

Proyek pengganti UAS MK Data Mining Kelas B 2025

Analisis dan Pemodelan Dataset Online Retail

Deskripsi Proyek:

Proyek akhir ini memberikan kesempatan kepada mahasiswa untuk menerapkan teknik data mining pada dataset nyata menggunakan Online Retail II Dataset dari UCI Machine Learning Repository yang dapat diakses melalui:

<https://archive.ics.uci.edu/dataset/502/online+retail+ii>

Dataset ini berisi catatan transaksi penjualan sebuah perusahaan retail dan mencakup informasi terkait invoice, produk, pelanggan, harga, kuantitas, serta waktu transaksi.

Ruang Lingkup Project:

1. Supervised Learning

Mahasiswa diminta untuk membangun model supervised learning berdasarkan target variabel yang **dibentuk sendiri** menggunakan atribut yang tersedia. Mahasiswa harus:

- Menentukan target variabel (kelas atau nilai prediksi) yang relevan dan dapat dibangun secara logis dari data.
- Menjelaskan alasan pemilihan target tersebut beserta argumen pendukung dari sisi data maupun konteks bisnis.
- Menjelaskan variabel / fitur apa saja yang digunakan, termasuk pertimbangan pemilihannya.
- Memilih dua model supervised learning (misalnya Artificial Neural Network, Logistic Regression, Random Forest, SVM, Linear Regression, dan lainnya sesuai kebutuhan tugas).
- Melakukan pelatihan dan evaluasi kedua model tersebut menggunakan metrik evaluasi yang sesuai dengan jenis problem (klasifikasi atau regresi).
- Membandingkan hasil kedua model dan memberikan interpretasi terhadap performanya.

2. Unsupervised Learning

Selain supervised learning, mahasiswa boleh melakukan pemodelan dengan unsupervised learning. Dalam bagian ini mahasiswa dapat:

- Mengolah dataset sehingga dapat dibentuk representasi yang sesuai untuk clustering atau association rule mining (ARM).
- Menerapkan metode unsupervised learning seperti k-means, hierarchical clustering, DBSCAN, atau teknik ARM seperti apriori atau FP-Growth.
- Menemukan pola, struktur, atau informasi yang tidak terlihat dari data mentah.
- Memberikan interpretasi yang jelas mengenai karakteristik cluster atau pola yang ditemukan.

Komponen Laporan

Laporan akhir disusun dalam format PDF dan harus memuat:

- a. Deskripsi dataset dan proses pembersihan data.
- b. Penjelasan pembentukan target variabel dan pertimbangan pemilihannya.
- c. Pemilihan fitur dan justifikasi (alasan pemilihan fitur).
- d. Proses pembangunan dua model supervised learning lengkap dengan evaluasi dan perbandingan kinerja.
- e. Proses pemodelan unsupervised learning beserta interpretasi hasil (jika dilakukan).
- f. Kesimpulan yang merangkum temuan utama dari kedua pendekatan.

Laporan akhir dapat mengikuti format berikut:

- Cover
- Ringkasan eksekutif
- Bab 1: Pendahuluan (berisi latar belakang, rumusan masalah, tujuan, manfaat)
- Bab 2: Metodologi (berisi poin a-e selain hasil evaluasi dan interpretasi)
- Bab 3: Hasil (berisi hasil dari poin d dan e)
- Kesimpulan: poin f
- Referensi, jika ada
- Lampiran: berisi link source code dan lampiran lainnya.

Ketentuan Tugas

- Tugas dilakukan secara kelompok beranggotakan 2 orang.
- Seluruh kode program wajib dikumpulkan dalam folder terpisah / link yang dituliskan pada laporan.
- Analisis harus orisinal dan setiap bentuk plagiarisme akan dikenakan sanksi akademik sesuai peraturan yang berlaku. Penggunaan generative AI harus dijelaskan pada bagian akhir laporan.

Kriteria Penilaian

Aspek Penilaian	Bobot
Eksplorasi dan pembersihan data	20%
Pemilihan target variable, atribut dan algoritma	20%
Pembuatan dan evaluasi model	30%
Kerapian, argumentasi, dan sistematika laporan	30%

---Selamat Bekerja---