



## 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

#### PREDIKSI NEGARA BERDASARKAN POLA PEMBELIAN (LEVEL hardcore 🧠🔥)

- **Dataset:** Project1-invoice-data.csv
- **Tujuan:** Berdasarkan fitur seperti Quantity, UnitPrice, InvoiceDate, bisa nggak kita klasifikasikan bahwa suatu transaksi itu berasal dari negara tertentu (contohnya UK atau lainnya)?.
- **Kolom dalam Dataset**
  1. **InvoiceNo:** Nomor transaksi.
  2. **StockCode:** Kode barang.
  3. **Description:** Nama barang.
  4. **Quantity:** Jumlah barang yang dibeli.
  5. **InvoiceDate:** Waktu transaksi terjadi.
  6. **UnitPrice:** Harga satuan barang.
  7. **CustomerID:** ID pelanggan.
  8. **Country:** Negara pelanggan (bisa jadi label klasifikasi).
- **Tahap 1: EDA**
  1. Jumlah transaksi per negara
  2. Distribusi Quantity dan UnitPrice
  3. Statistik deskriptif (mean, median, dsb.)
  4. Frekuensi produk populer
  5. Korelasi antar fitur numerik
  6. Visualisasi waktu (InvoiceDate → jam ramai?).
- **Tahap 2: Data Pre-processing**
  1. Drop kolom yang nggak relevan (misal: Description, StockCode)
  2. Feature engineering:
    - Dari InvoiceDate, ekstrak: Jam, Hari, atau Waktu (Pagi/Siang/Sore)
  3. Label encoding untuk Country
  4. Scaling untuk Quantity, UnitPrice
  5. Drop duplikasi & handle missing values
  6. Pisahkan data training/testing (misalnya 80:20 atau sesuai instruksi lomba).
- **Tahap 3: Problem Solving – Implementasi KNN Manual (No Library ML!)**
  1. Implementasikan algoritma KNN manual:
    - Hitung jarak Euclidean dari data baru ke seluruh data training
    - Ambil k tetangga terdekat
    - Voting mayoritas label



## PT UNIVERSAL BIG DATA

Ruko Modern Kav A16-A17, Jl Loncat Indah, Tasikmadu, Kota Malang 65143  
No. Telepon 0812-1212-2388, Email : [suratkit@gmail.com](mailto:suratkit@gmail.com)

---

- **Tahap 4: Evaluasi Model**

1. Evaluasi hasil klasifikasi:

- Confusion matrix
- Accuracy
- Precision
- Recall
- F1 Score

2. Boleh juga uji dengan **beberapa nilai k** untuk lihat pengaruhnya.

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

Library yang dilarang: **Semua library ML seperti xgboost, tensorflow, lightgbm, sklearn.**