

UJIAN AKHIR SEMESTER

ANALISIS BIG DATA

Hari/ Tanggal: Selasa, 10 Juni 2025

Dosen: Monica Cinthya, S.T., M.Kom.

Ujian Akhir Semester (UAS) mata kuliah *Big Data Analytic dan Data Mining* dilaksanakan dalam bentuk **kelompok (1 Kelompok Maksimal 4 Anggota)**. Mahasiswa diminta untuk mengembangkan sebuah pipeline analitik data yang mencakup proses **pengambilan data secara langsung (web scraping)**, **Data Preprocessing**, **Exploration Data Analysis (EDA)**, dan **pengembangan model klasifikasi berbasis machine learning**.

Proyek ini menekankan pemahaman dan penerapan konsep-konsep big data dan data mining secara terpadu.

Tahapan Proyek:

1. Data Collection

- Mahasiswa diminta melakukan scraping data dari sumber terbuka seperti berita online, forum, e-commerce, media sosial, atau situs review (misalnya: Twitter, Kompas, IMDb, Tokopedia).
- Data yang dikumpulkan harus memiliki unsur **teks** dan dilabeli berdasarkan **keyword utama** (misalnya: “politik”, “olahraga”, “hiburan”)
- Keyword tersebut akan digunakan sebagai **label klasifikasi**.
- Minimal 10 ribu data.

2. Data Preprocessing

- Pembersihan data (cleaning), normalisasi teks, penghapusan stopwords, tokenisasi, dan konversi ke bentuk numerik (TF-IDF, word embedding, dsb).
- Penanganan missing value **jika ada**.

3. Exploratory Data Analysis (EDA)

- Visualisasi distribusi data berdasarkan label, panjang teks, frekuensi kata kunci, dan korelasi fitur.
- Gunakan tools visualisasi seperti seaborn, matplotlib, atau plotly.

4. Pengembangan Model Machine Learning (Akan dijelaskan oleh dosen Data Mining)
5. Pelaporan dan Presentasi

Learning Outcome:

- Mahasiswa dapat menjelaskan dan memahami dataset hasil scraping dengan keyword sebagai label klasifikasi.
- Mahasiswa dapat menjelaskan Notebook EDA yang menyajikan insight dari data.
- Mahasiswa dapat menjelaskan model klasifikasi beserta evaluasinya.
- Mahasiswa dapat menjelaskan Laporan proyek yang menyeluruh dan sistematis.

Kriteria Penilaian Mata Kuliah Big Data Analysis:

1. Kualitas topik (10%)
2. Pemahaman Dataset (30%)
3. Pemahaman materi dan dapat menyajikan insight dari data (30%)
4. Laporan (25%)
5. PPT (5%)

***Presentasi dilakukan di hari Selasa, 10 Juni 2025. Wajib semua anggota hadir dan membawa laporan.**