

Mata Kuliah : Penggalan Data B
Kredit : 3 SKS
Dosen : Rizky Amelia, S.Si., M.Han

Hari/tanggal : Senin-Minggu, 21-27/10/2024
Waktu : 1 Week
Status : *Team Work*

Instruksi Kerja Ujian Tengah Semester (UTS)

1. Penentuan Metode Penggalan Data (.pdf)

Dalam mempelajari MK Penggalan Data (Data Mining), apa yang ingin kelompok anda capai dengan data mining? Isu apa yang ingin coba kelompok anda selesaikan dengan data mining, apakah itu berkaitan dengan klasifikasi, prediksi, deteksi anomali, atau lainnya? Jika sudah menentukan, selanjutnya;

- a) **Identifikasi metode:** berdasarkan tujuan yang sudah kelompok anda tentukan, pilihlah salah satu metode/teknik penggalan data yang tepat untuk digunakan (*Classification, Cluster Analysis, Regression, Association Rule, Anomaly Detection, Text Mining, Time Series Analysis, Dimensionality Reduction, atau Neural Networks (Deep Learning)*)!
- b) **Justifikasi pilihan:** tuliskan alasan mengapa kelompok anda memilih metode tersebut. Sertakan kelebihan dan kekurangan metode ini dibandingkan metode lain!

2. Review Jurnal (.pdf)

- a) **Pencarian Jurnal:** Cari dan pilih 3-5 jurnal yang relevan dengan metode/teknik penggalan data yang dipilih. Pastikan jurnal tersebut berkualitas dan terkini dalam 5 tahun belakang baik jurnal/prosiding internasional dan/atau jurnal nasional dengan level Sinta antara S1, S2, S3 dan S4.
- b) **Review/ringkasan Jurnal:** Buat ringkasan dari masing-masing jurnal kedalam tabel, mencakup:
 - Masalah yang diangkat serta tujuan dari penelitian
 - Dataset yang digunakan, atribut yang digunakan
 - Metode/teknik penggalan data yang digunakan
 - Algoritma/flowchart/alur penyelesaian seperti apa
 - Hasil yang diperoleh pada penelitian
 - Identifikasi kelemahan dan keterbatasan (seperti metode yang digunakan, ukuran sample yang diteliti, ruang lingkup penelitian, variable/data yang tidak dipertimbangkan, dsb)
 - Sertakan link masing-masing

3. Penentuan Gap dan Ide Topik Proyek Tugas Besar Kelompok (.pdf)

a) Menentukan Gap Penelitian

Berdasarkan analisis dari kelemahan dan keterbatasan masing-masing jurnal yang telah direview, tuliskan satu atau lebih gap penelitian yang dapat diidentifikasi. Gap ini bisa berupa area/data yang kurang diteliti, metode atau pendekatan yang belum diuji, konteks yang belum diperhatikan, dsb.

b) Ide/Topik Penelitian

Berdasarkan gap yang telah diidentifikasi, usulkan ide penelitian yang dapat menjadi fokus topik dari Proyek Tugas Besar kelompok anda. Ide ini boleh menggunakan isu/permasalahan/konteks lain diluar

topik jurnal yang direview (data) tetapi menggunakan metode/pendekatan yang berkaitan dengan jurnal yang direview atau menggunakan dataset yang sama namun metode/pendekatan yang berbeda, ataupun tidak keduanya, ide baru dari kelompok anda dipersilahkan namun tetap menggunakan jurnal yang telah direview sebagai referensi utama. Kemudian buatlah mind map/ kerangka pemikiran yang mencakup:

- Isu yang ingin di angkat
- Judul/topik penelitian yang diusulkan
- Tujuan dan rumusan dari penelitian
- Metode yang diusulkan
- Luaran/output yang diharapkan

4. Pencarian Dataset (Google Colab)

a) Pencarian Dataset

Setelah menentukan teknik/metode lalu kalian dapati gap serta ide/topik penelitian proyek kalian, carilah dataset yang relevan dan memenuhi syarat untuk metode yang digunakan dan permasalahan yang diangkat. Pastikan dataset tersebut berkualitas dan memiliki fitur yang diperlukan. Kalian bisa mencari di repositori data seperti Kaggle, UCI Machine Learning Repository, Google Scholar, Badan Pusat Statistik, Satu Data Indonesia atau sumber lain. Lalu buatlah deskripsi dataset, meliputi :

- Sumber Data (cantumkan link data atau sumber data)
- Format, jumlah data dan fitur
- Atribut pada dataset: tampilkan 10 kolom data beserta atributnya
- Analisis atribut (deskripsi, type data dan jumlah data)

b) Pra-Processing Dataset:

Setelah mendapatkan dataset, lakukan langkah-langkah pra-processing yang sesuai untuk mempersiapkan data sebelum penerapan teknik penggalian data.

- **Pembersihan Data:** lakukan pembersihan data, termasuk menghapus nilai yang hilang, menghapus duplikasi, dan memperbaiki kesalahan data.
- **Transformasi Data:** lakukan transformasi yang diperlukan, seperti normalisasi, standarisasi, encoding variabel kategorikal (konversi variabel kategorikal menjadi numerik jika diperlukan), dan pengurangan dimensi jika diperlukan.
- **Pembagian Data:** seperti memisahkan dataset menjadi set pelatihan, validasi, dan pengujian untuk evaluasi model yang lebih baik.
- **Penanganan Data Imbalance:** mengatasi ketidakseimbangan antara kelas dalam dataset, misalnya dengan menambah jumlah contoh dari kelas minoritas atau mengurangi kelas mayoritas.
- Dan proses pra-processing dataset lainnya sesuai kebutuhan dari penelitian kelompok anda.
- Buatlah visualisasi dari dataset dalam beberapa bentuk diagram seperti histogram, blox pot, dsb.

Buatlah dokumentasi proses dengan menambahkan catatan tentang setiap langkah pra-processing yang dilakukan. Jelaskan alasan di balik setiap langkah.