UBIG

Prediksi Harga Mobil

- **Tahap 1: EDA**

Peserta diminta untuk **menganalisis dataset CarPrice_Assignment.csv** dan menjawab pertanyaan berikut:

1. Berapa jumlah mobil berdasarkan **jenis bahan bakar (fueltype)** dan **sistem penggerak (drivewheel)**?
2. **Hitung rata-rata harga mobil berdasarkan:**
 1. Jenis bodi (carbody).
 2. Jumlah silinder (cylindernumber).
3. Buat **scatter plot antara horsepower dan price**. Apakah ada hubungan antara tenaga kuda dan harga mobil?
4. Gunakan **heatmap korelasi** untuk melihat hubungan antar fitur. Sebutkan **3 fitur yang paling berpengaruh terhadap harga mobil (price)**.
5. Apakah ada **outlier dalam harga mobil (price)**? Gunakan **IQR method & Z-score method** untuk mendeteksi outlier.



Ada fitur yang hampir tidak berpengaruh pada harga, tapi tetap ada di dataset! Cari tahu fitur mana! 🤔

- **Tahap 2: Data Pre-processing**

Peserta harus **menyiapkan dataset sebelum digunakan dalam Machine Learning**.

1. Pisahkan nama merek dari CarName → Buat kolom baru untuk merek mobil.
2. Ubah fitur kategorikal menjadi numerik menggunakan One-Hot Encoding (fueltype, carbody, drivewheel, dll).
3. Normalisasi fitur numerik (enginesize, horsepower, curbweight, dll) menggunakan StandardScaler.
4. Lakukan feature selection → Hanya gunakan fitur yang paling relevan dengan harga (price).
5. Pisahkan dataset menjadi 80% training set dan 20% testing set.



"CarName" bisa dipecah menjadi merek mobil, dan itu penting! 🚗💡

- **Tahap 3: Implementasi Machine Learning (KNN)**

Peserta akan mengimplementasikan **K-Nearest Neighbors (KNN)** untuk **regresi** dalam memprediksi harga mobil.

1. Bangun model KNN Regressor untuk memprediksi harga (price).



PT UNIVERSAL BIG DATA

Ruko Modern Kav A16-A17, Jl Loncat Indah, Tasikmadu, Kota Malang 65143
No. Telepon 0812-1212-2388, Email : suratkita@gmail.com

2. Gunakan Elbow Method untuk menemukan nilai K terbaik.
3. Evaluasi model dengan Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R^2 Score.
4. Bandingkan performa KNN dengan Linear Regression dan Decision Tree Regressor! Model mana yang paling akurat?
5. Prediksi harga untuk mobil berikut:
 1. Merek: Toyota
 2. Jenis bodi: Sedan
 3. Jumlah Silinder: 4
 4. Kapasitas Mesin: 2000 cc
 5. Tenaga Kuda: 150 HP
 6. Berat Kosong: 2800 lbs
 7. Bahan Bakar: Gas
 8. Sistem Penggerak: FWD
 9. Konsumsi BBM dalam Kota: 25 mpg
 10. Konsumsi BBM di Jalan Tol: 30 mpg



Tidak semua fitur berguna untuk prediksi harga! Hapus fitur yang tidak relevan! 🤖

• Tahap 4: Evaluasi Model

Peserta harus **menganalisis performa model dan mencari cara untuk meningkatkannya.**

1. Bandingkan hasil prediksi dengan harga asli menggunakan scatter plot.
2. Analisis error dalam model → Apakah ada pola kesalahan tertentu?
3. Coba optimasi model dengan tuning hyperparameter atau mencoba fitur tambahan.
4. Kesimpulan: Apakah KNN cukup baik untuk prediksi harga mobil? Jika tidak, model apa yang lebih cocok?



Harga mobil sangat dipengaruhi oleh fitur tertentu, tapi ada juga fitur yang tidak berpengaruh sama sekali! 🤖

Library yang diperbolehkan: **numpy, pandas, matplotlib, seaborn**