



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

Latihan Soal LKS AI UBIG

KLASIFIKASI INVOICE PELANGGAN MENGGUNAKAN KNN MANUAL TANPA LIBRARY ML (LEVEL HARDCORE 🧠🔥)

- **Dataset:** Project1-invoice-data.csv
- **Tujuan:** Memprediksi apakah suatu invoice akan dibayar (Paid) atau tidak (Unpaid) berdasarkan karakteristik pelanggan dan transaksi, menggunakan algoritma KNN manual tanpa library ML.
- **Tantangan:**
 - KNN harus diimplementasikan dari nol.
 - Tidak diperbolehkan menggunakan scikit-learn atau library ML.
 - Peserta harus mengelola fitur kategorikal dan waktu secara mandiri.
- **Kolom dalam Dataset**
 1. **CustomerID:** ID unik pelanggan.
 2. **InvoiceNo:** Nomor invoice transaksi.
 3. **Gender:** Jenis kelamin pelanggan (Male / Female).
 4. **Age:** Umur pelanggan.
 5. **PaymentMethod:** Metode pembayaran (Cash, Debit, E-wallet, dsb).
 6. **InvoiceAmount:** Total nilai belanja (dalam Rupiah).
 7. **InvoiceDate:** Tanggal transaksi.
 8. **Status:** Keterangan status invoice (Paid / Unpaid) → Ini akan dijadikan target klasifikasi.
- **Tahap 1: EDA**
 1. Tampilkan struktur dataset dan jumlah data tiap kelas Quality (Good vs Bad).
 2. Visualisasikan distribusi setiap fitur numerik menggunakan histogram atau boxplot.
 3. Hitung statistik deskriptif (mean, median, std) untuk semua fitur numerik.
 4. Deteksi outlier menggunakan metode Z-score dan tampilkan datanya.
 5. Hitung korelasi antar fitur dan tampilkan dalam bentuk heatmap.
 6. Tulis insight dari hasil analisis yang bisa berguna untuk pemodelan KNN.
- **Tahap 2: Data Pre-processing**
 1. Tangani missing values dan hapus duplikasi jika ditemukan.
 2. Ubah Gender dan PaymentMethod menjadi nilai numerik dengan encoding manual.
 3. Ubah tanggal InvoiceDate menjadi fitur baru, misalnya: DayOfWeek, Month, atau IsWeekend.
 4. Normalisasi fitur numerik menggunakan Min-Max Scaling secara manual.
 5. Ubah kolom Status: Paid → 1, Unpaid → 0.
 6. Bagi dataset menjadi training (80%) dan testing (20%) secara manual.
- **Tahap 3: Problem Solving – Implementasi KNN Manual (No Library ML!)**
 1. Implementasikan fungsi euclidean_distance() antar dua data.



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. Buat fungsi `knn_predict()`:
 - Menghitung jarak terhadap data training
 - Menentukan K tetangga terdekat
 - Mengambil mayoritas label sebagai hasil prediksi
3. Uji nilai $K = 3, 5, 7$ dan bandingkan performanya.
4. Prediksi label untuk data testing dan simpan hasilnya.

- **Tahap 4: Evaluasi Model**

1. Buat confusion matrix dari hasil prediksi (tanpa library ML).
2. Hitung Accuracy, Precision, Recall, dan F1-score secara manual.
3. Bandingkan performa model untuk tiap nilai K dan berikan kesimpulan.

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

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



🔥 **LEVEL HARDCORE – KNN full manual, tanpa jalan pintas. Siap uji logika dan ketelitianmu!**