ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Perbandingan Akurasi SVM dan Naive Bayes dalam Klasifikasi Opini Isu Internasional

[1]Karina Dinda Artanti

[1]Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur
[1]22082010221@student.upnjatim.ac.id

Abstrak*— Perkembangan media sosial seperti YouTube telah menjadi ruang publik digital bagi masyarakat Indonesia dalam merespons berbagai kebijakan, termasuk isu strategis internasional. Volume opini publik yang besar dan tidak terstruktur membutuhkan pendekatan komputasional berupa analisis sentimen agar dapat diekstrak menjadi informasi yang bermanfaat. Penelitian ini bertujuan membandingkan performa tingkat akurasi antara algoritma Support Vector Machine (SVM) dan Naive Bayes dalam mengklasifikasikan sentimen opini publik. Sebanyak 1.555 komentar dikumpulkan melalui YouTube Data API v3 dan diperoleh 1.541 data valid setelah prapemrosesan, termasuk tahap stemming menggunakan pustaka Sastrawi, kemudian direpresentasikan menjadi vektor numerik menggunakan pembobotan TF-IDF. Pembagian data dilakukan secara acak dan stratified pada rasio 80:20. Untuk memastikan hasil tidak bergantung pada satu pembagian data tertentu, pengujian diperkuat dengan Stratified K-Fold Cross-Validation (k=5 dan k=10), serta pengujian pada kondisi data seimbang menggunakan teknik SMOTE. Hasil evaluasi menunjukkan SVM berbasis kernel linear mengungguli Naive Bayes secara signifikan dengan rata-rata akurasi validasi silang mencapai 91,82% (F1-macro 0,85), sedangkan Naive Bayes hanya memperoleh 71,77% (F1-macro 0,38). Standar deviasi antar-fold yang kecil (di bawah 4%) membuktikan kestabilan model SVM. Penerapan SMOTE meningkatkan performa SVM menjadi 92,73%, sementara pada Naive Bayes justru menyingkap bahwa akurasi awalnya bersifat semu akibat kegagalan mendeteksi kelas minoritas. Hasil komparasi ini menyimpulkan bahwa SVM merupakan algoritma yang lebih optimal dan tangguh untuk diimplementasikan pada sistem pemantauan kecenderungan opini masyarakat di media sosial.

Kata Kunci— **Analisis Sentimen, *YouTube*, *Support Vector Machine*, *Naive Bayes*, TF-IDF.**

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi di era digital telah mengubah cara masyarakat dalam menyuarakan opini dan aspirasinya. Media sosial kini tidak hanya berfungsi sebagai sarana interaksi sosial, tetapi juga telah bertransformasi menjadi ruang publik digital untuk merespons berbagai kebijakan pemerintah, termasuk dalam merespons isu-isu strategis berskala internasional [1]. Salah satu platform yang paling masif digunakan oleh masyarakat Indonesia adalah YouTube. Melalui kolom komentar YouTube, masyarakat kerap memberikan tanggapan yang beragam mulai dari dukungan, kritik, hingga sikap netral terhadap pernyataan resmi pemerintah, seperti isu keterlibatan atau rencana keluarnya Indonesia dari forum Board of Peace [2].

Banyaknya volume komentar yang terus bertambah setiap detiknya menghasilkan data teks yang sangat besar dan tidak terstruktur. Pemantauan opini publik secara manual tentu tidak lagi efektif dan efisien. Oleh karena itu, diperlukan sebuah pendekatan komputasional melalui bidang Natural Language Processing (NLP), khususnya teknik Analisis Sentimen (Analisis Opini), untuk mengotomatisasi proses ekstraksi informasi dan pemetaan sentimen masyarakat secara cepat dan akurat [3]. Analisis sentimen memungkinkan pemerintah untuk memahami kecenderungan pandangan publik sebagai bentuk evaluasi dalam mewujudkan tata kelola pemerintahan yang cerdas (smart governance).

Dalam melakukan klasifikasi sentimen, pendekatan Machine Learning telah terbukti memberikan tingkat akurasi yang tinggi. Terdapat berbagai algoritma yang umum digunakan, dua di antaranya adalah Support Vector Machine (SVM) dan Naive Bayes. Algoritma SVM dikenal sangat tangguh dalam menangani data teks berdimensi tinggi (high-dimensional space) dan mampu menemukan hyperplane optimal untuk memisahkan kelas data [4]. Di sisi lain, Naive Bayes sering menjadi pilihan populer karena efisiensi komputasinya yang tinggi dan pendekatannya yang berbasis probabilitas teorema Bayes, meskipun memiliki asumsi independensi antar fitur yang kuat [5].

Beberapa penelitian terdahulu telah mengaplikasikan kedua metode tersebut pada berbagai kasus klasifikasi teks, namun performa dari masing-masing algoritma sangat bergantung pada karakteristik dataset yang digunakan, terutama pada teks tidak baku khas komentar media sosial masyarakat Indonesia. Selain itu, evaluasi yang hanya bertumpu pada satu kali pembagian data dan mengabaikan kondisi ketidakseimbangan kelas berpotensi menghasilkan kesimpulan yang kurang representatif. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengkomparasikan tingkat akurasi antara algoritma SVM dan Naive Bayes dalam mengklasifikasikan sentimen masyarakat pada kolom komentar YouTube terkait isu internasional, dengan pengujian yang diperkuat melalui validasi silang dan skenario data seimbang. Hasil komparasi ini diharapkan dapat memberikan rekomendasi algoritma terbaik untuk diimplementasikan pada sistem cerdas pemantauan opini publik.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

II. LANDASAN TEORI

A. *Text Mining*

Text mining, atau penambangan teks, merupakan proses ekstraksi pola berupa informasi dan pengetahuan yang bermanfaat dari sekumpulan dokumen teks yang tidak terstruktur [6]. Berbeda dengan *data mining* tradisional yang umumnya mengolah data numerik terstruktur dalam basis data, *text mining* berhadapan dengan kompleksitas bahasa alami manusia. Tujuan utama dari *text mining* adalah untuk mengubah data teks yang masif menjadi format yang lebih terstruktur agar dapat dianalisis lebih lanjut menggunakan berbagai algoritma pembelajaran mesin (*machine learning*). Salah satu penerapan paling populer dari *text mining* dalam menganalisis opini publik di media sosial adalah analisis sentimen [7].

B. *Analisis Sentimen*

Analisis sentimen, atau *opinion mining*, merupakan salah satu cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengekstrak, mengidentifikasi, dan mengklasifikasikan opini atau teks ke dalam kategori polaritas tertentu, seperti positif, negatif, maupun netral [8]. Dalam konteks media sosial seperti YouTube, analisis sentimen berperan penting untuk mengubah data teks yang tidak terstruktur menjadi informasi kuantitatif yang dapat digunakan untuk mengukur respons masyarakat terhadap suatu fenomena atau kebijakan [9]. Tahapan utama dalam analisis sentimen meliputi pengumpulan data, prapemrosesan teks (*preprocessing*), ekstraksi fitur, dan klasifikasi menggunakan model *machine learning*.

C. *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang bekerja dengan mencari *hyperplane* atau bidang pemisah terbaik yang memaksimalkan *margin* antara dua atau lebih kelas data dalam ruang berdimensi tinggi [10]. Pada kasus klasifikasi teks, kernel linear sering digunakan karena kemampuannya memisahkan data teks yang direpresentasikan secara *sparse* (seperti hasil ekstraksi TF-IDF) secara efisien. Persamaan umum fungsi klasifikasi linier pada SVM dirumuskan sebagai berikut:

$$f(x) = sign(w.x + b) \quad \textbf{(1)}$$

di mana w merepresentasikan vektor bobot, x adalah vektor fitur dari data masukan, dan b merupakan nilai bias. SVM sangat efektif dalam menangani *dataset* dengan fitur yang sangat banyak, menjadikannya pilihan utama dalam klasifikasi teks bahasa alami [11].

D. *Naive Bayes*

Algoritma *Naive Bayes* merupakan metode klasifikasi probabilistik sederhana yang didasarkan pada Teorema Bayes dengan asumsi independensi atau ketidakterikatan antar fitur (kata) dalam sebuah teks [12]. Meskipun asumsi ini sering kali tidak berlaku sempurna pada struktur kalimat manusia, *Naive Bayes* tetap menunjukkan komputasi yang cepat dan performa yang kompetitif pada berbagai tugas klasifikasi sentimen dasar. Persamaan Teorema Bayes dalam klasifikasi didefinisikan sebagai:

$$P(c|X) = \frac{P(X|c)P(c)}{P(X)} \quad \textbf{(2)}$$

Dalam persamaan tersebut, $P(c|X)$ adalah probabilitas *posterior* dari kelas c berdasarkan *predictor* X, $P(c)$ adalah probabilitas *prior* kelas, $P(X|c)$ adalah kemungkinan prediktor dari kelas yang diberikan, dan $P(X)$ adalah probabilitas *prior* prediktor [13].

E. *TF-IDF*

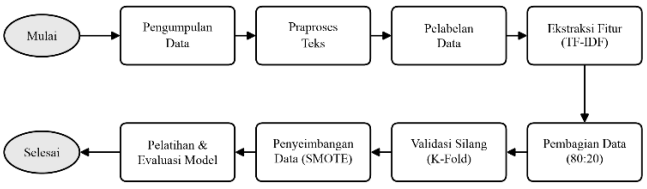
Sebelum data teks dapat diklasifikasikan oleh algoritma *machine learning*, teks tersebut harus dikonversi menjadi representasi numerik. Pembobotan TF-IDF digunakan untuk mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen terhadap seluruh korpus [14]. Nilai TF-IDF meningkat sebanding dengan frekuensi kemunculan kata dalam sebuah dokumen (TF), namun diimbangi oleh frekuensi kemunculan kata tersebut di seluruh korpus (IDF), sehingga kata hubung yang sering muncul tidak mendominasi bobot fitur. Pembobotan ini dirumuskan sebagai:

$$Wi,j = tfi,j \times log(\frac{N}{dfi}) \quad \textbf{(3)}$$

di mana Wi,j adalah bobot kata i pada dokumen j, tfi,j adalah frekuensi kata i dalam dokumen j, N adalah total dokumen, dan dfi adalah jumlah dokumen yang mengandung kata i [15].

III. METODE PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan komputasional yang sistematis untuk mengubah data teks tidak terstruktur menjadi model klasifikasi sentimen. Secara garis besar, alur metode penelitian ini terdiri dari enam tahapan utama: pengumpulan data, prapemrosesan teks, pelabelan data, ekstraksi fitur, klasifikasi, dan evaluasi model.



Gbr 1. Metode Penelitian

A. *Pengumpulan Data*

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa opini publik (komentar) yang diambil dari platform YouTube. Pengambilan data dilakukan secara otomatis melalui YouTube Data API v3 menggunakan skrip bahasa pemrograman Python. Total data mentah yang berhasil dikumpulkan berjumlah 1.555 komentar dari video yang membahas isu kebijakan luar negeri Indonesia, dan setelah

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

melalui tahap prapemrosesan serta penyaringan komentar kosong diperoleh 1.541 data valid yang siap diolah.

B. *Praproses Teks*

Teks komentar dari media sosial umumnya mengandung banyak karakter *noise* dan kata tidak baku yang dapat menurunkan performa algoritma. Oleh karena itu, dilakukan prapemrosesan secara menyeluruh melalui lima tahapan berikut:

1. *Cleaning*: Membersihkan teks dari elemen yang tidak relevan dengan analisis, seperti URL/tautan, sebutan pengguna (*mention*), karakter angka, dan tanda baca.
2. *Case Folding*: Menyeragamkan seluruh karakter huruf di dalam teks menjadi huruf kecil (*lowercase*).
3. *Tokenizing*: Memecah kalimat utuh menjadi potongan kata tunggal (*token*) menggunakan pustaka *Natural Language Toolkit* (NLTK).
4. *Stopword Removal*: Menghilangkan kata sambung atau kata umum yang dianggap tidak memiliki bobot sentimen. Pada tahap ini, dilakukan pengecualian terhadap kata-kata negasi (seperti "tidak", "bukan", "belum") agar konteks polaritas kalimat negatif tidak berubah.
5. *Stemming*: Mengembalikan kata yang memiliki imbuhan (awalan, sisipan, atau akhiran) menjadi bentuk kata dasarnya. Proses ini menggunakan pustaka *Sastrawi* yang didesain khusus untuk pemrosesan bahasa Indonesia.

C. *Pelabelan Data*

Pelabelan data pada penelitian ini dilakukan secara otomatis menggunakan pendekatan lexicon-based dengan kamus NRC Lexicon (National Research Council Emotion Lexicon) untuk mengkategorikan komentar menjadi kelas "Positif", "Negatif", atau "Netral". Pendekatan ini diterapkan pada teks hasil pembersihan sebelum proses stemming, agar kata kunci bersentimen tidak kehilangan bentuk aslinya, kemudian skor sentimen dari kamus tersebut dijumlahkan untuk menentukan label akhir setiap komentar.

D. *Ekstraksi Fitur*

Data teks yang telah bersih dan berlabel perlu direpresentasikan ke dalam bentuk vektor numerik agar dapat diproses secara matematis oleh algoritma klasifikasi. Pembobotan dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menilai tingkat kepentingan suatu kata terhadap sebuah dokumen dalam keseluruhan korpus. Proses pembobotan TF-IDF ditempatkan di dalam pipeline pemodelan sehingga proses fit hanya dilakukan pada data latih, guna menghindari kebocoran data (data leakage) ke data uji.

E. *Pembagian Data*

Dataset yang telah direpresentasikan dalam bentuk vektor numerik TF-IDF dipartisi menjadi dua himpunan bagian yang saling lepas, yaitu data latih dan data uji, dengan rasio pembagian standar sebesar 80:20. Pembagian dilakukan secara acak dengan penetapan random state tertentu agar hasil penelitian dapat direproduksi, serta menerapkan stratified split sehingga proporsi setiap kelas sentimen (Positif, Netral, dan Negatif) tetap terjaga sama antara data latih dan data uji. Penerapan stratifikasi ini penting mengingat distribusi kelas pada penelitian ini tidak seimbang. Hasil verifikasi menunjukkan proporsi kelas pada data latih (Netral 67,78%, Positif 20,94%, Negatif 11,28%) hampir identik dengan proporsi pada data uji (Netral 67,64%, Positif 21,04%, Negatif 11,33%). Pembagian ini bertujuan untuk menghindari masalah overfitting serta memastikan model dapat diuji menggunakan data yang belum pernah dipelajari sebelumnya.

F. *Validasi Silang (K-Fold Cross-Validation)*

Pengujian yang hanya bertumpu pada satu kali pembagian data berpotensi menghasilkan nilai akurasi yang dipengaruhi oleh keberuntungan komposisi data. Oleh karena itu, untuk memastikan performa model tidak bergantung pada satu pembagian data tertentu, dilakukan pengujian menggunakan Stratified K-Fold Cross-Validation dengan dua skenario, yaitu k=5 dan k=10. Stratifikasi pada setiap fold menjaga proporsi kelas tetap seimbang sebagaimana distribusi aslinya. Pada setiap iterasi, proses pembobotan TF-IDF di-fit hanya pada data latih di masing-masing fold untuk mencegah kebocoran data. Kinerja model dilaporkan dalam bentuk nilai rata-rata (mean) beserta standar deviasi (Std) antar-fold dari akurasi serta F1-macro yang disajikan pada kolom terpisah, sehingga kestabilan model dapat terukur secara objektif.

G. *Penanganan Ketidakseimbangan Data*

Distribusi kelas pada dataset bersifat tidak seimbang, dengan kelas Netral sebagai mayoritas dan kelas Negatif sebagai minoritas. Kondisi ini berpotensi menyebabkan model cenderung memprediksi kelas mayoritas sehingga metrik akurasi menjadi kurang representatif. Untuk mengukur performa metode pada kondisi data seimbang, diterapkan teknik oversampling SMOTE (Synthetic Minority Over-sampling Technique) [16]. SMOTE bekerja dengan membangkitkan sampel sintetis pada kelas minoritas berdasarkan interpolasi terhadap k tetangga terdekatnya. Dalam penelitian ini SMOTE diterapkan hanya pada data latih di dalam setiap fold validasi silang melalui mekanisme pipeline, sehingga data uji tetap merepresentasikan distribusi asli dan tidak terjadi kebocoran data.

H. *Pelatihan Model*

Pelatihan model dilakukan menggunakan dua algoritma klasifikasi, yaitu Naive Bayes dan Support Vector Machine (SVM) dengan kernel linear. Algoritma Naive Bayes dilatih untuk mengklasifikasikan sentimen melalui pendekatan probabilistik dengan menghitung peluang kemunculan kata pada tiap kategori kelas. Sementara itu, algoritma SVM dilatih untuk memisahkan kelas sentimen dengan mencari hyperplane optimal pada ruang data berdimensi tinggi yang dihasilkan dari fitur TF-IDF. Kedua model dilatih pada dua skenario, yaitu menggunakan data asli dan menggunakan data yang telah diseimbangkan dengan SMOTE, guna mengukur pengaruh ketidakseimbangan kelas terhadap performa masing-masing algoritma.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

I. *Evaluasi Model*

Kinerja dari kedua model yang telah dilatih kemudian diuji untuk memastikan keandalannya. Evaluasi performa dilakukan secara komprehensif menggunakan representasi Confusion Matrix untuk melihat perbandingan hasil prediksi yang benar dan salah pada masing-masing kelas. Metrik pengukuran utama yang digunakan meliputi tingkat akurasi (accuracy), presisi (precision), recall, dan f1-score. Nilai F1-macro turut dilaporkan sebagai metrik utama pendamping akurasi, karena metrik ini memberikan bobot yang setara pada setiap kelas sehingga lebih adil dalam menilai performa model pada data yang tidak seimbang.

IV. HASIL DAN PEMBAHASAN

A. Hasil Praproses Teks

Tahapan Praproses Teks yang meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming telah berhasil mereduksi karakter noise dan menyederhanakan variasi kata pada dataset komentar YouTube. Penggunaan pustaka Sastrawi terbukti efektif dalam memotong berbagai afiksasi (imbuhan) bahasa Indonesia menjadi kata dasar, sehingga beban komputasi fitur berkurang dan konteks kalimat menjadi lebih padat. Hasil transformasi data sebelum dan sesudah tahap praproses dapat dilihat pada Gbr. 2.

	Komentar_Asli	Komentar_Clean
0	Halah gilirin lu tau ternyata si USA gak sepow...	halah gilir lu tau si usa gak sepower full lu ...
1	Cepat keluar se belum rudal Iran datang	cepat belum rudal iran
2	Udah kek keluar bae, terus minta maaf sama Pal...	udah kek bae maaf palestina iran udah join org...
3	Tram sudah jadi saudara tua	tram saudara tua
4	Harus keluar... Gak ada manfaatnya wis ro r...	gak manfaat wis ro rakyat e mdadog angel kongk...
5	Pikirin bangsa sendiri dulu carut marut malah ...	pikirin bangsa carut marut sok sok an emang do...

Gbr 2. Hasil Praproses Teks

B. Distribusi Kelas Dataset

Dari 1.541 data valid yang diperoleh, distribusi kelas sentimen menunjukkan kondisi yang tidak seimbang (imbalanced) sebagaimana disajikan pada Tabel I. Kelas Netral mendominasi dataset dengan proporsi 67,75%, sedangkan kelas Negatif hanya berjumlah 174 data atau 11,29%. Kondisi ketidakseimbangan inilah yang menjadi dasar dilakukannya pengujian tambahan menggunakan teknik penyeimbangan data.

TABEL I
DISTRIBUSI KELAS DATASET

Kelas Sentimen	Jumlah Data	Proporsi(%)
Netral	1.044	67,75
Positif	323	20,96
Negatif	174	11,29
Total	1.541	100,00

C. Hasil Evaluasi Model pada Skema Holdout 80:20

Tahap Pengujian tahap pertama dilakukan menggunakan himpunan data uji sebesar 20% dari total dataset, yakni setara dengan 309 baris komentar YouTube, dengan pembagian acak dan stratified. Hasil komparasi evaluasi performa dari kedua algoritma direkapitulasi pada Tabel II. SVM memperoleh akurasi sebesar 94% dengan F1-macro 0,88, sedangkan Naive Bayes hanya memperoleh akurasi 72% dengan F1-macro 0,39.

TABEL II
HASIL EVALUASI KINERJA MODEL PADA SKEMA HOLDOUT 80:20 (STRATIFIED)

Algoritma	Kelas	Precision	Recall	F1	Akurasi
SVM	Negatif	0,91	0,60	0,72	94%
SVM	Netral	0,92	1,00	0,96	94%
SVM	Positif	0,98	0,91	0,94	94%
Naive Bayes	Negatif	0,00	0,00	0,00	72%
Naive Bayes	Netral	0,72	1,00	0,83	72%
Naive Bayes	Positif	0,82	0,22	0,34	72%

D. Hasil Validasi Silang (K-Fold Cross-Validation)

Untuk memastikan hasil pada skema holdout bukan merupakan artefak dari satu pembagian data tertentu, dilakukan pengujian menggunakan Stratified K-Fold Cross-Validation dengan skenario k=5 dan k=10. Hasil pengujian disajikan pada Tabel III, di mana kolom Mean menunjukkan nilai rata-rata antar-fold dan kolom ±Std menunjukkan standar deviasi antar-fold dari metrik yang bersangkutan. Performa SVM terbukti sangat stabil dengan standar deviasi yang kecil, yaitu sebesar 0,0143 pada k=5 dan 0,0184 pada k=10. Kestabilan ini membuktikan bahwa tingkat akurasi tinggi yang diperoleh SVM konsisten pada berbagai komposisi pembagian data. Sebaliknya, Naive Bayes memperoleh nilai F1-macro yang sangat rendah (0,38) meskipun akurasinya mencapai 71,77%, yang mengindikasikan adanya kegagalan model dalam mengenali kelas minoritas.

TABEL III
HASIL STRATIFIED K-FOLD CROSS-VALIDATION (DATA ASLI)

Model	K	Mean Akurasi	±Std	Mean F1-Macro	±Std
SVM (linear)	5	0,9137	0,0143	0,8447	0,0206
SVM (linear)	10	0,9182	0,0184	0,8500	0,0359
Naive Bayes	5	0,7158	0,0064	0,3775	0,0141
Naive Bayes	10	0,7177	0,0110	0,3816	0,0223

E. Hasil Pengujian dengan Data Seimbang (SMOTE)

Pengujian selanjutnya dilakukan untuk mengukur performa kedua metode pada kondisi data seimbang menggunakan teknik SMOTE. Perbandingan performa sebelum dan sesudah penerapan SMOTE pada skenario 10-fold disajikan pada Tabel IV, sedangkan rincian metrik per-kelas disajikan pada Tabel V.

TABEL IV
PERBANDINGAN PERFORMA SEBELUM DAN SESUDAH SMOTE (10-FOLD)

Model	Akurasi Asli	Akurasi SMOTE	F1-Macro Asli	F1-Score SMOTE
SVM (linear)	0,9182	0,9273	0,8500	0,8749
Naive Bayes	0,7177	0,5691	0,3816	0,5353

Penerapan SMOTE meningkatkan performa SVM pada seluruh metrik, dengan akurasi naik dari 91,82% menjadi 92,73% dan

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

F1-macro dari 0,8500 menjadi 0,8749. Peningkatan paling penting terjadi pada kemampuan mendeteksi kelas minoritas, di mana recall kelas Negatif meningkat dari 0,59 menjadi 0,70. Sebaliknya, Naive Bayes menunjukkan perilaku yang berbeda. Pada data asli, Naive Bayes gagal total mendeteksi kelas minoritas dengan nilai recall Negatif sebesar 0,00, sehingga akurasi 71,77% yang diperolehnya sesungguhnya hanya berasal dari kecenderungan model menebak kelas mayoritas. Setelah penerapan SMOTE, Naive Bayes mulai mampu mendeteksi kelas Negatif dengan recall 0,67 dan F1-macro meningkat dari 0,38 menjadi 0,54, namun akurasi keseluruhannya justru menurun menjadi 56,91% akibat banyaknya kesalahan klasifikasi pada kelas mayoritas. Temuan ini menegaskan bahwa akurasi tinggi Naive Bayes pada data tidak seimbang bersifat semu.

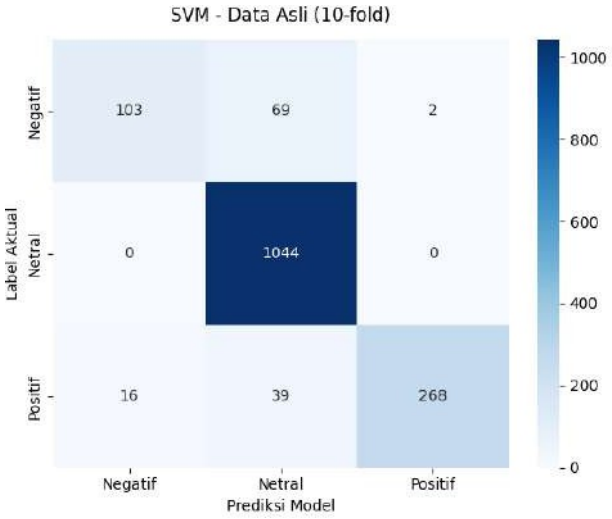
TABEL V
RINCIAN METRIK PER-KELAS (AGREGAT 10-FOLD)

Model	Skenario	Kelas	Precision	Recall	F1
SVM	Asli	Negatif	0,87	0,59	0,70
SVM	Asli	Netral	0,91	1,00	0,95
SVM	Asli	Positif	0,99	0,83	0,90
SVM	SMOTE	Negatif	0,83	0,70	0,76
SVM	SMOTE	Netral	0,93	0,99	0,96
SVM	SMOTE	Positif	0,98	0,85	0,91
Naive Bayes	Asli	Negatif	0,00	0,00	0,00
Naive Bayes	Asli	Netral	0,71	1,00	0,83
Naive Bayes	Asli	Positif	0,91	0,19	0,32
Naive Bayes	SMOTE	Negatif	0,26	0,67	0,38
Naive Bayes	SMOTE	Netral	0,94	0,47	0,62
Naive Bayes	SMOTE	Positif	0,47	0,85	0,60

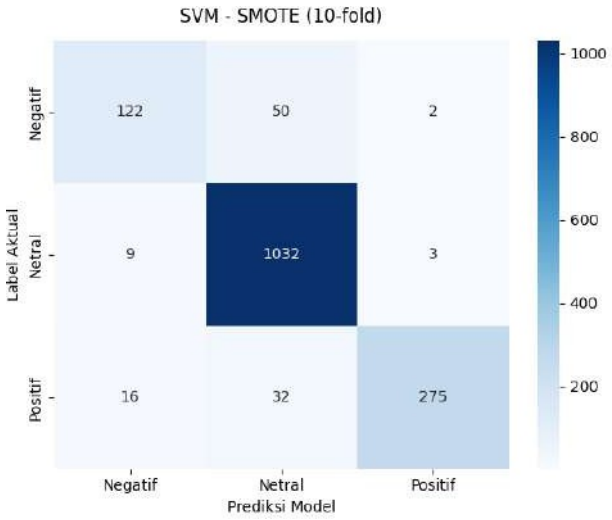
F. Analisis Perbandingan Performa

1) Analisis Performa SVM

Berdasarkan hasil pengujian, algoritma Support Vector Machine (SVM) berbasis kernel linear menunjukkan kinerja yang superior dan konsisten. Pada skema holdout, SVM mencapai akurasi 94%, sedangkan pada validasi silang 10-fold memperoleh rata-rata akurasi 91,82% dengan standar deviasi yang kecil. Kestabilan model ini terbukti dari kemampuannya mengklasifikasikan ketiga kelas sentimen secara proporsional, dengan presisi 0,99 pada kelas Positif dan recall 1,00 pada kelas Netral. Keunggulan paling menonjol terlihat pada kemampuannya menangani kelas minoritas: meskipun jumlah data kelas Negatif sangat terbatas, SVM tetap mampu menghasilkan presisi 0,87 dan recall 0,59 pada data asli, yang kemudian meningkat menjadi recall 0,70 setelah penerapan SMOTE. Kestabilan ini membuktikan bahwa SVM mampu membentuk hyperplane pemisah yang optimal antar kelas data berdimensi tinggi. Rincian distribusi prediksi model SVM dapat dilihat pada Gbr. 3 dan Gbr. 4.



Gbr 3. Confusion Matrix SVM – Data Asli (10-Fold)

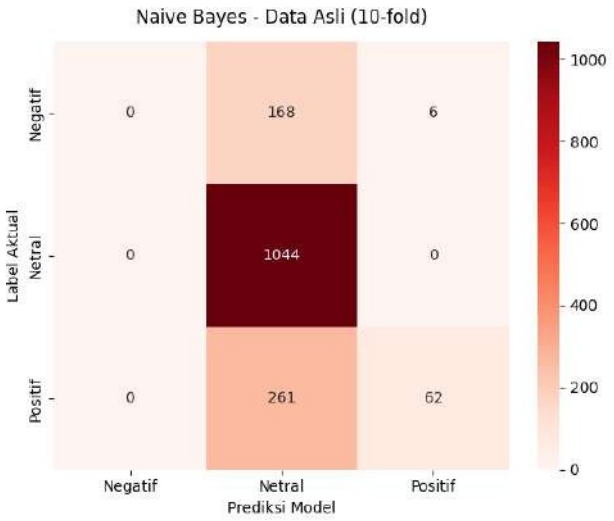


Gbr. 4 Confusion Matrix SVM – SMOTE (10-Fold)

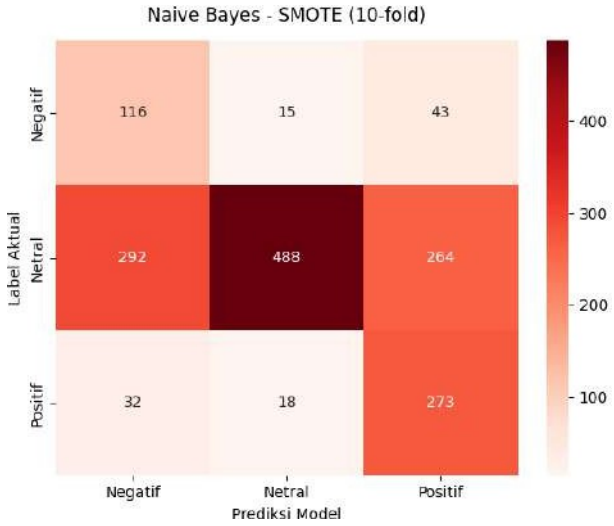
2) Analisis Performa Naive Bayes

Berbanding terbalik dengan SVM, algoritma Naive Bayes menghasilkan tingkat akurasi yang jauh lebih rendah, yakni 72% pada skema holdout dan 71,77% pada validasi silang 10-fold. Laporan klasifikasi menunjukkan adanya bias prediksi yang sangat kuat menuju kelas mayoritas (Netral), di mana nilai recall untuk sentimen Netral mencapai 1,00 namun mengorbankan kelas lainnya. Pada kelas minoritas (sentimen Negatif), Naive Bayes mengalami kegagalan total yang ditunjukkan dengan nilai presisi, recall, dan f1-score yang menyentuh angka 0,00. Setelah penerapan SMOTE, model mulai mampu mengenali kelas Negatif dengan recall 0,67, tetapi akurasi keseluruhannya menurun menjadi 56,91% karena meningkatnya kesalahan klasifikasi pada kelas mayoritas. Rincian penumpukan prediksi Naive Bayes dapat diamati pada Gbr. 5 dan Gbr. 6.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026



Gbr. 5 Confusion Matrix Naive Bayes – Data Asli (10-Fold)



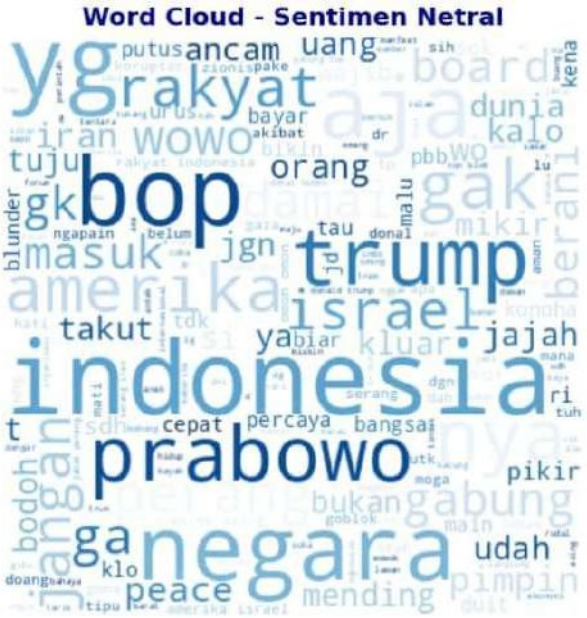
Gbr. 6 Confusion Matrix Naive Bayes – SMOTE (10-Fold)

Asumsi independensi kondisional pada Naive Bayes gagal merumuskan probabilitas kata bersentimen negatif akibat minimnya frekuensi data latih pada kelas tersebut. Akibatnya, model secara bias mengklasifikasikan komentar negatif ke dalam kelas mayoritas. Penerapan SMOTE memang memperbaiki sensitivitas Naive Bayes terhadap kelas minoritas, namun dengan konsekuensi penurunan akurasi yang signifikan, sehingga F1-macro-nya tetap jauh di bawah SVM. Di sisi lain, algoritma SVM terbukti sebagai model yang sangat tangguh (robust) karena mampu menemukan margin dan hyperplane pemisah yang optimal meskipun distribusi data antar kelas sentimen sangat timpang, dan merupakan satu-satunya algoritma yang performanya meningkat secara menyeluruh setelah penyeimbangan data. Oleh karena itu, SVM sangat direkomendasikan sebagai algoritma utama untuk implementasi sistem analisis sentimen opini publik yang kerap memiliki distribusi data tidak beraturan.

G. Visualisasi Hasil

Visualisasi hasil merupakan salah satu metode yang digunakan untuk menggambarkan data secara grafis guna memperoleh pemahaman yang lebih komprehensif terhadap pola dan karakteristik data sentimen yang dianalisis. Pada tahapan ini, eksplorasi data direpresentasikan melalui pemodelan Word Cloud yang menampilkan kata-kata paling sering muncul pada masing-masing kelas sentimen. Visualisasi ini memberikan gambaran kontekstual mengenai topik utama yang paling sering diperbincangkan oleh warganet berdasarkan polaritas opininya.

Gbr. 7 Visualisasi Word Cloud Kelas Positif



Gbr. 8 Visualisasi Word Cloud Kelas Netral

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Gbr. 9 Visualisasi Word Cloud Kelas Negatif

V. KESIMPULAN

Penelitian ini telah berhasil mengimplementasikan dan mengkomparasikan kinerja algoritma Support Vector Machine (SVM) dan Naive Bayes dalam menganalisis sentimen opini publik pada kolom komentar YouTube terkait isu internasional. Berdasarkan pengujian Stratified K-Fold Cross-Validation, algoritma SVM dengan kernel linear terbukti memiliki performa yang jauh lebih superior dengan rata-rata akurasi 91,82% dan F1-macro 0,85, sedangkan Naive Bayes hanya mampu mencapai akurasi 71,77% dengan F1-macro 0,38. Standar deviasi antar-fold yang kecil membuktikan bahwa performa SVM stabil dan tidak dipengaruhi oleh komposisi pembagian data tertentu.

Evaluasi metrik secara terperinci mengungkapkan bahwa Naive Bayes sangat rentan terhadap kondisi ketidakseimbangan kelas (imbalanced data), dibuktikan dengan ketidakmampuannya memprediksi sentimen kelas minoritas di mana nilai presisi dan recall-nya menyentuh angka nol. Pengujian pada kondisi data seimbang menggunakan SMOTE menunjukkan bahwa SVM semakin membaik dengan akurasi meningkat menjadi 92,73% dan F1-macro 0,87, sementara pada Naive Bayes justru menyingkap bahwa akurasi awalnya bersifat semu karena peningkatan sensitivitas terhadap kelas minoritas dibayar dengan penurunan akurasi menjadi 56,91%. Dengan demikian, SVM terbukti lebih unggul sekaligus lebih tangguh baik pada kondisi data asli maupun data seimbang, sehingga sangat direkomendasikan sebagai model utama dalam pengembangan sistem pemantauan opini publik. Untuk penelitian selanjutnya, disarankan untuk memperkaya leksikon kamus sentimen dengan variasi bahasa tidak baku dan kata negasi slang agar klasifikasi kelas netral dan negatif menjadi lebih representatif, serta mengeksplorasi teknik penyeimbangan data lainnya maupun model berbasis deep learning.

REFERENSI

[1] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375-384, 2024.

[2] R. A. Firsttama, A. A. Arifiyanti, and D. S. Y. Kartika, "Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 2, pp. 282-285, 2024.

[3] M. P. Firdaus and D. Trisnawarman, "Public Sentiment Analysis of the Public Housing Savings Program Using the IndoBERT Lite Model on YouTube Comments," *Jurnal Ilmu Komputer*, vol. 5, pp. 359-368, 2025.

[4] A. Widiyanti and D. A. Megawaty, "Perbandingan Algoritma K-Nearest Neighbor dan Support Vector Machine Pada Pengenalan Pola Tulisan Tangan," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1451, 2024.

[5] R. Kurniawan and S. Anwar, "Analisis Sentimen Penggemar Treasure Di Karnaval Mandiri Menggunakan Naïve Bayes," *Jurnal Informatika*, vol. 4, pp. 203-213, 2025.

[6] R. Ardiyansyah, et al., "Sentimen Komentar Youtube Dengan Sentiment Intensity Analyzer Dari NLTK," *Jurnal Teknik Informatika*, pp. 31-36, 2024.

[7] D. Fernando, et al., "Comparison of Machine Learning Algorithms in the Role of Hepatitis Patient Disease Classification," *Journal of Computer Science*, vol. 4, no. 2, pp. 159-170, 2021.

[8] L. Rao, "Sentiment Analysis of English Text with Multilevel Features," *Scientific Programming*, vol. 2022, 2022.

[9] A. H. K. S. Harahap, "Klasifikasi Sentimen Masyarakat Terhadap Revisi Undang-Undang Menggunakan Naïve Bayes Classifier," *Bulletin of Computer Science Research*, vol. 5, no. 4, pp. 594-603, 2025.

[10] J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analysis Of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naïve Bayes And K-Nearest Neighbor Methods," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 347-356, 2022.

[11] M. Nouval, "Analisis Sentimen Menggunakan Algoritma Support Vector Machine," *Jurnal Teknik Informatika*, vol. 4, no. 4, pp. 298-307, 2026.

[12] Z. Mesanda and M. Ikhsan, "Perbandingan Support Vector Machine dan Naïve Bayes Terkait Kepuasan Pengguna," *Journal of Computer System and Informatics*, vol. 5, no. 4, pp. 555-562, 2024.

[13] R. A. E. V. T. Sapanji, D. Hamdani, and P. Harahap, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, pp. 202-211, 2023.

[14] O. Salwa, D. Fadhillah, and J. H. Jaman, "Perbandingan Naive Bayes, Support Vector Machine, Logistic Regression dan Random Forest," *Jurnal Komputer*, vol. 13, no. 1, 2025.

[15] B. Amal, et al., "Penilaian Mahasiswa Terhadap Keefektifan Penggunaan Google Colab Dalam Pembelajaran Coding," *Kesatria: Jurnal Penerapan Sistem Informasi*, vol. 6, no. 1, pp. 66-76, 2025.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[17] S. Rohimah, S. Kenia, and P. Loka, "Analyzing Customer Sentiment Towards Marketplace Reviews Using Classification Algorithms," *Jurnal Sistem Informasi*, vol. 2, pp. 23-30, 2025.