Tugas Klasifikasi Teks

Deskripsi:

Buatlah sebuah aplikasi klasifikasi teks sederhana yang akan mengklasifikasikan sekumpulan teks Bahasa Indonesia ke dalam beberapa kelas. Tiap kelompok silakan mengadaptasi beberapa aplikasi klasifikasi teks untuk menentukan tema tugas, misal:

- Klasifikasi sentimen. Contoh: klasifikasi review kendaraan di forum otomotif ke dalam kelas positif dan negatif

- Klasifikasi gender. Contoh: klasifikasi gender penulis komentar di forum online.

- Deteksi stance. Contoh: klasifikasi teks di kolom komentar berita online, apakah masuk ke kelas: pendukung capres-cawapres nomor 1, atau capres-cawapres nomor 2.

- Deteksi hate-speech. Contoh: klasifikasi teks di kolom komentar berita online, apakah termasuk dalam hate speech

- Klasifikasi berita koran online menurut topiknya.

Masing-masing kelompok membangun sendiri dataset-nya, baik untuk data pelatihan maupun data pengujian (membangun dalam hal ini adalah melakukan prapemrosesan dan melakukan pelabelan, tidak memanfaatkan data yang sudah siap untuk dipakai). Antar kelompok boleh memilih tema tugas yang sama, tetapi harus menggunakan dataset yang berbeda. Jumlah dokumen dalam corpus minimal 100 dokumen.

Deliverables:

1. Softcopy program beserta komentar yang mudah dimengerti

2. Softcopy corpus

3. Petunjuk cara eksekusi program

4. Laporan singkat yang berisi:

a. Keterangan sumber corpus, tema yang dipilih.

b. Teknik yang dipakai, fitur yang dipilih, teknik pengujian dan evaluasi kinerja

c. Analisis terhadap hasil pengujian dan evaluasi kinerja

Komponen penilaian:

1. Kebenaran, kelengkapan dan kejelasan program (nilai maksimum: 70)

2. Laporan (nilai maksimum: 30)

3. Bonus: penerapan metode tambahan untuk meningkatkan kinerja misalnya smoothing, multinomial classification (nilai maksimum: 20)

Dikumpulkan 15 Mei 2019