



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

SIMULASI LKS AI UBIG

KLASIFIKASI PENDAPATAN TOKO KOPI MENGGUNAKAN KNN MANUAL TANPA LIBRARY ML (LEVEL SUPRA HARDCORE 🧠🔥)

- **Dataset:** coffee_shop_revenue.csv
- **Tujuan:** Bangun model klasifikasi KNN dari nol (tanpa sklearn) untuk memprediksi Revenue_Level dari coffee shop. Kategori Revenue_Level:
 - Low = Revenue < Q1
 - Medium = $Q1 \leq \text{Revenue} < Q3$
 - High = Revenue $\geq Q3$.
- **Tantangan:**
 - Tidak diperbolehkan menggunakan scikit-learn atau library ML lainnya.
 - Semua perhitungan distance, prediksi KNN, evaluasi model, dan validasi dilakukan secara manual.
 - Peserta harus memahami implementasi algoritmik KNN dan menguji berbagai skenario nilai K serta metrik jarak.
- **Kolom dalam Dataset**
 1. **Number_of_Customers_Per_Day:** jumlah pelanggan harian.
 2. **Average_Order_Value:** rata-rata transaksi per pelanggan.
 3. **Operating_Hours_Per_Day:** jam operasional toko dalam sehari.
 4. **Number_of_Employees:** jumlah pegawai yang bekerja.
 5. **Marketing_Spend_Per_Day:** pengeluaran harian untuk marketing.
 6. **Location_Foot_Traffic:** estimasi jumlah orang yang melintas di sekitar lokasi.
 7. **Daily_Revenue:** total pendapatan harian (digunakan sebagai target awal klasifikasi).
- **Tahap 1: EDA**
 1. Distribusi tiap fitur (histogram/boxplot).
 2. Korelasi antar fitur (heatmap).
 3. Scatterplot 3D antara Customers, Order_Value, dan Revenue.
 4. Deteksi outlier (IQR dan Z-Score).
 5. Buat analisis naratif: 3-5 insight dari data.



- **Temukan fitur yang berkorelasi tinggi namun memberi insight berbeda.**
- **Apakah ada fitur yang terlihat "tidak penting" tapi justru menentukan?**



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

• Tahap 2: Data Pre-processing

1. Tangani missing value dan duplikasi.
2. Buat kolom Revenue_Level (Low, Medium, High).
3. Lakukan normalisasi manual untuk semua fitur numerik.
4. Buat fitur baru Customer_Value = Customers × Order_Value.
5. (Bonus) Buat fitur Revenue_per_Employee.
6. Split data 80:20 tanpa train_test_split.
7. Implementasikan PCA manual untuk mengurangi dimensi, dan bandingkan hasilnya.

• Tahap 3: Problem Solving – Implementasi KNN Manual (No Library ML!)

1. Implementasi fungsi:
 - euclidean_distance(a, b)
 - manhattan_distance(a, b)
 - get_k_nearest_neighbors(...)
 - predict_class(...)
2. Prediksi dengan nilai k = 3, 5, 7, 9
3. Bandingkan hasil prediksi dengan berbagai kombinasi:
 - Normalisasi + Euclidean
 - Tanpa Normalisasi + Manhattan
 - PCA + Euclidean



- Apakah kecilnya k membuat model overfit?
- Apa efek PCA terhadap akurasi?
- Adakah fitur berlebihan (redundant)?

• Tahap 4: Evaluasi Model

1. Buat **confusion matrix** manual.
2. Hitung:
 1. Accuracy
 2. Precision
 3. Recall
 4. F1-Score
3. Analisis:
 1. Mana kelas yang paling sering diprediksi salah?
 2. Apakah terjadi bias terhadap kelas tertentu?



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

Final Boss: Real-World Prediction

Berikut data real dari coffee shop yang baru dibuka:

```
Number_of_Customers_Per_Day = 160
Average_Order_Value = 8.2
Operating_Hours_Per_Day = 10
Number_of_Employees = 3
Marketing_Spend_Per_Day = 130
Location_Foot_Traffic = 840
```

1. Prediksikan Revenue_Level dengan KNN buatanmu.
2. Jelaskan **mengapa** model memilih level tersebut.
3. **Bonus:** Buat visualisasi pembandingan dengan 5 data training terdekat.

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

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



NO SKLEARN. NO SHORTCUTS. 🔥 LEVEL: SUPRA HARDCORE – Untuk yang benar-benar tahan uji logika dan stamina ngoding.