



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

Knn Bank Marketing

- **Tahap 1: EDA**

Peserta diminta untuk mengeksplorasi dataset bank.csv dan menjawab pertanyaan berikut:

1. Berapa jumlah total data dan berapa banyak kolom yang tersedia?
2. Apakah ada nilai yang kosong atau hilang dalam dataset?
3. Hitung statistik dasar (mean, median, min, max) untuk kolom numerik (age, balance, duration, dll).
4. Tampilkan distribusi kelas pada kolom target (y).
5. Buat visualisasi histogram atau boxplot untuk melihat distribusi age, balance, dan duration.
6. Apakah ada outlier dalam data? Gunakan metode IQR untuk mendeteksi outlier pada kolom balance.

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

Berdasarkan hasil EDA, lakukan data pre-processing untuk menyiapkan dataset sebelum digunakan dalam Machine Learning.

1. Ubah kolom kategorikal (job, marital, education, dll) menjadi nilai numerik menggunakan teknik encoding.
2. Normalisasi atau standarisasi data numerik (age, balance, duration) agar skala lebih seimbang.
3. Pisahkan dataset menjadi fitur (X) dan target (y).
4. Bagi dataset menjadi training set (80%) dan testing set (20%).

- **Tahap 3: Implementasi Machine Learning (KNN)**

Setelah melakukan pre-processing, peserta akan mengimplementasikan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi.

1. Bangun model KNN untuk memprediksi apakah seseorang akan membuka deposito (y).
2. Coba beberapa nilai K dan gunakan Elbow Method untuk menemukan nilai K terbaik.
3. Tampilkan hasil prediksi menggunakan Confusion Matrix.
4. Bandingkan performa model dengan metrik Accuracy, Precision, Recall, dan F1-Score.

- **Tahap 4: Evaluasi Model**

Peserta diminta mengevaluasi performa model dan mencari cara meningkatkan akurasinya.

1. Analisis confusion matrix dan interpretasikan hasilnya.
2. Bandingkan performa KNN dengan algoritma lain (misalnya, Decision Tree atau Logistic Regression).
3. Coba teknik feature selection atau hyperparameter tuning untuk meningkatkan performa.
4. Buat kesimpulan akhir: Apakah model sudah cukup baik untuk digunakan?

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