

# **Tugas Besar Penambangan Data**



**Mata Kuliah: Penambangan Data (IF25-32025)**

Dosen Pengampu : Meida Cahyo Untoro, S.Kom., M.Kom.

**Institut Teknologi Sumatera  
Fakultas Teknologi Industri  
Teknik Informatika  
2025**

## 1. Latar Belakang

Pada tahap pembelajaran sebelumnya, mahasiswa telah melakukan *review* jurnal terkait implementasi metode-metode yang dapat diterapkan dalam menangani berbagai tantangan dalam *data mining*. Topik-topik tersebut meliputi:

- *Anomaly / Outlier Detection*
- *Missing Value*
- *Redundancy / Inconsistency / Feature Selection / Dimensionality Reduction*
- *Imbalanced Dataset* (Over/Under/Hybrid Sampling)

Dengan dasar pemahaman tersebut, setiap kelompok diwajibkan untuk mengimplementasikan hasil temuan dari **JURNAL UTAMA** ke dalam studi kasus nyata. Pelaksanaan tugas besar ini dilakukan secara **berkelompok** sesuai dengan pembagian yang tercantum pada dokumen berikut:

[\*\*\[Klik Disini\] Daftar Kelompok Tugas Besar Penambangan Data\*\*](#)

## 2. Ruang Lingkup Pengerjaan

Tujuan utama tugas besar ini adalah melakukan implementasi metode *preprocessing* pada karakteristik data yang berbeda. Mahasiswa diwajibkan untuk:

1. **Pemilihan Metode:** Memilih satu atau lebih metode penyelesaian masalah data berdasarkan hasil review **JURNAL UTAMA** yang telah dilakukan sebelumnya.
2. **Penggunaan Dataset:** Melakukan implementasi metode tersebut terhadap **2 (dua) jenis dataset** yang berbeda (Dataset 1 dan Dataset 2) yang telah disediakan di E-Learning.
3. **Tahapan Pengerjaan:**
  - a. **Fokus Utama - Preprocessing Data:** Menerapkan metode penanganan masalah data sesuai jurnal rujukan (misal: teknik imputasi spesifik atau deteksi *outlier*). Ini adalah inti tugas besar untuk menangani permasalahan kualitas data.

- b. **Validasi dengan Machine Learning Sederhana:** Menerapkan algoritma standar (misal: Naive Bayes atau Decision Tree) semata-mata sebagai alat ukur. Tujuannya bukan membangun model terbaik, melainkan untuk menguji apakah data hasil *preprocessing* lebih berkualitas dibanding sebelumnya.
- c. **Evaluasi Efektivitas Metode:** Membandingkan hasil akurasi antara metode usulan (dari jurnal) dengan metode konvensional (standar). Analisis difokuskan pada apakah metode *preprocessing* tersebut terbukti lebih efektif dalam meningkatkan kualitas data.

#### 4. Analisis Komparatif (Critical Thinking):

- Membandingkan hasil penerapan metode antara Dataset 1 dan Dataset 2.
- Menentukan pada dataset mana metode tersebut bekerja lebih efektif.
- Memberikan analisis mendalam mengenai alasan perbedaan hasil tersebut berdasarkan karakteristik data.

### 3. Output yang Dikumpulkan

Sebagai syarat kelulusan tugas besar, setiap kelompok wajib mengumpulkan:

#### 1. Kode Program (Repository Github)

- Kode program wajib dimasukkan ke dalam repository (Github/GitLab, dll).
- Pastikan pengaturan repository **tidak private** agar dapat diakses oleh tim penilai.
- Buat repository github semenarik mungkin.

#### 2. Laporan

- Format detail laporan akan diinformasikan kemudian (menyusul).
- Mahasiswa wajib memantau pembaruan informasi terkait format laporan di E-Learning secara berkala.

#### 3. Video Presentasi

- **Durasi:** Maksimal 3 menit.

- **Ketentuan Konten:**

- Bukan video tutorial langkah coding.
- Fokus pada alur proses, hasil, dan *insight*.
- Kreatif, informatif, dan mudah dipahami.
- [\[Klik Disini\] Contoh video yang baik](#)

## 4. Ketentuan Tambahan

- Pastikan Jurnal Utama yang digunakan memiliki reputasi yang baik dan terpercaya seperti Scopus dll.s
- Metode yang dipilih **wajib** diterapkan pada kedua dataset untuk melihat konsistensi performa.
- **Batas Waktu Pengumpulan:** Tugas dikumpulkan paling lambat tanggal **5 Desember 2025** melalui E-Learning.