



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN  
RISET, DAN TEKNOLOGI

UNIVERSITAS PADJADJARAN

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PROGRAM STUDI S-1 TEKNIK INFORMATIKA

Jl. Raya Bandung-Sumedang Km. 21 Jatinangor 45363 Telp./Fax. 022 7798983  
<http://informatika.unpad.ac.id>, e-mail : [informatika@unpad.ac.id](mailto:informatika@unpad.ac.id)

---

UJIAN TENGAH SEMESTER GANJIL 2023/2024

Mata Kuliah : Data Mining  
Dosen : Dr. Intan Nurma Yulita, M.T  
Kelas : A & B  
Hari/Tanggal : Kamis, 19 Oktober 2023  
Sifat : *Take Home*

**Soal**

Buatlah penerapan klasterisasi teks berbahasa Indonesia berdasarkan data *tweet* untuk suatu produk baik barang maupun jasa. Tugas dikerjakan secara berkelompok yang terdiri dari 4 orang. Tugas disajikan dalam bentuk makalah dan juga aplikasi sederhana berbasis web. Format makalah harus mengikuti kaidah penulisan dalam konferensi IEEE. Adapun poin-poin yang perlu diperhatikan, antara lain:

- Abstrak  
Bagian ini harus merangkum isi Bab 1 sampai dengan Bab 5 yaitu deskripsi singkat mengenai permasalahan, metode, metodologi hasil, dan kesimpulan.
- Bab 1. Pendahuluan  
Bagian ini menjelaskan latar belakang permasalahan, pembahasan penyelesaian masalah melalui praproses teks dan klasterisasi serta alasan penggunaan algoritma yang digunakan untuk menyelesaikan permasalahan.
- Bab 2. Metode  
Bagian ini menjelaskan teori mengenai seluruh algoritma yang digunakan secara detail, serta parameter evaluasi yang digunakan.
- Bab 3. Metodologi  
Jelaskan tahap-tahap dalam menyelesaikan permasalahan secara teknis dan detail. Lebih baik jika tahapan juga digambarkan dalam bentuk *flow chart*. Tidak perlu mencantumkan terlalu teknis mekanisme penggunaan tools/aplikasi yang digunakan. Jangan lupa menjelaskan deskripsi dataset yang digunakan.
- Bab 4. Hasil dan Analisis  
Lakukan sejumlah eksperimen untuk menemukan parameter yang optimal untuk algoritma yang digunakan. Serta lakukan analisis terhadap hasil yang diperoleh.
- Bab 5. Kesimpulan  
Berikan kesimpulan mengenai eksperimen yang dilakukan dalam menyelesaikan permasalahan.
- Referensi  
Gunakan minimal 10 referensi dari konferensi/jurnal.

**Catatan:**

- Cantumkan sitasi referensi, tabel, gambar sesuai format IEEE.
- Makalah berfokus pada penyelesaian masalah menggunakan praproses teks dan klasterisasi sehingga bukanlah panduan teknis menggunakan tools.
- Minimal ada lima halaman dalam makalah
- Maksimal *similarity* dalam makalah adalah 30%

Makalah dikumpulkan di [reguler.live.unpad.ac.id](http://reguler.live.unpad.ac.id) dengan format \*.doc/\*.docx paling lambat pada hari Minggu, 29 Oktober 2023 pukul 23.59. Dataset, file \*.txt yang berisi nama dan NPM anggota, *code* aplikasi, serta *screenshot* bukti aplikasi berjalan juga perlu diunggah. Adapun *code* harus menyertakan komentar pada barisan *code* yang disajikan. Cukup satu orang perwakilan kelompok yang mengumpulkan tugas.