[T02] Tugas Individu - Menghapus Duplikasi dan Menangani Missing Value

Deskripsi Singkat

Kerjakan tugas mandiri untuk memahami dan menerapkan **teknik dasar pembersihan data (data cleaning)** menggunakan **NumPy dan Pandas**.
Fokus utama tugas ini adalah:

1. Mengidentifikasi dan **menghapus data duplikat**,
2. Mendeteksi dan **menangani data hilang (missing values)** menggunakan berbagai strategi pengisian,
3. Menulis laporan singkat bergaya **IEEE conference** di Overleaf mengenai hasil dan analisis pembersihan data.
4. Tautan template: https://www.overleaf.com/latex/templates/ieee-conference-template/grfzhhncsfqn

Dataset

Gunakan dataset dari repo dosen:
https://github.com/virgantara/machine-learning-course/blob/master/week05/data/data_kotor.csv

Aturan & Batasan

- **Gunakan hanya:** numpy, pandas
- **Tidak boleh** memakai library tambahan (e.g., sklearn, missingno, fancyimpute, dll)
- Kode harus **reproducible**, rapi, dan diberi komentar.
- Simpan hasil *cleaning* dalam file CSV baru:
  student_scores_cleaned.csv
- Laporan wajib dibuat dengan **template IEEE Conference (Overleaf)**.
- **Penamaan file tidak sama dengan format, nilai = 0**
- **Terlambat mengumpulkan, nilai = 0**

Output yang Dikumpulkan

1. **Kode Python:**
   T02_NIM_cleaning.py
2. **Laporan PDF IEEE-style:**
   Minimal 3 halaman berisi:
   - Abstract
   - Introduction
   - Methods (deteksi & penghapusan duplikasi, teknik imputation)
   - Results (tabel ringkasan & cuplikan hasil cleaning)
   - Discussion (analisis dampak cleaning)
   - Conclusion
3. CSV hasil bersih: student_scores_cleaned.csv
4. Penamaan laporan, T02_NIM_Laporan.pdf
5. **Taruh ketiga file tersebut dalam satu file zip dengan format T02_NIM.zip**

Spesifikasi Eksperimen:
1. Deteksi dan Penghapusan Duplikasi
2. Menangani Missing Values
3. Validasi Data

Struktur Laporan (IEEE Format)

Bagian	Isi yang Diharapkan
Abstract	Ringkasan tujuan cleaning dan hasil utama
Introduction	Pentingnya data cleaning dalam ML
Methods	Langkah-langkah duplikasi dan imputasi
Results	Jumlah data sebelum/sesudah cleaning, tabel perbandingan
Discussion	Analisis pengaruh pembersihan data terhadap kualitas dataset
Conclusion	Ringkasan hasil dan rekomendasi

Rubrik Penilaian Tugas Individu [T02]

Aspek	Deskripsi	Kriteria Penilaian	Bobot (%)
1. Implementasi Kode (Cleaning)	Kode mampu menghapus duplikasi, mendeteksi & menangani missing value sesuai metode.	**Excellent (90–100):** Semua langkah berfungsi, hasil sesuai logika, kode bersih dan terdokumentasi. **Good (75–89):** Ada kekurangan minor, tapi hasil benar. **Fair (60–74):** Hanya satu teknik cleaning diterapkan. **Poor (<60):** Kode error atau tidak sesuai instruksi.	40
2. Eksperimen & Hasil (CSV & Output)	File hasil cleaning valid, jumlah data konsisten, perubahan terdokumentasi.	**Excellent:** Hasil lengkap & benar, file CSV bersih. **Good:** Minor kesalahan. **Fair:** Hasil tidak konsisten. **Poor:** Tidak ada output valid.	20
3. Analisis Data Cleaning	Mahasiswa menjelaskan dampak penghapusan duplikasi & imputasi terhadap dataset.	**Excellent:** Analisis mendalam (mis. perubahan distribusi, potensi bias). **Good:** Analisis logis tapi singkat. **Fair:** Hanya menjelaskan langkah tanpa analisis. **Poor:** Tidak ada pembahasan.	20
4. Kualitas Laporan IEEE (LaTeX)	Struktur sesuai IEEE, 4 halaman, bahasa formal, tabel/gambar jelas.	**Excellent:** Lengkap dan rapi. **Good:** Struktur benar tapi kurang halus. **Fair:** Tidak lengkap atau format salah. **Poor:** Tidak pakai format IEEE.	15
5. Kerapihan & Etika Akademik	Penamaan variabel, komentar kode, orisinalitas, gaya penulisan.	**Excellent:** Kode rapi dan orisinal. **Good:** Cukup rapi. **Fair:** Minim komentar.	5

Aspek	Deskripsi	Kriteria Penilaian	Bobot (%)
		Poor: Plagiasi.	