



## 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](mailto:suratkita@gmail.com)

### Latihan Soal LKS AI UBIG

#### Laptop Price Prediction – Kesulitan 70%

- **Penjelasan Kolom Dataset**

- **Company:** Nama perusahaan pembuat laptop. Contoh: Dell, HP, Apple.
- **Product:** Nama tipe atau model dari laptop tersebut.
- **TypeName:** Jenis atau kategori laptop. Contoh: Ultrabook, Notebook, Gaming, dll.
- **Inches:** Ukuran layar laptop dalam satuan inci.
- **ScreenResolution:** Resolusi layar laptop, mencakup dimensi piksel dan informasi tambahan seperti touchscreen atau panel IPS.
- **Cpu:** Spesifikasi prosesor yang digunakan, termasuk merk dan kecepatan.  
Contoh: Intel Core i7 2.5GHz.
- **Ram:** Kapasitas RAM. Format aslinya seperti “8GB”, perlu dikonversi ke angka.
- **Memory:** Kapasitas dan tipe penyimpanan (HDD / SSD). Bisa terdapat lebih dari satu jenis penyimpanan, contoh: “128GB SSD + 1TB HDD”.
- **Gpu:** Jenis GPU (kartu grafis) yang digunakan dalam laptop.
- **OpSys:** Sistem operasi yang terinstal di laptop, contoh: Windows 10, macOS.
- **Weight:** Berat laptop, format aslinya “2.1kg” dan perlu dikonversi ke angka.
- **Price\_euros:** Harga laptop dalam satuan Euro. Kolom ini akan digunakan sebagai target klasifikasi, dikategorikan menjadi kelas Low, Medium, dan High.

- **Tahap 1: EDA**

1. Tampilkan struktur dataset (``info()``, jumlah baris/kolom).
2. Identifikasi dan tangani **missing values** dan **duplikasi data**.
3. Hitung statistik deskriptif (``mean``, ``median``, ``min``, ``max``) pada fitur numerik.
4. Visualisasikan distribusi ``Price_euros`` dengan **Histogram** dan **Boxplot**.
5. Tambahkan kolom baru ``Price_Class`` dengan klasifikasi sebagai berikut:
  - Low
  - Medium
  - High
6. Hitung jumlah data untuk masing-masing kelas dan tampilkan dengan **Bar Chart**.
7. Buat korelasi antar fitur numerik dan tampilkan menggunakan **Heatmap**.

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

1. Bersihkan data: tangani missing values, hapus duplikasi.
2. Ubah RAM, Weight, dan Memory ke tipe numerik (GB, KG, dll).
3. Ekstrak kecepatan CPU dari kolom Cpu ke kolom numerik baru (ProcessorSpeed).
4. Pisahkan Memory menjadi SSD dan HDD.



## 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](mailto:suratkita@gmail.com)

---

5. Lakukan encoding manual (misal dengan dict atau replace()) untuk kolom kategorikal seperti Company, OpSys, dll.
6. Normalisasi fitur numerik (gunakan rumus min-max normalization atau z-score manual).
7. Tentukan fitur (X) dan target (y = Price\_Class).
8. Bagi data menjadi training dan testing (contoh: 80:20).

- **Tahap 3: Problem Solving Implementasi (KNN Manual)**

Peserta mengimplementasikan algoritma **K-Nearest Neighbors (KNN)** secara **manual (tanpa library ML)** untuk memprediksi kelas harga laptop. Tugas Anda:

1. Implementasikan KNN secara manual.
  1. Hitung jarak Euclidean antar data.
  2. Ambil k tetangga terdekat.
  3. Gunakan mayoritas voting untuk menentukan kelas (Low, Medium, High).
2. Buat fungsi predict\_knn(data\_training, data\_test, k) untuk klasifikasi.
3. Uji model terhadap data testing.
4. Hitung secara manual:
  - Accuracy
  - Confusion Matrix
  - Precision, Recall, dan F1-Score (rumus manual)

- **Tahap 4: Evaluasi Model**

Peserta melakukan evaluasi model KNN dan eksplorasi pendekatan lain untuk meningkatkan akurasi klasifikasi. Tugas Anda:

1. Tampilkan hasil evaluasi model KNN:
  1. Confusion Matrix
  2. Precision, Recall, F1-Score
2. Coba bandingkan hasil KNN untuk nilai k berbeda (misalnya k = 3, 5, 7) dan simpulkan.
3. Tampilkan visualisasi hasil prediksi vs kelas sebenarnya.
4. Simpulkan:
  1. Berapa nilai k terbaik?
  2. Apa kelebihan & kekurangan model KNN manual?

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

Library yang tidak diperbolehkan: **scikit-learn, tensorflow, keras, xgboost, dll.**