



Analisis Sentimen Efektivitas Pembelajaran Daring Menggunakan Hybrid Naive Bayes dan SVM

Faried Prayugo Dwipa Putra^{1*}, Al-Khowarizmi²

^{1,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: ^{1*}fariedprayugodwipaputra@gmail.com, ²alkhowarizmi@umsu.ac.id

Abstrak

Pembelajaran daring berkembang pesat pascapandemi COVID-19, namun efektivitasnya masih menimbulkan pro dan kontra di kalangan masyarakat Indonesia, yang banyak diungkapkan melalui opini pada aplikasi X. Volume opini yang besar dan bersifat tidak terstruktur menyulitkan pemantauan kecenderungan sentimen publik secara manual sehingga diperlukan metode klasifikasi otomatis. Penelitian ini bertujuan menganalisis sentimen publik mengenai efektivitas pembelajaran daring di Indonesia menggunakan pendekatan *hybrid Multinomial Naive Bayes* dan *Support Vector Machine* (SVM) melalui mekanisme soft voting. Data penelitian berupa 2.655 tweet berbahasa Indonesia yang dikumpulkan melalui crawling dan dilabeli ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif, kemudian diproses melalui *case folding*, *cleansing*, penanganan negasi, *stopword removal*, stemming, dan pembobotan TF-IDF, lalu dibagi dengan proporsi 75:25 menggunakan stratifikasi kelas. Hasil pengujian menunjukkan model hybrid mencapai *accuracy* sebesar 76,05%, *precision* berbobot 69,54%, *recall* berbobot 76,05%, dan F1-score berbobot 72,10%, lebih tinggi dibandingkan model tunggal Naive Bayes (62,90%) dan SVM (70,43%). Meskipun demikian, model masih mengalami kesulitan mengenali kelas netral akibat ketidakseimbangan jumlah data antarkelas. Penelitian ini menyimpulkan bahwa pendekatan hybrid soft voting mampu meningkatkan performa klasifikasi sentimen secara keseluruhan, tetapi penanganan kelas minoritas masih memerlukan perbaikan lebih lanjut.

Kata Kunci: analisis sentimen, pembelajaran daring, Naive Bayes, Support Vector Machine, TF-IDF.

PENDAHULUAN

Perubahan pola pembelajaran pascapandemi COVID-19 mendorong penggunaan pembelajaran daring sebagai salah satu alternatif utama dalam pelaksanaan pendidikan di Indonesia. Model pembelajaran ini memungkinkan proses belajar mengajar berlangsung tanpa keharusan bertatap muka secara langsung di ruang kelas. Meskipun penggunaannya terus meningkat, penerapan pembelajaran daring masih menghadapi berbagai kendala, seperti kesiapan peserta didik, keterbatasan fasilitas, kualitas jaringan internet, serta kemampuan menyesuaikan diri dengan lingkungan belajar berbasis digital, sebagaimana dicatat pada sejumlah kajian implementasi pembelajaran daring di lapangan (Murzal dan Ridwan, 2021; Nurulita dan Hakim, 2021). Kondisi tersebut menyebabkan efektivitas pembelajaran daring masih menjadi persoalan yang terus dipertanyakan oleh berbagai kalangan.

Penerapan pembelajaran daring menimbulkan beragam tanggapan dari masyarakat yang banyak disampaikan melalui media sosial, termasuk aplikasi X. Unggahan pengguna pada platform tersebut kerap menggambarkan pengalaman maupun penilaian mereka terhadap pelaksanaan pembelajaran daring, baik berupa dukungan, keluhan, maupun informasi yang bersifat netral. Kumpulan opini tersebut berpotensi dimanfaatkan sebagai sumber data untuk melihat kecenderungan persepsi masyarakat, tetapi volumenya yang besar dan sifatnya yang tidak terstruktur menyulitkan proses pengamatan secara manual. Oleh karena itu, diperlukan suatu metode yang mampu mengelompokkan opini tersebut ke dalam kategori sentimen tertentu secara otomatis dan terukur.

Analisis sentimen merupakan cabang pengolahan bahasa alami yang digunakan untuk mengidentifikasi kecenderungan opini suatu teks ke dalam kategori positif, negatif, atau netral. Salah satu algoritma yang umum digunakan dalam klasifikasi teks adalah Naive Bayes, yaitu metode berbasis probabilitas yang mengasumsikan independensi antarfitur sehingga relatif sederhana dan efisien dalam menangani data teks berdimensi tinggi. Di sisi lain, *Support Vector Machine* (SVM) bekerja dengan mencari hyperplane optimal yang mampu memisahkan kelas-kelas sentimen dan umumnya menghasilkan

akurasi yang kompetitif pada data hasil ekstraksi fitur seperti *Term Frequency–Inverse Document Frequency* (TF-IDF). Karena keduanya memiliki karakteristik yang saling melengkapi, penelitian ini menggabungkan kedua algoritma melalui pendekatan hybrid dengan mekanisme *soft voting*, sehingga prediksi akhir mempertimbangkan probabilitas keluaran dari kedua model secara bersamaan.

Sejumlah penelitian terdahulu telah menerapkan algoritma Naive Bayes dan SVM untuk klasifikasi teks maupun analisis sentimen pada berbagai konteks. Santoso dkk. (2020) membandingkan performa Naive Bayes dan SVM dalam analisis sentimen terhadap produk SecBowl di *Twitter* dan menemukan bahwa Naive Bayes memberikan proses klasifikasi yang lebih cepat, sedangkan SVM memberikan akurasi yang lebih tinggi, terutama pada data berukuran besar. Di luar konteks analisis sentimen, algoritma Naive Bayes juga banyak diterapkan pada berbagai domain klasifikasi lain, seperti klasifikasi karakteristik kepribadian siswa (Martantoh dan Yanih, 2022) dan penentuan tingkat kenyamanan pasien rumah sakit (Nizam Fadli dkk., 2021), yang menunjukkan fleksibilitas algoritma tersebut pada berbagai jenis data. Fajar Fiandaru dan Romando Sipayung (2025) menerapkan metode Naive Bayes untuk menganalisis sentimen terhadap Undang-Undang ITE pada *Twitter* dan memperoleh akurasi hingga 94% dengan dominasi sentimen netral. Zusrotun dkk. (2022) menerapkan Naive Bayes untuk analisis sentimen terhadap pembelajaran daring dan memperoleh akurasi sebesar 74,08%. Siti Aisah dkk. (2024) menerapkan algoritma SVM untuk analisis sentimen ulasan aplikasi Al-Qur'an digital, sedangkan Syahri dan Puspita (2025) menerapkan SVM untuk klasifikasi data mahasiswa berprestasi berbasis data mining. Studi lain oleh Laia dkk. (2025) menerapkan Naive Bayes dengan TF-IDF untuk menganalisis sentimen publik pada platform X, sedangkan Oktafiani dkk. (2025) menerapkan Naive Bayes untuk analisis sentimen layanan akademik dengan akurasi 86,74%.

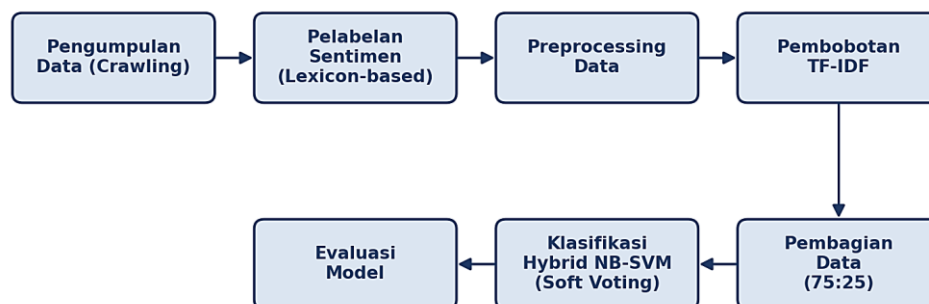
Berdasarkan tinjauan tersebut, sebagian besar penelitian terdahulu masih menerapkan algoritma Naive Bayes atau SVM secara tunggal, sedangkan penelitian yang menggabungkan kedua algoritma melalui pendekatan hybrid pada konteks opini publik mengenai efektivitas pembelajaran daring di aplikasi X masih terbatas. Kesenjangan tersebut menjadi dasar bagi penelitian ini untuk mengembangkan model klasifikasi sentimen yang menggabungkan Multinomial Naive Bayes dan Linear SVM melalui mekanisme *soft voting*, dengan harapan memperoleh performa klasifikasi yang lebih baik dibandingkan penggunaan masing-masing algoritma secara tunggal. Kebaruan dalam penelitian ini terletak pada penerapan pendekatan hybrid tersebut secara khusus pada data opini publik berbahasa Indonesia mengenai efektivitas pembelajaran daring yang bersumber dari aplikasi X, disertai penanganan kata negasi untuk menjaga makna sentimen pada tahap praproses.

Permasalahan utama dalam penelitian ini adalah bagaimana opini publik yang tidak terstruktur pada aplikasi X dapat diklasifikasikan secara otomatis ke dalam kategori sentimen positif, negatif, dan netral, serta bagaimana performa pendekatan hybrid Naive Bayes dan SVM dibandingkan dengan penggunaan masing-masing algoritma secara tunggal. Sebagai solusi, penelitian ini menerapkan alur kerja yang mencakup pengumpulan data, pelabelan sentimen berbasis leksikon, praproses teks, pembobotan TF-IDF, klasifikasi menggunakan pendekatan hybrid dengan mekanisme *soft voting*, serta evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan F1-score. Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis sentimen publik mengenai efektivitas pembelajaran daring di Indonesia pada aplikasi X serta membandingkan performa algoritma Naive Bayes, SVM, dan pendekatan hybrid keduanya dalam melakukan klasifikasi sentimen tersebut.

METODE

Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional, dilaksanakan melalui tujuh tahapan utama, yaitu pengumpulan data, pelabelan sentimen, praproses data, pembobotan fitur menggunakan TF-IDF, pembagian data latih dan data uji, klasifikasi menggunakan pendekatan hybrid Naive Bayes dan SVM, serta evaluasi model. Alur tahapan penelitian secara keseluruhan disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan bahwa proses penelitian dimulai dari pengumpulan data *tweet* berbahasa Indonesia melalui *crawling*, dilanjutkan dengan pelabelan sentimen berbasis leksikon, praproses teks, pembobotan TF-IDF, pembagian data, klasifikasi menggunakan pendekatan hybrid dengan mekanisme *soft voting*, hingga evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan F1-score.

Dataset

Data penelitian merupakan data primer berupa *tweet* berbahasa Indonesia yang dihimpun dari aplikasi X (dahulu Twitter) mengenai efektivitas pembelajaran daring di Indonesia. Objek penelitian adalah opini publik pengguna aplikasi X yang berkaitan dengan topik tersebut. Pengumpulan data dilakukan melalui proses *crawling* menggunakan kata kunci “kuliah daring”, “belajar *online*”, dan “sekolah daring” pada rentang tahun 2020 hingga awal 2026, dengan memanfaatkan *tool crawler* berbasis Python. Setiap data hasil *crawling* terdiri atas atribut nama pengguna, tanggal unggahan, isi *tweet*, jumlah *like*, *reply count*, dan jumlah *repost*.

Dari total 3.779 data hasil *crawling*, dilakukan seleksi terhadap data duplikat dan data yang tidak relevan dengan topik pembelajaran daring, sehingga diperoleh 2.761 data. Data tersebut kemudian dilabeli ke dalam tiga kelas sentimen (positif, netral, negatif) menggunakan metode berbasis leksikon (*lexicon-based*), yaitu setiap *tweet* diberi skor sentimen berdasarkan bobot kata bermuatan sentimen yang kemudian dipetakan ke salah satu kelas beserta nilai *confidence score*. Data yang teridentifikasi mengandung sarkasme, yaitu *tweet* yang secara tekstual tampak berlawanan dengan makna sebenarnya, sebanyak 106 data, kemudian dihapus dari dataset karena berpotensi menurunkan kualitas pelatihan model. Rincian jumlah data pada setiap tahap seleksi disajikan pada Tabel 1.

Tabel 1. Jumlah Data pada Setiap Tahap Seleksi

Keterangan	Jumlah Data
Total data hasil <i>crawling</i>	3.779
Data setelah penghapusan duplikat/tidak relevan	2.761
Data terindikasi sarkasme (dihapus)	106
Total data yang digunakan dalam penelitian	2.655

Berdasarkan Tabel 1, dataset akhir yang digunakan dalam penelitian berjumlah 2.655 data. Distribusi data tersebut ke dalam tiga kelas sentimen disajikan pada Tabel 2.

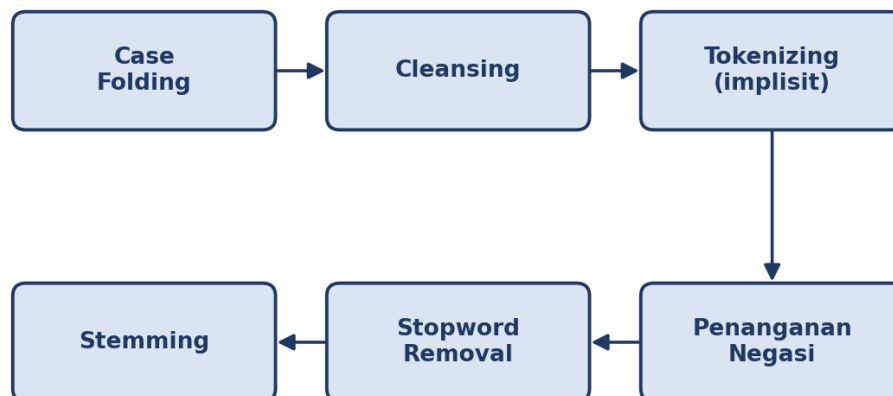
Tabel 2. Distribusi Data per Kelas Sentimen

Kelas Sentimen	Jumlah Data	Persentase
Positif	486	18,31%
Netral	179	6,74%
Negatif	1.990	74,95%
Total	2.655	100%

Sebagaimana ditunjukkan pada Tabel 2, distribusi kelas sentimen pada dataset penelitian ini tidak seimbang (*imbalanced*), dengan kelas negatif mendominasi sebesar 74,95% dari keseluruhan data, sedangkan kelas netral hanya sebesar 6,74%. Visualisasi distribusi data disajikan pada Gambar 4 (subbagian Distribusi Dataset). Kondisi ini menjadi pertimbangan penting dalam proses pelatihan model karena berpotensi menyebabkan model lebih mudah mempelajari kelas mayoritas.

Preprocessing Data

Data yang telah dilabeli selanjutnya melalui tahap praproses (*preprocessing*) menggunakan pustaka Sastrawi untuk *stopword removal* dan *stemming* berbahasa Indonesia, serta modul regular expression bawaan Python untuk tahap *cleansing*. Tahapan praproses yang diterapkan pada fungsi *preprocess text()* terdiri atas enam langkah, sebagaimana digambarkan pada Gambar 2.



Gambar 2. Alur Preprocessing Data

1. *Case Folding* — mengubah seluruh huruf kapital pada teks menjadi huruf kecil agar tidak terjadi perbedaan perlakuan terhadap kata yang sama akibat perbedaan kapitalisasi.
2. *Cleansing* — menghapus atribut yang tidak relevan pada teks, seperti tautan (URL), *mention* (@username), tagar (#), angka, tanda baca, dan spasi berlebih, menggunakan ekspresi reguler.
3. *Tokenizing* — memecah kalimat menjadi satuan kata (token) berdasarkan spasi; pada implementasi sistem, proses ini dilakukan secara implisit oleh fungsi `split()` serta secara internal oleh pustaka Sastrawi pada tahap stopword removal dan stemming.
4. Penanganan Kata Negasi — kata negasi bahasa Indonesia (baik baku maupun gaya percakapan, seperti “tidak”, “gak”, “belum”, “bukan”) digabungkan dengan kata setelahnya menjadi satu token, misalnya “tidak efektif” menjadi “negasi_efektif”, agar makna negasi tidak hilang pada tahap stopword removal dan stemming. Kamus normalisasi kata tidak baku/slang secara umum (misalnya “gpp” menjadi “tidak apa-apa”) belum diimplementasikan secara eksplisit pada kode sistem saat ini.
5. *Stopword Removal* — menghapus kata-kata umum yang tidak memberikan kontribusi berarti terhadap makna sentimen menggunakan daftar 126 kata stopword bawaan pustaka Sastrawi (*Stop Word Remover Factory*).
6. *Stemming* — mengembalikan kata ke bentuk dasarnya dengan menghilangkan imbuhan, menggunakan pustaka Sastrawi (*StemmerFactory*).

Contoh hasil masing-masing tahap praproses disajikan pada Tabel 5 (subbagian Hasil Preprocessing Data).

Pembobotan TF-IDF

Data hasil praproses selanjutnya direpresentasikan ke dalam bentuk vektor numerik melalui pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) menggunakan `TfidfVectorizer` dari pustaka Scikit-learn, dengan parameter `max_features=5000`, `ngram_range=(1,2)` (unigram dan bigram), `min_df=2`, `max_df=0,9`, dan `sublinear_tf=True`. Konfigurasi tersebut menghasilkan 5.000 kosakata unik (vocabulary) sebagai fitur klasifikasi, mencapai batas maksimum yang dikonfigurasi pada vectorizer. Perbandingan hasil perhitungan TF-IDF secara manual dan otomatis disajikan pada Tabel 6 (subbagian Hasil Pembobotan TF-IDF).

Pembagian Dataset

Data hasil pembobotan TF-IDF dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split` dari pustaka Scikit-learn, dengan proporsi 75:25 (`test_size=0,25`), pembagian acak (`random_state=42`), dan stratifikasi berdasarkan kelas label (`stratify=y`) agar proporsi setiap kelas sentimen pada data latih dan data uji tetap terjaga. Hasil pembagian data disajikan pada Tabel 3.

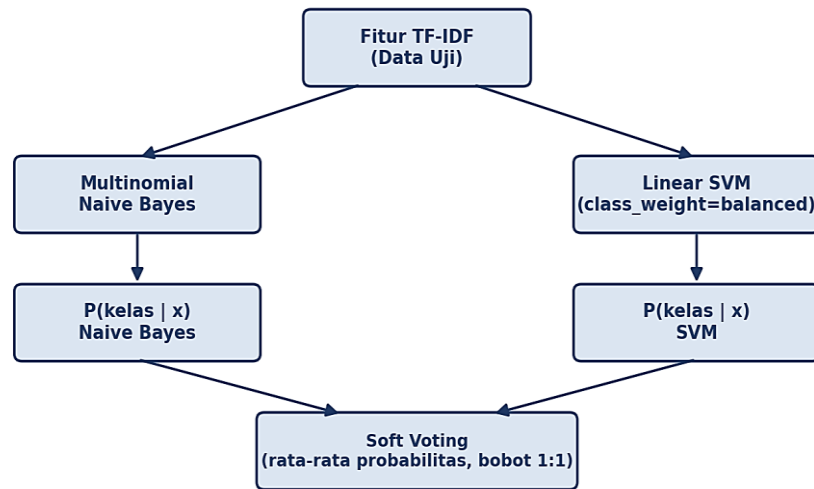
Tabel 3. Pembagian Data Latih dan Data Uji

Kelas Sentimen	Data Latih	Data Uji	Total
Positif	365	121	486
Netral	134	45	179
Negatif	1.492	498	1.990
Total	1.991	664	2.655

Klasifikasi Hybrid Naive Bayes dan SVM

Naive Bayes merupakan algoritma klasifikasi probabilistik yang mengasumsikan independensi antarfitur; berdasarkan probabilitas yang diperoleh melalui Teorema Bayes, dokumen ditempatkan pada kelas dengan nilai probabilitas tertinggi (Martantoh dan Yanih, 2022; Nizam Fadli dkk., 2021). *Support Vector Machine* (SVM) bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas sentimen pada ruang fitur, dan umumnya efektif pada data berdimensi tinggi seperti hasil ekstraksi fitur TF-IDF.

Proses klasifikasi pada penelitian ini menggunakan pendekatan hybrid yang menggabungkan Multinomial Naive Bayes dan Linear SVM melalui mekanisme *soft voting* menggunakan *VotingClassifier* dengan parameter `voting=“soft”`. Kedua algoritma dilatih secara terpisah menggunakan representasi fitur TF-IDF yang sama pada data latih, kemudian probabilitas keluaran dari kedua model dirata-ratakan dengan bobot kontribusi yang setara (1:1) untuk menentukan kelas akhir berdasarkan nilai probabilitas gabungan tertinggi. Arsitektur mekanisme hybrid ini disajikan pada Gambar 3.



Gambar 3. Arsitektur Klasifikasi Hybrid Naive Bayes dan SVM (*Soft voting*)

Untuk mengurangi pengaruh ketidakseimbangan kelas pada data latih, pelatihan model menerapkan pembobotan sampel menggunakan `compute_sample_weight(class_weight="balanced")`, agar kontribusi data dari setiap kelas menjadi lebih proporsional. Parameter yang digunakan pada masing-masing algoritma disajikan pada Tabel 4.

Tabel 4. Parameter Model Hybrid Naive Bayes dan SVM

Algoritma	Parameter	Nilai
Multinomial Naive Bayes	alpha	1,0 (default)
Linear SVM	C	1,0 (default)
Linear SVM	kernel	linear
Linear SVM	class_weight	balanced
<i>Soft voting</i>	bobot (weights)	setara / 1:1 (default)

Model yang telah dilatih disimpan menggunakan `joblib.dump` ke dalam berkas `model.pkl` dan `vectorizer.pkl` agar dapat digunakan kembali pada proses prediksi tanpa perlu dilakukan pelatihan ulang.

Evaluasi Model

Evaluasi kinerja model dilakukan dengan menghitung *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan F1-score pada data uji. *Confusion matrix* merupakan tabel yang membandingkan label sebenarnya dengan hasil prediksi model, sehingga jumlah prediksi benar maupun kesalahan klasifikasi pada setiap kelas dapat diamati (Syahri dan Puspita, 2025). *Accuracy* menunjukkan proporsi prediksi yang berhasil diklasifikasikan dengan benar terhadap seluruh data uji (Clara dkk., 2021), dihitung dengan rumus:

$$Accuracy = (\text{Jumlah Prediksi Benar} / \text{Jumlah Total Data Uji}) \times 100\% \quad (1)$$

Precision mengukur proporsi prediksi benar pada suatu kelas dibandingkan dengan seluruh data yang diprediksi sebagai kelas tersebut:

$$Precision = TP / (TP + FP) \times 100\% \quad (2)$$

Recall mengukur kemampuan model dalam mengenali seluruh data yang sebenarnya termasuk dalam suatu kelas:

$$Recall = TP / (TP + FN) \times 100\% \quad (3)$$

F1-score merupakan rata-rata harmonis antara *precision* dan *recall*, digunakan untuk menyeimbangkan kedua metrik tersebut (Pinem dan Putra, 2025):

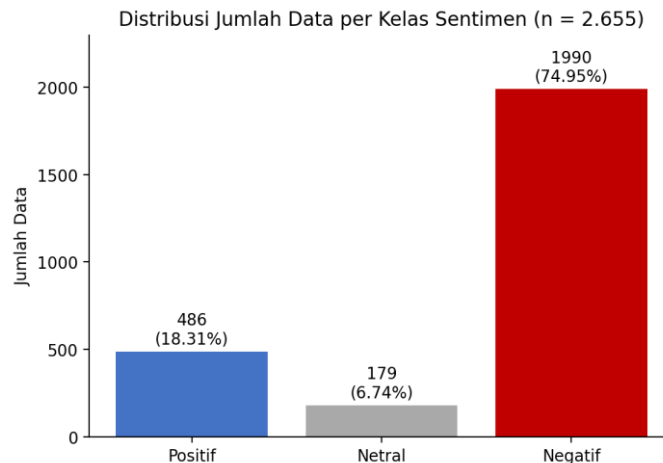
$$F1\text{-Score} = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

Evaluasi dilakukan pada tingkat per kelas, rata-rata makro (*macro average*), dan rata-rata berbobot (*weighted average*). Hasil evaluasi model hybrid selanjutnya dibandingkan dengan hasil evaluasi model Multinomial Naive Bayes tunggal dan model Linear SVM tunggal untuk mengetahui pengaruh pendekatan hybrid terhadap performa klasifikasi sentimen.

HASIL DAN PEMBAHASAN

Distribusi Dataset

Proses pengumpulan dan seleksi data menghasilkan 2.655 data yang digunakan dalam penelitian, sebagaimana telah disajikan pada Tabel 1 dan Tabel 2 (subbagian Dataset). Visualisasi distribusi jumlah data pada setiap kelas sentimen disajikan pada Gambar 4.



Gambar 4. Distribusi Jumlah Data per Kelas Sentimen

Gambar 4 menegaskan bahwa dataset penelitian bersifat tidak seimbang (*imbalanced*), dengan kelas negatif mendominasi sebesar 74,95% dari keseluruhan data, sedangkan kelas netral hanya sebesar 6,74%. Kondisi ini menjadi salah satu faktor yang memengaruhi kemampuan model dalam mengenali kelas minoritas, sebagaimana dibahas lebih lanjut pada subbagian Pembahasan.

Hasil Preprocessing Data

Contoh perubahan data pada setiap tahap praproses, sebagaimana diuraikan pada subbagian Preprocessing Data, disajikan pada Tabel 5.

Tabel 5. Contoh Hasil Setiap Tahap Preprocessing

Tahap	Input	Output
Case Folding	Kuliah <i>ONLINE</i> terus bikin capek	kuliah <i>online</i> terus bikin capek
Cleansing	Kuliah <i>ONLINE</i> terus bikin capek, ada yg tau <i>link zoom</i> hari ini? https://zoom.us/j/12345 #kuliahdaring @dosen123	ada yg tau <i>link zoom</i> hari ini
Tokenizing	kuliah <i>ONLINE</i> terus bikin capek yg tau <i>link zoom</i> hari ini	['kuliah','online','terus','bikin','capek','yg','tau','link','zoom','hari','ini']
Penanganan Negasi	belajar daring itu tidak efektif sama sekali	belajar daring itu negasi_efektif sama sekali
Stopword Removal	kuliah <i>ONLINE</i> terus bikin capek yg tau <i>link zoom</i> hari ini	Kuliah <i>online</i> terus bikin capek yg tau <i>link zoom</i> hari
Stemming	belajar daring negasi_efektif sama sekali kuota rbbulan negasi cukup	ajar daring negasi efektif sama sekali kuota rbbulan negasi cukup

Seluruh contoh pada Tabel 5 diambil langsung dari data hasil pengolahan sistem sebagaimana didokumentasikan pada skripsi. Tahap penanganan negasi memastikan makna “tidak efektif” tidak berubah menjadi “efektif” setelah kata negasi “tidak” dihapus pada tahap *stopword removal*, sehingga polaritas sentimen kalimat tetap terjaga.

Hasil Pembobotan TF-IDF

Perbandingan hasil perhitungan TF-IDF secara manual (menggunakan formula $TF = \text{jumlah kemunculan term pada dokumen} / \text{total token dalam dokumen}$ dan $IDF = \ln(N/df)$, dengan $N = 2.655$ dokumen) dan hasil perhitungan otomatis oleh *TfidfVectorizer* (yang menerapkan smoothing pada IDF, skala sublinear pada TF, serta normalisasi vektor L2 per dokumen) disajikan pada Tabel 6.

Tabel 6. Perbandingan Perhitungan TF-IDF Manual dan Otomatis

Term	TF	DF	IDF	TF-IDF Manual	TF-IDF Otomatis
daring	1	298	2,1871	0,2430	0,1440
mbg	1	340	2,0553	0,1370	0,1121
hemat	1	64	3,7253	0,1129	0,1175
kuota	1	52	3,9330	0,0959	0,1174

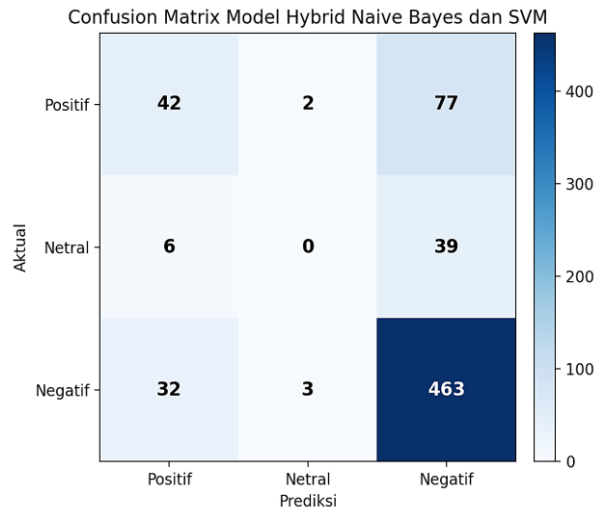
Sebagaimana ditunjukkan pada Tabel 6, nilai TF-IDF otomatis tidak identik dengan perhitungan manual meskipun berada pada rentang skala yang berdekatan, karena *TfidfVectorizer* menerapkan smoothing pada IDF ($IDF = \ln((1+N)/(1+df)) + 1$), skala sublinear pada TF ($1 + \log(TF)$), serta normalisasi vektor secara L2 per dokumen.

Confusion Matrix

Model hybrid Naive Bayes dan SVM dilatih menggunakan 1.991 data latih dan diuji terhadap 664 data uji (Tabel 3). Hasil pengujian dalam bentuk *confusion matrix* disajikan pada Tabel 7 dan Gambar 5.

Tabel 7. *Confusion Matrix* Model Hybrid Naive Bayes dan SVM

Aktual \ Prediksi	Positif	Netral	Negatif
Positif	42	2	77
Netral	6	0	39
Negatif	32	3	463



Gambar 5. Confusion Matrix Model Hybrid Naive Bayes dan SVM

Berdasarkan *confusion matrix* pada Tabel 7, jumlah prediksi benar diperoleh dari penjumlahan nilai diagonal utama, yaitu 42 data positif, 0 data netral, dan 463 data negatif, sehingga $accuracy = (42 + 0 + 463) / 664 \times 100\% = 76,05\%$.

Hasil Pengujian Model

Hasil evaluasi lengkap berupa *precision*, *recall*, dan F1-score untuk setiap kelas disajikan pada Tabel 8.

Tabel 8. Hasil Evaluasi Model Hybrid Naive Bayes dan SVM per Kelas

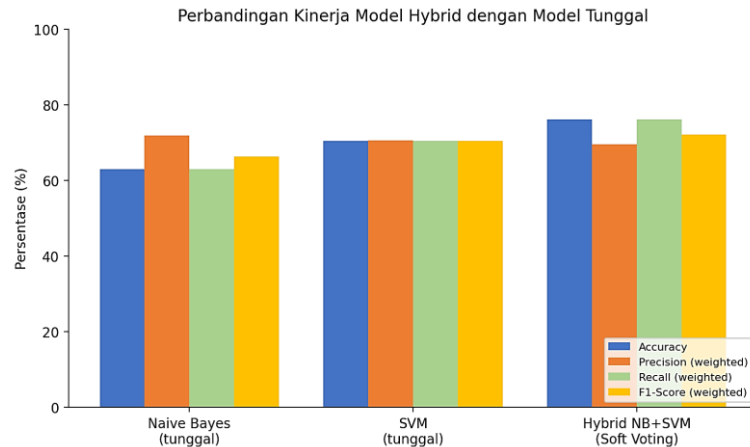
Kelas Sentimen	<i>Precision</i>	<i>Recall</i>	F1-Score
Positif	52,50%	34,71%	41,79%
Netral	0,00%	0,00%	0,00%
Negatif	79,97%	92,97%	85,98%
<i>Macro Average</i>	44,16%	42,56%	42,59%
<i>Weighted Average</i>	69,54%	76,05%	72,10%

Seperti disajikan pada Tabel 8, kelas negatif memperoleh *precision*, *recall*, dan F1-score tertinggi, masing-masing sebesar 79,97%, 92,97%, dan 85,98%, sedangkan kelas netral memperoleh nilai 0% pada ketiga metrik tersebut karena seluruh 45 data uji pada kelas ini tidak berhasil diprediksi sebagai netral. Sebagai perhitungan contoh pada kelas negatif: $precision = 463 / (463 + 39 + 77) \times 100\% = 79,97\%$; $recall = 463 / (463 + 3 + 32) \times 100\% = 92,97\%$; dan $F1-score = 2 \times (79,97\% \times 92,97\%) / (79,97\% + 92,97\%) = 85,98\%$.

Perbandingan kinerja model hybrid dengan model tunggal Naive Bayes dan SVM disajikan pada Tabel 9 dan Gambar 6.

Tabel 9. Perbandingan Kinerja Model Hybrid dengan Model Tunggal

Model	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	F1-Score
Multinomial Naive Bayes (tunggal)	62,90%	71,80%	62,90%	66,35%
Linear SVM (tunggal)	70,43%	70,50%	70,43%	70,42%
Hybrid Naive Bayes + SVM (<i>Soft voting</i>)	76,05%	69,54%	76,05%	72,10%



Gambar 6. Perbandingan Kinerja Model Hybrid dengan Model Tunggal

Sebagaimana disajikan pada Tabel 9 dan Gambar 6, model hybrid Naive Bayes dan SVM memperoleh *accuracy* sebesar 76,05%, lebih tinggi dibandingkan model Naive Bayes tunggal sebesar 62,90% dan model SVM tunggal sebesar 70,43%. Namun demikian, nilai *weighted precision* model hybrid (69,54%) sedikit lebih rendah dibandingkan model Naive Bayes tunggal (71,80%), meskipun tetap lebih tinggi dari SVM tunggal (70,50%). Kondisi ini menunjukkan bahwa peningkatan performa model hybrid tidak terjadi secara merata pada seluruh metrik evaluasi.

Perbandingan dengan Penelitian Terdahulu

Untuk melihat posisi hasil penelitian ini terhadap penelitian sejenis, Tabel 10 merangkum perbandingan *accuracy* pada penelitian-penelitian terdahulu yang telah disebutkan pada Pendahuluan, sepanjang angka tersebut dilaporkan secara eksplisit oleh masing-masing penelitian.

Tabel 10. Perbandingan dengan Penelitian Terdahulu

No.	Penelitian	Metode	Accuracy	Catatan
1	Santoso dkk. (2020)	Naive Bayes vs SVM (tunggal)	–	SVM lebih akurat, NB lebih cepat, topik produk <i>SecBowl</i> di Twitter
2	Fajar Fiandaru dan Sipayung (2025)	Naive Bayes	94%	Topik UU ITE, dominasi kelas netral
3	Zusrotun dkk. (2022)	Naive Bayes	74,08%	Topik belajar <i>ONLINE</i> di Twitter
4	Laia dkk. (2025)	Naive Bayes + TF-IDF	–	Topik platform X
5	Oktafiani dkk. (2025)	Naive Bayes	86,74%	Topik layanan akademik
6	Penelitian ini	Hybrid Naive Bayes + SVM (<i>Soft voting</i>)	76,05%	Topik pembelajaran daring di aplikasi X, tiga kelas sentimen

Berdasarkan Tabel 10, *accuracy* yang diperoleh pada penelitian ini (76,05%) berada pada rentang yang sebanding dengan penelitian Zusrotun dkk. (2022) yang memperoleh akurasi 74,08% pada topik serupa, tetapi lebih rendah dibandingkan penelitian Fajar Fiandaru dan Sipayung (2025) yang mencapai 94% maupun penelitian Oktafiani dkk. (2025) yang mencapai 86,74%. Perbedaan ini dapat mengindikasikan bahwa tingkat ketidakseimbangan kelas serta karakteristik topik yang lebih beragam pada dataset penelitian ini turut memengaruhi performa model, sehingga akurasi yang tinggi pada satu domain tidak selalu dapat direplikasi pada domain lain dengan distribusi data yang berbeda.

Pembahasan

Hasil evaluasi menunjukkan bahwa pendekatan hybrid melalui mekanisme *soft voting* memberikan peningkatan performa klasifikasi secara keseluruhan dibandingkan penggunaan Naive Bayes atau SVM secara tunggal, khususnya pada metrik *accuracy* dan *weighted recall*. Peningkatan ini sejalan dengan karakteristik kedua algoritma yang saling melengkapi, yaitu Naive Bayes yang efisien menangani fitur berdimensi tinggi dan SVM yang lebih baik dalam menentukan batas pemisah antarkelas, sehingga penggabungan probabilitas keluaran keduanya menghasilkan keputusan klasifikasi yang lebih stabil pada kelas mayoritas.

Meskipun demikian, model hybrid belum mampu mengenali kelas netral, yang ditunjukkan oleh nilai *precision*, *recall*, dan F1-score sebesar 0% pada kelas tersebut. Kondisi ini kemungkinan besar berkaitan dengan distribusi data yang tidak seimbang, di mana kelas netral hanya memiliki 134 data latih atau sekitar 6,74% dari total data, jauh lebih sedikit dibandingkan kelas negatif yang mendominasi.

Penerapan pembobotan sampel melalui `class_weight="balanced"` pada kedua algoritma belum sepenuhnya mampu mengatasi ketidakseimbangan tersebut, sehingga model cenderung mengarahkan prediksi kelas netral ke kelas mayoritas, terutama kelas negatif, sebagaimana terlihat pada Tabel 7 bahwa 39 dari 45 data netral justru diprediksi sebagai negatif.

Temuan ini juga menegaskan bahwa nilai *accuracy* yang tinggi tidak selalu mencerminkan kemampuan model dalam mengenali seluruh kelas secara seimbang, sehingga interpretasi hasil perlu mempertimbangkan *Macro Average* yang memberikan bobot setara pada setiap kelas (Tabel 8). Kontribusi utama penelitian ini terletak pada pembuktian bahwa pendekatan hybrid *soft voting* antara Naive Bayes dan SVM dapat meningkatkan performa klasifikasi secara agregat pada opini publik berbahasa Indonesia mengenai pembelajaran daring, sekaligus mengungkap keterbatasan pendekatan tersebut dalam menangani kelas minoritas. Keterbatasan penelitian ini di antaranya adalah belum diterapkannya normalisasi kata tidak baku dan slang secara eksplisit pada tahap praproses, serta belum diterapkannya teknik penyeimbangan data seperti *oversampling*, sehingga kemampuan model dalam mengenali kelas netral masih perlu ditingkatkan pada penelitian selanjutnya.

KESIMPULAN

Penelitian ini mengembangkan model klasifikasi sentimen publik mengenai efektivitas pembelajaran daring di Indonesia pada aplikasi X menggunakan pendekatan hybrid Multinomial Naive Bayes dan *Support Vector Machine* melalui mekanisme *soft voting*, dengan data penelitian berupa 2.655 tweet berbahasa Indonesia yang telah melalui tahap pelabelan berbasis leksikon, praproses teks, dan pembobotan TF-IDF. Hasil pengujian menunjukkan bahwa model hybrid memperoleh *accuracy* sebesar 76,05%, *weighted recall* sebesar 76,05%, dan *weighted F1-score* sebesar 72,10%, lebih tinggi dibandingkan model Naive Bayes tunggal (62,90%) dan model SVM tunggal (70,43%), sehingga menjawab tujuan penelitian untuk membandingkan performa ketiga pendekatan tersebut dalam mengklasifikasikan sentimen publik.

Namun demikian, model hybrid masih mengalami kesulitan dalam mengenali kelas netral akibat ketidakseimbangan jumlah data antarkelas, yang ditunjukkan oleh nilai *precision*, *recall*, dan *F1-score* sebesar 0% pada kelas tersebut, sehingga nilai *accuracy* yang tinggi belum sepenuhnya menggambarkan kemampuan model dalam mengenali seluruh kelas sentimen secara seimbang. Berdasarkan temuan tersebut, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data seperti *oversampling* serta memperkuat tahap praproses melalui normalisasi kata tidak baku dan penanganan konteks bahasa yang lebih kompleks, agar performa klasifikasi pada kelas minoritas dapat ditingkatkan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Sumatera Utara, khususnya Fakultas Ilmu Komputer dan Teknologi Informasi Program Studi Sistem Informasi, dosen pembimbing, serta seluruh pihak yang telah memberikan fasilitas, bimbingan, masukan, dan dukungan selama penelitian hingga artikel ini selesai.

DAFTAR PUSTAKA

- Aisah, I. S., Irawan, B., & Suprpti, T. (2023). Algoritma Support Vector Machine (SVM) untuk analisis sentimen ulasan aplikasi Al-Qur'an digital. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3759–3765. <https://doi.org/10.36040/jati.v7i6.8263>
- Andriani, F., & Hasibuan, W. R. (2026). Analisis sentimen program makan bergizi gratis pada media sosial X menggunakan metode Naïve Bayes. *Indonesian Journal of Multidisciplinary Scientific Studies (IJOMSS)*, 4(5), 328–335.
- Clara, S., Prianto, D. L., Al Habsi, R., Lumbantobing, E. F., & Chamidah, N. (2021). Implementasi seleksi fitur pada algoritma klasifikasi *machine learning* untuk prediksi penghasilan pada *Adult Income Dataset*. *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2(1), 741–747.
- Fadli, M. N., Damanik, I. S., & Irawan, E. (2021). Penerapan metode Naive Bayes dalam menentukan tingkat kenyamanan pada rumah sakit terhadap pasien. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(3), 117–122. <https://doi.org/10.30865/klik.v2i3.297>
- Fiandaru, D. F., & Sipayung, Y. R. (2025). Penerapan metode klasifikasi Naive Bayes untuk analisis sentimen terhadap Undang-Undang ITE di media sosial Twitter. *ELSE (Elementary School Education Journal)*, 9(2), 264–[halaman akhir]. <https://doi.org/10.30651/else.v9i2.24343>

- Laia, M. M., Hasan, F. N., & Kuntoro, A. Y. (2025). Analisis sentimen program makan gratis pada platform X menggunakan algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 13(2), 258–264. <https://doi.org/10.33884/jif.v13i02.10427>
- Martantoh, E., & Yanih, N. (2022). Implementasi metode Naïve Bayes untuk klasifikasi karakteristik kepribadian siswa di sekolah MTs Darussa'adah menggunakan PHP MySQL. *Jurnal Teknologi Sistem Informasi*, 3(2), 166–175. <https://doi.org/10.35957/jtsi.v3i2.2896>
- Murzal, M., & Ridwan, R. (2021). Pembelajaran daring pada masa pandemi COVID-19: Telaah terhadap implementasi daring di MI Fathul Akbar NW Tempos Desa Banyu Urip. *El Midad*, 13(2), 86–100. <https://doi.org/10.20414/elmidad.v13i2.4338>
- Nurulita, A., Hakim, A., & Nurhikmah, H. (2021). Pelaksanaan pembelajaran daring di masa COVID-19 di SMP Negeri 3 Marioriwawo. *Journal of Educational Technology, Curriculum, Learning, and Communication*, 1(3), 132–137. <https://doi.org/10.26858/jetclc.v1i3.22411>
- Pinem, T. H., & Putra, Z. P. (2025). Evaluasi kinerja algoritma klasifikasi *deep learning* dalam prediksi diabetes. *Jurnal Ilmiah FIFO*, 17(1), 17–28. <https://doi.org/10.22441/fifo.2025.v17i1.003>
- Syahri, R., & Puspita, D. (2025). Classification of outstanding students using Support Vector Machine (SVM) based on data mining. *JITE (Journal of Informatics and Telecommunication Engineering)*, 9(1). <https://doi.org/10.31289/jite.v9i1.13191>
- Zusrotun, O. P., Murti, A. C., & Fiati, R. (2022). Analisis sentimen terhadap belajar online pada media sosial Twitter menggunakan algoritma Naive Bayes. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 11(3), 310–319. <https://doi.org/10.23887/janapati.v11i3.49160>