



Analisis Sentimen Publik terhadap Kenaikan UKT pada Platform X Menggunakan Support Vector Machine

Ahmad Agung Hadikosyah^{1*}, Wilda Rina Hasibuan²

^{1*,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
Email: ^{1*}agunghk154@gmail.com, ²wildarina@umsu.ac.id

Abstrak

Kenaikan Uang Kuliah Tunggal (UKT) pada 2024 memicu perdebatan luas di media sosial X yang menghasilkan data teks tidak terstruktur. Penelitian ini memetakan sentimen publik terhadap isu tersebut, membandingkan tiga kernel Support Vector Machine (SVM), yaitu linear, polynomial, dan Radial Basis Function (RBF), dan mengimplementasikan model terbaik ke prototipe web. Data berupa 1.500 postingan berbahasa Indonesia periode Mei–Juli 2024 yang dipraproses, dilabeli otomatis dengan InSet Lexicon, dan dibobotkan dengan TF-IDF. Evaluasi memakai hold-out 80:20 dan 10-fold stratified cross validation dengan F1-macro, uji t berpasangan, dan uji Wilcoxon, serta Multinomial Naive Bayes sebagai pembanding. Sentimen negatif mendominasi (84,27%), diikuti positif (11,93%) dan netral (3,80%). Tema keberatan terbanyak adalah tuntutan transparansi penetapan kelompok UKT dan beban finansial keluarga. Pada validasi silang, F1-macro kernel RBF, linear, dan polynomial adalah 0,6571, 0,6520, dan 0,6479, tetapi selisihnya tidak signifikan ($p > 0,05$) sehingga ketiganya setara pada data ini. Kelas netral paling sulit diprediksi ($F1 \leq 0,38$). Prototipe lolos 12 skenario black-box dengan waktu respons rata-rata 10,5 milidetik, namun hanya 4 dari 10 teks baru yang sesuai penilaian manusia karena semua diprediksi negatif. Label manual dan data yang lebih beragam diperlukan sebelum model digunakan di luar konteks datasetnya.

Kata Kunci: Analisis Sentimen, Uang Kuliah Tunggal, Platform X, Support Vector Machine, TF-IDF.

PENDAHULUAN

Uang Kuliah Tunggal (UKT) adalah sistem pembiayaan yang menetapkan satu tarif per semester bagi mahasiswa perguruan tinggi negeri (PTN) sesuai kemampuan ekonomi keluarga. Pada 2024, Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 2 Tahun 2024 tentang Standar Satuan Biaya Operasional Pendidikan Tinggi (SSBOPT) diterbitkan sebagai dasar peningkatan anggaran bagi PTN, dan penyesuaian UKT di sejumlah PTN menuai masukan dan kritik dari masyarakat. Menindaklanjutinya, pada 27 Mei 2024 Mendikbudristek mengumumkan pembatalan kenaikan UKT tahun ajaran 2024/2025 (Kemendikbudristek, 2024).

Perdebatan tersebut berlangsung luas di media sosial X. Pada korpus penelitian ini, tagar #TolakKenaikanUKT, #UKTMahal, dan #TolakUKT masing-masing muncul pada 406, 406, dan 338 postingan. Teks media sosial yang pendek, informal, dan tidak terstruktur sulit dibaca secara manual dalam jumlah besar. Karena itu analisis sentimen otomatis, yang mengklasifikasikan opini ke kelas positif, negatif, atau netral, lazim dipakai untuk memetakan opini publik (Koto dan Rahmaningtyas, 2017).

Support Vector Machine (SVM) banyak dipilih untuk klasifikasi teks karena ruang fitur teks berdimensi tinggi dan jarang umumnya dapat dipisahkan secara linear (Joachims, 1998), dengan dasar teoretis berupa pemisahan margin maksimum (Cortes dan Vapnik, 1995). Fungsi kernel memungkinkan pemisahan non-linear, tetapi kernel yang paling sesuai bergantung pada karakteristik data.

Pada analisis sentimen berbahasa Indonesia, SVM sering dibandingkan dengan Naive Bayes. Kurniawan dkk. (2023) melaporkan SVM lebih akurat daripada Naive Bayes pada sentimen marketplace di Twitter, dan Ningsih dkk. (2024) melaporkan akurasi 70,82% untuk SVM dan 63,02% untuk Naive Bayes pada sentimen penggunaan mobil listrik. Keduanya menilai keunggulan berdasarkan akurasi. Pada isu UKT, Gulo dan Purwanto (2024) menerapkan Naive Bayes pada 250 tweet periode 3–26 Mei 2024 dengan akurasi 80% tetapi recall kelas positif hanya 40%. Raiya dkk. (2025) menerapkan LSTM pada komentar YouTube dan Instagram dengan akurasi 84% tetapi F1 kelas positif 0,41. Sumartha (2025) memakai IndoBERT pada data X dan melaporkan mayoritas opini bersifat negatif. Arah penelitian juga bergerak ke model berbasis transformer seperti IndoBERTweet (Koto dkk., 2021). Dua studi UKT yang

melaporkan metrik per kelas menunjukkan pola yang sama: akurasi tinggi, tetapi kinerja pada kelas minoritas rendah. Pola ini menandakan bahwa akurasi saja tidak memadai pada data yang timpang (He dan Garcia, 2009; Sokolova dan Lapalme, 2009).

Dari studi yang berhasil ditelusuri penulis, belum ada kajian isu kenaikan UKT 2024 pada platform X yang (1) membandingkan tiga kernel SVM secara berdampingan dengan F1-macro, (2) menguji perbedaannya secara statistik, dan (3) melanjutkannya ke prototipe aplikasi yang diuji pada teks di luar dataset. Kebaruan artikel ini terletak pada kombinasi tersebut, bukan pada klaim bahwa satu kernel lebih unggul.

Permasalahan penelitian ini ialah bagaimana distribusi sentimen publik terhadap kenaikan UKT, bagaimana kinerja ketiga kernel SVM dan apakah perbedaannya bermakna, apa karakteristik tema opini pada hasil klasifikasi, dan bagaimana model terbaik diimplementasikan serta diuji. Tiga hipotesis diajukan. H1: kernel RBF menghasilkan F1-macro lebih tinggi daripada kernel linear dan polynomial. H2: SVM dengan kernel terbaik menghasilkan F1-macro lebih tinggi daripada Multinomial Naive Bayes. H3: sentimen negatif mendominasi distribusi kelas pada Mei–Juli 2024. Tujuan penelitian adalah memetakan distribusi sentimen, mengukur dan membandingkan kinerja kernel SVM, menganalisis tema opini, serta mengimplementasikan dan menguji prototipe aplikasi web.

METODE

Penelitian ini bersifat kuantitatif dengan desain eksperimen komputasional yang deskriptif-komparatif. Semua model dibandingkan pada dataset, praproses, dan metrik yang sama. Nilai acak ditetapkan tetap (`random_state = 42`) agar hasil dapat direproduksi (Pedregosa dkk., 2011). Alur penelitian disajikan pada Gambar 1.



Gambar 1. Alur tahapan penelitian

Pengumpulan data. Postingan publik dikumpulkan dengan web crawling [SEBUTKAN ALAT, VERSI, DAN TANGGAL PENGAMBILAN DATA] pada 1 Mei–31 Juli 2024 dengan filter bahasa Indonesia dan sembilan kata kunci: "UKT", "kenaikan UKT", "UKT naik", "UKT mahal", "biaya kuliah naik", "#TolakUKT", "#TolakKenaikanUKT", "#UKTMahal", dan "UKT 2024". Atribut yang disimpan hanya id teranonimkan, teks, tanggal, jumlah suka, dan jumlah repost. Nama akun, foto profil, dan lokasi tidak disimpan. Penyaringan menghapus teks duplikat, postingan berbahasa asing, serta postingan berisi tautan saja atau spam. Sampel akhir diambil acak proporsional per bulan hingga 1.500 postingan. Dataset teranonimkan dan kode tersedia di [TAUTAN REPOSITORY].

Praproses teks. Enam langkah dijalankan berurutan: pembersihan (mention, tagar, URL, emoji, dan tanda baca), case folding, normalisasi kata tidak baku dengan kamus, tokenisasi, penghapusan stopwords, dan stemming dengan pustaka Sastrawi (lihat Tala, 2003). Stopword memakai daftar bahasa Indonesia NLTK (757 kata) dengan pengecualian kata negasi "tidak", "bukan", "belum", dan "jangan".

Pelabelan. Label dibuat otomatis dengan InSet Lexicon (Koto dan Rahmaningtyas, 2017), yang memuat 3.609 kata positif dan 6.609 kata negatif berbobot -5 sampai $+5$. Skor dokumen adalah jumlah bobot token yang ditemukan dalam leksikon. Skor > 0 berlabel positif, skor < 0 negatif, dan skor $= 0$ netral. Tidak ada validasi oleh pelabel manusia.

Ekstraksi fitur. Teks diubah menjadi vektor TF-IDF (Salton dan Buckley, 1988) dengan unigram dan bigram, `max_features = 5.000`, `min_df = 2`, `max_df = 0,90`, sublinear tf, dan normalisasi L2. Vectorizer dilatih hanya pada data latih di dalam Pipeline Scikit-learn untuk mencegah kebocoran data.

$$\text{TF-IDF}(t,d) = (1 + \ln f(t,d)) \times \text{idf}(t) \quad (1)$$

Klasifikasi. Klasifikasi tiga kelas memakai kelas SVC Scikit-learn [VERIFIKASI: strategi multikelas yang benar-benar dipakai di kode] dengan tiga kernel:

$$K_{\text{linear}}(x, x') = x \cdot x' \quad (2)$$

$$K_{\text{poly}}(x, x') = (\gamma x \cdot x' + r)^d \quad (3)$$

$$K_{\text{RBF}}(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4)$$

Karena dua kelas berproporsi di bawah 15%, `class_weight = "balanced"` diterapkan pada seluruh model SVM. Multinomial Naive Bayes dilatih tanpa pembobotan kelas sebagai pembanding. Hiperparameter dioptimalkan dengan GridSearchCV 5-fold pada data latih dengan skor F1-macro berikut.

Tabel 1. Ruang parameter grid search

Model	Parameter yang diuji
SVM linear	$C \in \{0,01; 0,1; 1; 10; 100\}$
SVM polynomial	$C \in \{0,1; 1; 10\}; \gamma \in \{\text{scale, auto}\}; \text{derajat} \in \{2, 3\}; \text{coef0} \in \{0, 1\}$
SVM RBF	$C \in \{0,1; 1; 10; 100\}; \gamma \in \{\text{scale, auto, 0,001; 0,01; 0,1}\}$
Multinomial NB	$\alpha \in \{0,01; 0,1; 0,5; 1,0; 2,0\}$

Evaluasi. Dua skema dipakai: hold-out stratified 80:20 (1.200 data latih dan 300 data uji) dan 10-fold stratified cross validation pada seluruh dataset sebagai skema utama. Metrik dihitung dari confusion matrix tiga kelas. Metrik keputusan adalah F1-macro karena kelas tidak seimbang.

$$\text{Akurasi} = (\sum \text{TPc}) / N \quad (5)$$

$$\text{Precisionc} = \text{TPc} / (\text{TPc} + \text{FPc}) \quad (6)$$

$$\text{Recallc} = \text{TPc} / (\text{TPc} + \text{FNc}) \quad (7)$$

$$\text{F1c} = (2 \times \text{Precisionc} \times \text{Recallc}) / (\text{Precisionc} + \text{Recallc}) \quad (8)$$

$$\text{F1-macro} = (\text{F1positif} + \text{F1negatif} + \text{F1netral}) / 3 \quad (9)$$

Perbedaan F1-macro antarmodel pada sepuluh lipatan yang sama diuji dengan uji t berpasangan dan uji Wilcoxon ($\alpha = 0,05$), disertai ukuran efek d untuk sampel berpasangan ($d = t/\sqrt{n}$). Karena lipatan saling tumpang tindih, uji ini bersifat indikatif (Dietterich, 1998), dan Wilcoxon dipakai sebagai alternatif nonparametrik (Demšar, 2006).

Prototipe aplikasi web. Prototipe dikembangkan sebagai sarana demonstrasi dan pengujian fungsional, sebagaimana lazim pada pengembangan sistem berbasis prototipe (Maharani dkk., 2025). Arsitekturnya klien–server tiga lapis: antarmuka React.js (dengan Vite, Tailwind CSS, dan Axios), REST API Flask, dan model terlatih yang diserialisasi (.pkl). Praproses dan TF-IDF di server identik dengan tahap pelatihan. Endpoint `/api/predict` mengembalikan label dan skor keyakinan berupa normalisasi softmax atas nilai decision function, yang bukan probabilitas terkalibrasi. Pengujian meliputi 12 skenario black-box, pengukuran waktu respons pada 30 permintaan berturut-turut, dan uji coba pada 10 kalimat bertema UKT buatan peneliti di luar dataset (4 negatif, 3 positif, 3 netral menurut penilaian manusia).

HASIL DAN PEMBAHASAN

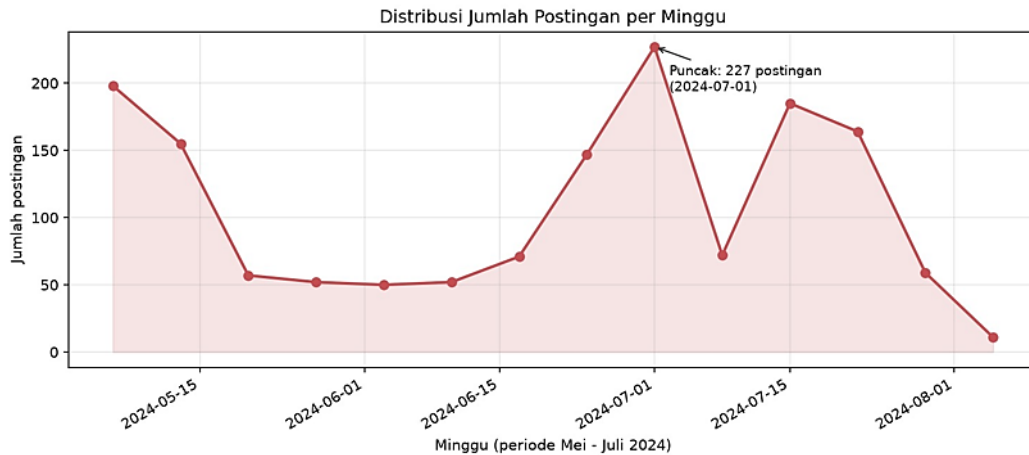
Karakteristik data dan praproses

Crawling menghasilkan 2.085 postingan mentah. Setelah penyaringan tersisa 2.000 postingan, dan pengambilan sampel proporsional menghasilkan 1.500 postingan (490 Mei, 504 Juni, 506 Juli 2024) (Tabel 2). Kata kunci "ukt" paling produktif (1.799 kemunculan), sedangkan frasa utuh "ukt mahal" tidak ditemukan.

Tabel 2. Rekapitulasi penyaringan data

Tahap	Masuk	Dibuang	Sisa
Duplikat teks identik	2.085	77	2.008
Postingan berbahasa asing	2.008	3	2.005
Hanya URL dan spam	2.005	5	2.000
Sampling proporsional per bulan	2.000	500	1.500

Jumlah postingan per minggu memuncak tiga kali: 198 pada minggu pertama Mei, 227 pada minggu yang berakhir 1 Juli, dan 185 pada pertengahan Juli. Jumlah terendah, 50–57 per minggu, terjadi dari pertengahan Mei sampai awal Juni (Gambar 2). Praproses menurunkan total token dari 31.385 menjadi 17.131 (–45,4%), kosakata unik dari 1.570 menjadi 1.007 (–35,9%), dan rata-rata token per dokumen dari 20,92 menjadi 11,42. Daftar stopword bawaan NLTK ternyata memuat kata "naik" dan "buat", sehingga keduanya ikut terhapus.



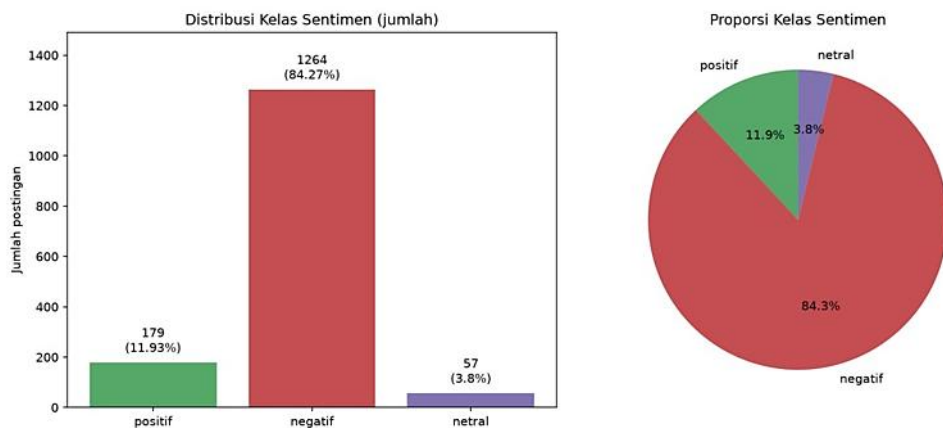
Gambar 2. Distribusi jumlah postingan per minggu

Distribusi sentiment

Pelabelan InSet menghasilkan dominasi kelas negatif (Tabel 3, Gambar 3). Kelas netral dibentuk secara residual dari postingan yang berskor nol.

Tabel 3. Distribusi kelas sentimen hasil pelabelan InSet Lexicon

Kelas	Jumlah	Persentase	Rentang skor	Rerata skor
Positif	179	11,93%	1 s.d. 11	3,17
Negatif	1.264	84,27%	-42 s.d. -1	-10,71
Netral	57	3,80%	0	0,00
Total	1.500	100%	—	—



Gambar 3. Distribusi kelas sentimen pada dataset

Fitur dan penyetelan model

Pembagian hold-out menghasilkan 1.200 data latih (143 positif, 1.011 negatif, 46 netral) dan 300 data uji (36, 253, 11). Kosakata TF-IDF berjumlah 1.644 term (543 unigram dan 1.101 bigram), sehingga batas 5.000 tidak pernah mengikat. Matriks latih berukuran 1.200×1.644 dengan kejarangan 98,84% dan rata-rata 19,03 fitur aktif per dokumen. Parameter terbaik hasil grid search disajikan pada Tabel 4. Selisih F1-macro validasi silang antarkernel pada tahap ini kurang dari 0,006.

Tabel 4. Parameter terbaik dan skor validasi silang 5-fold pada data latih

Model	Parameter terbaik	F1-macro CV-5	Waktu (detik)
SVM linear	$C = 1$	0,6434	4,23
SVM polynomial	$C = 1$; derajat = 2; $\gamma = \text{scale}$; $\text{coef0} = 1$	0,6413	8,11
SVM RBF	$C = 10$; $\gamma = 0,1$	0,6383	6,48
Multinomial NB	$\alpha = 0,01$	0,6114	0,33

Kinerja model

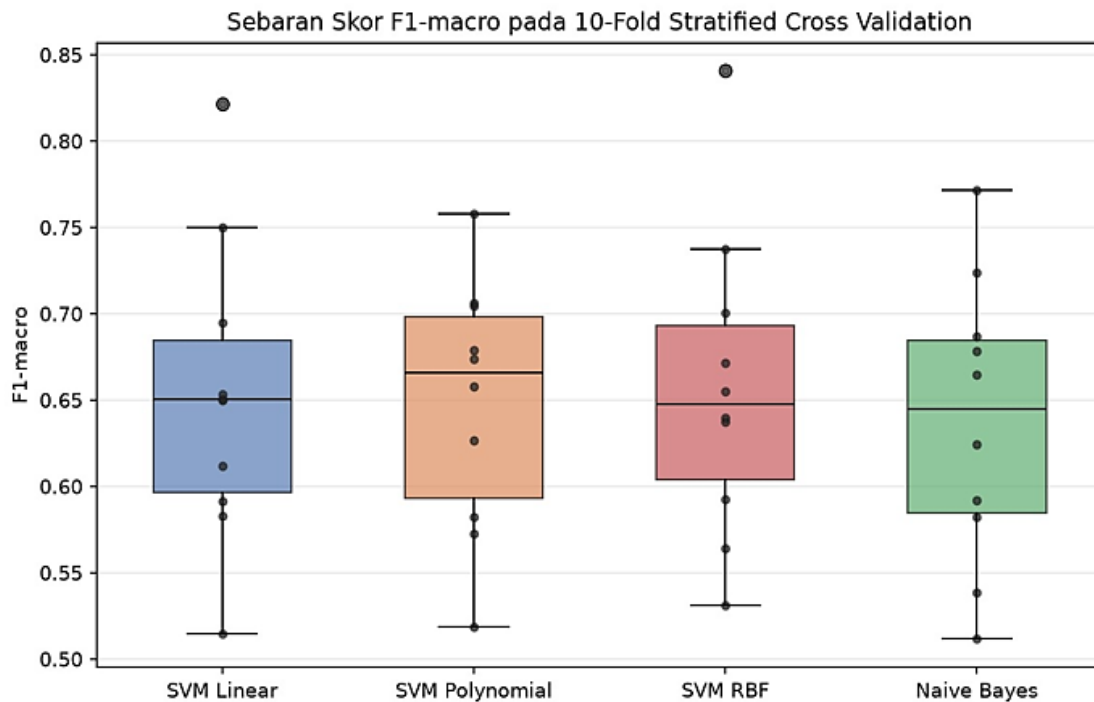
Pada hold-out, akurasi berkisar 0,8700–0,9167, tetapi F1-macro hanya 0,6416–0,6962, dan Naive Bayes tertinggi (Tabel 5). Pada 10-fold cross validation, urutannya berubah: kernel RBF tertinggi (F1-macro 0,6571), disusul linear, polynomial, dan Naive Bayes (Tabel 6). Simpangan baku antarlipatan (0,0689–0,0846) jauh lebih besar daripada selisih antarmodel (Gambar 4).

Tabel 5. Kinerja pada data uji hold-out 80:20

Model	Akurasi	Precision-macro	Recall-macro	F1-macro
SVM linear	0,8700	0,6043	0,7060	0,6416
SVM polynomial	0,9033	0,6572	0,6743	0,6628
SVM RBF	0,8933	0,6441	0,7152	0,6726
Multinomial NB	0,9167	0,7621	0,6716	0,6962

Tabel 6. Rerata $\pm$ simpangan baku pada 10-fold stratified cross validation

Model	Akurasi	Precision-macro	Recall-macro	F1-macro
SVM linear	0,8847 $\pm$ 0,0329	0,6329 $\pm$ 0,0881	0,6975 $\pm$ 0,0854	0,6520 $\pm$ 0,0832
SVM polynomial	0,9133 $\pm$ 0,0186	0,6777 $\pm$ 0,1206	0,6528 $\pm$ 0,0574	0,6479 $\pm$ 0,0689
SVM RBF	0,9007 $\pm$ 0,0261	0,6434 $\pm$ 0,1005	0,6848 $\pm$ 0,0753	0,6571 $\pm$ 0,0846
Multinomial NB	0,9087 $\pm$ 0,0242	0,6765 $\pm$ 0,1233	0,6238 $\pm$ 0,0726	0,6373 $\pm$ 0,0783

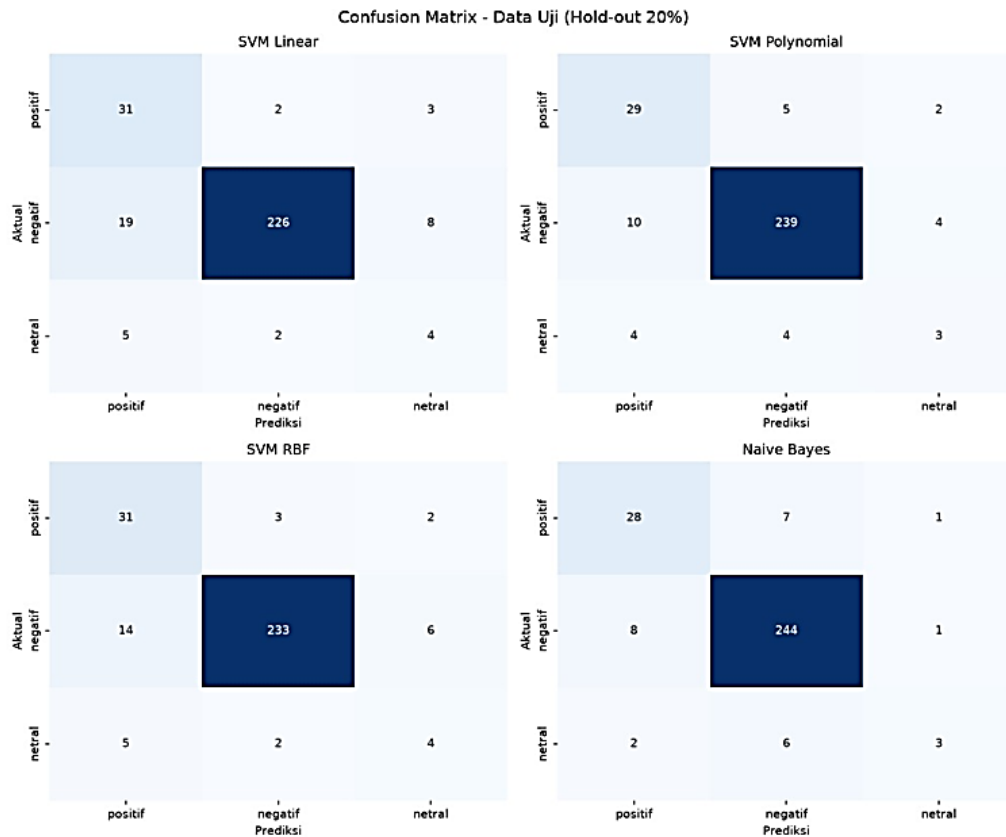


Gambar 4. Sebaran F1-macro pada sepuluh lipatan validasi silang

Per kelas, kelas negatif diprediksi sangat baik, kelas positif sedang, dan kelas netral paling rendah (Tabel 7). Model SVM RBF melakukan 32 kesalahan dari 300 data uji (10,67%): negatif diprediksi positif 14, negatif diprediksi netral 6, dan netral diprediksi positif 5. Hanya 4 dari 11 dokumen netral dikenali benar (Gambar 5).

Tabel 7. F1-score per kelas pada data uji hold-out

Model	Positif (n = 36)	Negatif (n = 253)	Netral (n = 11)
SVM linear	0,6813	0,9358	0,3077
SVM polynomial	0,7342	0,9541	0,3000
SVM RBF	0,7209	0,9491	0,3478
Multinomial NB	0,7568	0,9569	0,3750



Gambar 5. Confusion matrix SVM RBF pada data uji hold-out

Uji signifikansi

Tidak ada pasangan model yang terlihat berbeda signifikan (Tabel 8). Ukuran efek berkisar diantara 0,081–0,257.

Tabel 8. Uji signifikansi perbedaan F1-macro pada sepuluh lipatan

Pasangan model	Selisih F1	t	p (uji t)	p (Wilcoxon)	d
RBF vs linear	+0,0050	0,399	0,6989	0,4316	0,126
RBF vs polynomial	+0,0092	0,681	0,5127	0,7695	0,215
Linear vs polynomial	+0,0042	0,256	0,8034	0,8457	0,081
RBF vs Naive Bayes	+0,0197	0,813	0,4374	0,4922	0,257
Linear vs Naive Bayes	+0,0148	0,767	0,4626	0,3223	0,243
Polynomial vs Naive Bayes	+0,0106	0,512	0,6208	0,6953	0,162

Efisiensi

Waktu pelatihan beserta grid search adalah 4,23 detik (linear), 8,11 (polynomial), 6,48 (RBF), dan 0,33 (Naive Bayes). Waktu inferensi per dokumen adalah 0,1070; 0,1112; 0,1442; dan 0,0224 milidetik. Jumlah support vector adalah 631 (linear), 532 (polynomial), dan 549 (RBF).

Tema opini. Model RBF yang diterapkan pada seluruh dataset memprediksi 1.220 postingan negatif (81,33%), 208 positif (13,87%), dan 72 netral (4,80%). Ini estimasi dalam sampel karena sebagian data adalah data latih. Jumlah postingan yang diprediksi negatif per bulan adalah 388 dari 490 (79,2%) pada Mei, 419 dari 504 (83,1%) pada Juni, dan 413 dari 506 (81,6%) pada Juli. Analisis isi berbasis kata pencari menghasilkan tema pada Tabel 9. Satu postingan dapat memuat lebih dari satu tema. Kelas netral tidak ditafsirkan karena berukuran kecil dan heterogen.

Tabel 9. Tema dominan pada kelas hasil prediksi SVM RBF

Kelas	Tema	Jumlah	Persentase
Negatif (n = 1.220)	Tuntutan transparansi dan keadilan penetapan	330	27,05%
	Beban finansial keluarga	305	25,00%
	Informasi resmi, pengumuman, dan prosedur	253	20,74%
	Ketidaksesuaian dengan mutu layanan	184	15,08%
Positif (n = 208)	Informasi resmi, pengumuman, dan prosedur	84	40,38%
	Tuntutan transparansi dan keadilan penetapan	68	32,69%
	Dukungan, apresiasi, dan penerimaan kebijakan	52	25,00%
	Ketidaksesuaian dengan mutu layanan	46	22,12%

Prototipe aplikasi

Model SVM RBF ($C = 10$, $\gamma = 0,1$) diimplementasikan pada arsitektur pada Gambar 6, dan tampilan antarmukanya pada Gambar 7. Seluruh 12 skenario black-box berhasil, termasuk penanganan masukan kosong (HTTP 400), teks melebihi 1.000 karakter (400), teks tanpa kata bermakna setelah praproses (422), dan endpoint tidak dikenal (404). Dari 30 permintaan berturut-turut, waktu respons rata-rata 10,5 milidetik (median 9,7; minimum 5,8; maksimum 24,3). Pada uji teks baru, hanya 4 dari 10 prediksi sesuai penilaian manusia (Tabel 10). Skor keyakinan berada pada rentang sempit, 68,7–76,1%. Dari 69 token hasil praproses kesepuluh teks, 14 token (20,3%) tidak ada dalam kosakata model, dan tiga teks tanpa token di luar kosakata pun tetap diprediksi negatif.

Gambar 7. Antarmuka prototipe aplikasi

Tabel 10. Hasil uji coba pada sepuluh teks baru di luar dataset

Penilaian manusia	Jumlah teks	Diprediksi negatif	Sesuai
Negatif	4	4	4
Positif	3	3	0
Netral	3	3	0
Total	10	10	4

Pengujian hipotesis

H1 hanya diterima secara arah: F1-macro RBF (0,6571) lebih tinggi daripada linear (0,6520) dan polynomial (0,6479), tetapi tidak signifikan ($p = 0,6989$ dan $0,5127$). H2 tidak didukung: RBF unggul atas Naive Bayes pada validasi silang (0,6571 vs 0,6373; $p = 0,4374$), tetapi arahnya berbalik pada hold-out (0,6726 vs 0,6962). H3 diterima: kelas negatif mencapai 84,27% dan tetap dominan pada ketiga bulan.

Pembahasan

Pertama, dominasi sentimen negatif sejalan dengan studi isu UKT lain. Gulo dan Purwanto (2024) menemukan 189 dari 250 tweet (75,6%) negatif pada 3–26 Mei 2024, dan Sumartha (2025) melaporkan mayoritas opini negatif pada data X. Raiya dkk. (2025) menemukan 83,5% kelas negatif pada data uji komentar YouTube dan Instagram. Label studi tersebut berbeda (penurunan atau kenaikan minat kuliah), sehingga besaran antarstudi tidak dapat dibandingkan langsung. Tema keberatan terbanyak berkaitan dengan prosedur penetapan kelompok UKT (27,05%) di samping beban finansial (25,00%). Ini menunjukkan keberatan publik tidak semata soal kemampuan membayar. Kelas "positif" perlu dibaca hati-hati: 40,38% isinya informasi resmi berdiskusi baik, dan hanya 25,00% berupa dukungan atau penerimaan eksplisit. Temuan ini bersifat indikatif karena berasal dari satu platform, satu periode, dan label otomatis, dan pengguna X bukan sampel representatif masyarakat.

Kedua, akurasi menyesatkan pada data ini. Akurasi 87,00–91,67% pada hold-out hampir sama dengan proporsi kelas negatif (84,27%), sedangkan F1-macro hanya 0,64–0,70. Pola yang sama terlihat pada Gulo dan Purwanto (2024) serta Raiya dkk. (2025), yaitu akurasi tinggi dengan kinerja kelas minoritas rendah. Karena itu keputusan perbandingan model didasarkan pada F1-macro.

Ketiga, ketiga kernel SVM berkinerja setara. Selisih F1-macro terbesar (0,0092) jauh lebih kecil daripada simpangan baku antarlipatan, dan semua uji tidak signifikan dengan ukuran efek sangat kecil. Hal ini wajar karena representasi TF-IDF berdimensi tinggi dan sangat jarang (98,84%), tempat pemisahan linear sudah memadai (Joachims, 1998), sehingga kernel non-linear tidak menambah keuntungan berarti. Karena setara, kernel dapat dipilih berdasarkan efisiensi, dan kernel linear lebih ringan. Perbedaan waktu inferensi per dokumen bersifat mikro dan fluktuatif sehingga hanya arahnya yang layak dibaca. Perbandingan dengan Naive Bayes juga tidak konsisten antarskema evaluasi. Naive Bayes unggul pada satu partisi hold-out dan jauh lebih cepat, sedangkan SVM unggul tipis pada validasi silang tanpa signifikansi. Temuan ini memberi nuansa pada laporan keunggulan SVM berbasis akurasi (Kurniawan dkk., 2023; Ningsih dkk., 2024): keunggulan tersebut tidak otomatis berlaku jika diuji pada metrik yang peka kelas minoritas dan dengan uji statistik.

K keempat, kualitas label membatasi seluruh hasil. InSet menjumlahkan bobot kata tanpa memahami struktur kalimat sehingga kalimat faktual yang memuat kata bermuatan negatif dapat berlabel negatif, sedangkan keluhan berdiskusi sopan atau tanpa kata terdaftar dapat berlabel netral. Karena itu sebagian "kesalahan model" sesungguhnya adalah galat label yang diwariskan. Kelas netral hanya 57 dokumen (11 pada data uji), sehingga metriknya tidak stabil dan menjadi sumber utama besarnya simpangan baku. Penghapusan "naik" dan "buat" oleh stopwords memutus sebagian bigram bermakna, dan bigram secara umum hanya sedikit menembus fitur berbobot tertinggi pada teks pendek ini.

Kelima, prototipe berhasil sebagai demonstrasi teknis, tetapi model di dalamnya belum dapat diandalkan. Keberhasilan 12 skenario black-box membuktikan fungsi perangkat lunak, bukan validitas model. Semua prediksi teks baru jatuh ke kelas negatif dengan keyakinan pada rentang sempit. Ini menunjukkan model kembali ke kecenderungan kelas mayoritas ketika sinyal pembeda tidak tersedia, antara lain karena 20,3% token di luar kosakatanya. Prototipe belum layak dipakai untuk pengambilan keputusan.

Keterbatasan

(1) Label otomatis tanpa validasi manusia. (2) Kelas netral kecil dan residual. (3) Stopword bawaan menghapus kata bermakna domain. (4) Hiperparameter dipilih pada data latih yang sebagian juga muncul sebagai data uji pada validasi silang seluruh dataset, dan lipatan saling tumpang tindih. Estimasi validasi silang karenanya sedikit optimistis dan uji signifikansi bersifat indikatif. (5) Generalisasi pada teks baru rendah. (6) Data berasal dari satu platform, satu periode, dan satu isu. (7) Prototipe belum diuji pada pengguna nyata, dan skor keyakinannya tidak terkalibrasi.

KESIMPULAN

Sentimen publik terhadap kenaikan UKT pada X periode Mei–Juli 2024 didominasi sentimen negatif (84,27%), dengan keberatan terutama pada transparansi penetapan kelompok UKT dan beban finansial keluarga. Pada validasi silang 10 lipatan, F1-macro kernel RBF, linear, dan polynomial adalah 0,6571, 0,6520, dan 0,6479, dan perbedaannya tidak signifikan secara statistik. Ketiga kernel dapat dianggap setara pada data ini, begitu pula SVM terhadap Multinomial Naive Bayes (0,6373). Prototipe web berfungsi pada seluruh 12 skenario pengujian dengan waktu respons rata-rata 10,5 milidetik, tetapi hanya 4 dari 10 teks baru yang diprediksi sesuai penilaian manusia, karena seluruhnya diprediksi negatif. Model karena itu belum layak dipakai di luar distribusi data latihnya.

Saran untuk penelitian lanjutan: (1) validasi manual sebagian label oleh dua pelabel dengan Cohen's Kappa; (2) penyeimbangan kelas, misalnya SMOTE hanya pada data latih di setiap lipatan, atau pengumpulan data netral dan positif; (3) kurasi daftar stopwords; (4) validasi silang bersarang (nested) untuk penyeteraan dan evaluasi; (5) perluasan data dan ragam ekspresi; (6) perbandingan dengan model transformer seperti IndoBERTweet (Koto dkk., 2021) pada dataset yang sama; dan (7) kalibrasi probabilitas pada prototipe.

DAFTAR PUSTAKA

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7, 1–30.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923. <https://doi.org/10.1162/089976698300017197>
- Gulo, N. K. J., & Purwanto. (2024). Analisis sentimen masyarakat terhadap kebijakan kenaikan Uang Kuliah Tunggal (UKT) pada media sosial X menggunakan metode Naïve Bayes Classifier. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 3(2), 714–722. <https://senafiti.budiluhur.ac.id/index.php/senafiti/index>

- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In C. Nédellec & C. Rouveirol (Eds.), *Machine Learning: ECML-98* (pp. 137–142). Springer. <https://doi.org/10.1007/BFb0026683>
- Kemendikbudristek. (2024, 27 Mei). *Mendikbudristek batalkan kenaikan UKT* [Siaran pers]. <https://vokasi.kemdikbud.go.id/read/b/mendikbudristek-batalkan-kenaikan-ukt>
- Koto, F., Lau, J. H., & Baldwin, T. (2021). IndoBERTweet: A pretrained language model for Indonesian Twitter with effective domain-specific vocabulary initialization. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 10660–10668. <https://aclanthology.org/2021.emnlp-main.833/>
- Koto, F., & Rahmaningtyas, G. Y. (2017). InSet lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *2017 International Conference on Asian Language Processing (IALP)*, 391–394. <https://doi.org/10.1109/IALP.2017.8300625>
- Kurniawan, I., Hananto, A. L., Hilabi, S. S., Hananto, A., Priyatna, B., & Rahman, A. Y. (2023). Perbandingan algoritma Naive Bayes dan SVM dalam sentimen analisis marketplace pada Twitter. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 10(1), 731–740. <https://doi.org/10.35957/jatisi.v10i1.3582>
- Maharani, D., Atmia, K., & Novita, S. (2025). Blynk-based air handling unit (AHU) air filter monitoring system prototype at Juwata Tarakan Airport Terminal. *Indonesian Journal of Multidisciplinary Scientific Studies*, 3(5), 392–397. <https://doi.org/10.33151/ijomss.v3i5.696>
- Ningsih, W., Alfiana, B., Rahmaddeni, R., & Wulandari, D. (2024). Perbandingan algoritma SVM dan Naïve Bayes dalam analisis sentimen Twitter pada penggunaan mobil listrik di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 556–562. <https://doi.org/10.57152/malcom.v4i2.1253>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Raiya, M. S., Khamil, M. R. P., Fadillah, N., & Saputra, R. A. (2025). Implementasi algoritma Long Short-Term Memory pada isu kenaikan Uang Kuliah Tunggal terhadap minat kuliah mahasiswa. *Jurnal Informatika Terpadu*, 11(1), 37–43. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Sumartha, D. K. (2025). Implementasi IndoBERT untuk analisis sentimen opini publik terhadap kebijakan kenaikan UKT di era pemerintahan. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1), 867–875. <https://doi.org/10.23960/jitet.v13i3S1.7880>
- Tala, F. Z. (2003). *A study of stemming effects on information retrieval in Bahasa Indonesia* [Master's thesis, Universiteit van Amsterdam].