



UJIAN TENGAH SEMESTER GENAP TA 2025/2026

Mata Kuliah : Kecerdasan Buatan
Prodi/Semester : Teknik Informatika
Hari/Tanggal : Senin/ 06 November 2025
Nama Dosen : Muhamad Fatchan, S.Kom., M.kom
Kelas : **TI.22.A.SE.1**

Laporan Kelompok

Nama : Novianto Andi Hardiansyah (312210316)
Zalfa Dewi Zahrani (312210320)
Dafa Alfiana Erlangga (312210467)

1. Latar Belakang

Konten bermuatan kebencian dan kata-kata kasar pada media sosial berdampak negatif bagi pengguna. Diperlukan sistem moderasi otomatis berbasis Kecerdasan Buatan (AI) yang mampu mengklasifikasikan teks menjadi “ujaran kebencian/abusive” atau “bukan”. Pada tugas ini, saya membangun model klasifikasi biner dengan fitur TF-IDF dan membandingkan beberapa algoritma ringan (LinearSVC & MultinomialNB), kemudian mengevaluasi kinerjanya dengan metrik standar klasifikasi.

2. DATA YANG DIGUNAKAN

- Nama file: Dataset Ujaran.csv
- Ukuran: 24.783 baris, 2 kolom: tweet (teks) dan class (label).
- Skema label asli (class): 0 = hate, 1 = abusive, 2 = normal.
- Pemetaan label biner untuk tugas ini:
- $\text{label_binary} = 1$ jika $\text{class} \in \{0, 1\}$ (ujaran kebencian/abusive), dan 0 jika $\text{class} = 2$ (non-hate).
- Distribusi (hasil baca data): class: 0=1430, 1=19190, 2=4163 \rightarrow biner: 1=20620, 0=4163 (terdapat imbalance).

3. METODOLOGI

3.1 Pra-Pemrosesan (Preprocessing)

- Lowercase, hapus URL, angka, dan tanda baca.
- Stopword ringan (Inggris) untuk mengurangi kata umum.
- (Opsional) normalisasi slang/emoji tidak dilakukan pada baseline ini.

3.2 Ekstraksi Fitur

- **TF-IDF** dengan parameter: n-gram (1–2), min_df=2, max_features=40.000.
- Alasan: representasi sederhana-kuat untuk teks pendek (tweet)

3.3 Pembagian Data

- Train/Test = **80/20**, **stratified** terhadap label biner.

3.4 Model Yang dibandingkan

- LinearSVC (SVM linear): kuat untuk data teks high-dimensional.
- Multinomial Naive Bayes: baseline cepat untuk teks.

3.5 Metrik Evaluasi

- Accuracy, Precision, Recall, F1 (kelas positif = 1) dan Confusion Matrix.
- Interpretasi Singkat: Precision tinggi \rightarrow sedikit *false positive* (tidak mudah salah nuduh). Recall tinggi \rightarrow sedikit *false negative* (hate tidak banyak yang lolos). F1 menyeimbangkan Precision–Recall.

4. HASIL DAN ANALISIS

4.1 Tabel Ringkasan performa

Classification Report:				
	precision	recall	f1-score	support
0	0.88	0.76	0.82	833
1	0.95	0.98	0.97	4124
accuracy			0.94	4957
macro avg	0.92	0.87	0.89	4957
weighted avg	0.94	0.94	0.94	4957



4.2 Analisa Singkat

- Dataset imbalanced
- LinearSVC cenderung unggul pada teks pendek TF-IDF.
- Potensi error: sarkasme, slang, atau konteks yang hilang.
- Ide perbaikan: `class_weight='balanced'`, tuning hyperparameter, normalisasi slang, dan model transformer (BERT/IndoBERT) jika ingin akurasi lebih tinggi.

5. Kesimpulan

Baseline klasifikasi ujaran kebencian pada tweet berhasil dibangun dengan TF-IDF dan dievaluasi menggunakan Accuracy, Precision, Recall, F1 serta Confusion Matrix. menunjukkan model cukup efektif sebagai baseline moderasi otomatis. Pekerjaan lanjutan disarankan pada penanganan imbalance, perluasan preprocessing, dan eksplorasi model berbasis transformer.