



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

Evaluasi & Strategi Persiapan Akhir LKS AI Provinsi

1. Kebutuhan Minimum: Skill Teknis yang Harus Dimiliki

A. Pemahaman Algoritma KNN dari Nol

- Bisa menjelaskan dengan runut: cara kerja KNN, perhitungan jarak, pengambilan K tetangga, dan proses voting.
- Implementasi euclidean_distance() secara manual.
- Punya fungsi knn_predict() yang fleksibel untuk testing set banyak baris.

B. Data Normalization Manual

- Harus paham dan bisa terapkan:
 - **Min-Max Scaling**
 - **Z-Score Normalization**
- Tahu kenapa normalisasi penting buat KNN (karena sensitif terhadap skala fitur).

C. Split Data Secara Manual

- Paham cara membagi data training dan testing (misalnya pakai slicing).
- Harus bisa menghitung rasio dengan benar, misalnya 80:20 atau 60:20:20 (train:val:test).

D. Evaluasi Model Tanpa Bantuan Library

- Harus bisa membuat Confusion Matrix, Accuracy, Precision, Recall, dan F1-score secara manual dari y_true dan y_pred.
- Tahu kapan harus pakai F1 daripada Accuracy (jika data imbalance).

E. Representasi Masalah & Kreativitas

- Paham bagaimana membuat klasifikasi (misal: Low, Medium, High) dari data numerik → pakai kuartil atau domain knowledge.
- Bisa menjelaskan secara logis kenapa data dipotong atau direkayasa (feature engineering).
- Misal: fitur Revenue_per_Customer dari Daily_Revenue / Number_of_Customers.

2. Kesalahan Umum Selama Bootcamp (yang HARUS dihindari)

Berikut beberapa kekurangan yang sering terlihat dan bisa jadi **penghambat skor maksimal**:

1. ✗ Terlalu fokus pada coding, lupa analisa.

→ Nilai “analytical solutions” dan “proposed solutions” jadi kosong.

2. ✗ Tidak ada pembagian kerja dalam tim.

→ 2 orang kerja satu laptop, tapi tidak efektif. **Solusi:**

- 1 orang fokus coding / implementasi
- 1 orang bantu analisis, mencatat insight, menyiapkan kesimpulan evaluasi.



3. **✗ Tidak terbiasa menjelaskan hasil analisis.**
→ Harus bisa jelaskan: “Kenapa fitur ini penting?”, “Kenapa data ini perlu dibuang?”, “Apa insight yang ditemukan dari distribusi data?”
4. **✗ Terlalu banyak waktu habis di tahap awal (EDA atau Preprocessing)** → Ingat: **6 jam di tahap Machine Learning**, ini waktu emas!
Harus sempat bereksperimen dengan nilai K berbeda, dan menjelaskan kenapa pilih K tertentu.

3. Strategi Optimal untuk Lomba (2 Hari, 4 Tahap, Offline Mode)

Tahap	Durasi	Strategi Ideal
EDA	2 Jam	Cepat analisa fitur & visualisasi. Simpan insight untuk digunakan nanti. Fokus juga ke deteksi outlier.
Preprocessing	2 Jam	Lakukan normalisasi & encoding manual. Buat fitur turunan jika sempat. Split data manual.
Machine Learning (KNN)	6 Jam	Fokus implementasi fungsi KNN sendiri. Uji beberapa nilai K, tampilkan hasil evaluasinya.
Evaluasi Model	2 Jam	Buat metrik evaluasi sendiri (manual). Simpulkan hasil. Beri usulan perbaikan / analisis tambahan.

4. Tips Hadapi Mode Offline (Tanpa Internet)

- Sudah hafal struktur KNN dan rumus evaluasi.
- Sudah siapkan template fungsi manual KNN di kepala atau catatan.
- Harus bisa debug tanpa ChatGPT/Stack Overflow.
- Pastikan tim memahami pembagian peran & kolaborasi via LAN.
- Buat dokumentasi model + insight secara manual di file .txt/.md/.ipynb selama kerja.

5. E. Target Skor Tertinggi (Nilai 4 dari Juri)

- Untuk dapat nilai “4”:
 - Masalah direpresentasikan dengan baik
 - Solusi analitis berbasis data dilakukan (EDA, korelasi, feature importance)
 - Proposed solution masuk akal, dibuktikan dengan hasil model, dan ada penjelasan/logika kenapa memilih K tertentu, atau kenapa fitur tertentu dibuang/digunakan
 - Tambahan: fitur turunan, insight unik, hasil evaluasi yang tajam



CHECKLIST PERSIAPAN AKHIR LKS AI 2025 – KNN MANUAL (LEVEL PROVINSI)

1. UMUM

- Mengerti alur lomba (2 hari, 4 tahap: EDA, Preprocessing, ML, Evaluasi).
- Paham aturan offline: tanpa internet, hanya LAN sharing file.
- Tim sudah latihan kerja berpasangan dan bagi tugas (1 coding, 1 analisis/evaluasi).
- Laptop sudah bersih dari koneksi internet dan alat komunikasi.

2. Tahap I – EDA (2 Jam)

- Menampilkan jumlah data, tipe data, dan distribusi label.
- Visualisasi minimal 3 fitur numerik (histogram, boxplot).
- Korelasi fitur ditampilkan (heatmap atau matriks korelasi).
- Outlier dicek (Z-score / IQR).
- Insight dari data dicatat (hubungan antar fitur dan target).

3. Tahap II – Preprocessing (2 Jam)

- Missing values dicek dan ditangani.
- Fitur kategorikal diencoding secara manual.
- Fitur numerik dinormalisasi manual (Min-Max / Z-Score).
- Label dikonversi menjadi angka jika perlu (contoh: Good = 1, Bad = 0).
- Split data training & testing dilakukan tanpa sklearn.
- Simpan hasil preprocessing dalam variabel yang rapi.

4. Tahap III – KNN Manual (6 Jam)

- Fungsi euclidean_distance() dibuat sendiri.
- Fungsi knn_predict() dibuat (input: training set, data uji, nilai k).
- Eksperimen minimal 3 nilai k (contoh: k=3, 5, 7).
- Hasil prediksi untuk seluruh data testing dicatat.
- Feature baru (jika ada) diuji apakah meningkatkan akurasi.

5. Tahap IV – Evaluasi Model (2 Jam)

- Confusion matrix dihitung tanpa sklearn.
- Manual hitung Accuracy, Precision, Recall, F1-score.
- Membandingkan hasil evaluasi tiap nilai k.
- Simpulkan performa model, dan beri saran perbaikan jika perlu.

6. Bonus + Nilai Tambah

- Menyertakan feature engineering atau transformasi kreatif.
- Menuliskan analytical insight dan alasan pemilihan fitur / metode.
- Memberikan saran solusi dari hasil model (proposed solution).
- Menyediakan hasil akhir dalam bentuk presentasi / laporan sederhana.



PT UNIVERSAL BIG DATA

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



Kalau nanti hasilnya belum sesuai harapan, bukan berarti kalian gagal. Mungkin kalian sedang disiapkan untuk sesuatu yang lebih besar. Yang penting: kalian sudah ngoding, bukan ngeluh.