Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 4 No. 3 (2026) pp: 1067-1075
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Komparasi Kinerja Metode Naïve Bayes Dan SVM Untuk Analisis Sentimen Opini Publik Di Youtube David Gadgetin Terhadap Iphone 17 Pro

Febrina Rahmadianti Putri[1], Miftahul Jannah[2], Restiayu Sekar Utami[3], Shintawati Khoirunnisa[4], Hanif Zaidan Sinaga[5]
[1,2,3,4,5] Program Studi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Ibn Khaldun Bogor,
[1]febrhdn@gmail.com, [2]miftahuljannahtibs@gmail.com, [3]restiayusekartmi@gmail.com, [4]its.shintakhoirunnisa@gmail.com, [5*]hanif.zaidan@gmail.com

Abstrak
Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam analisis sentimen opini publik pada komentar YouTube terkait penggunaan iPhone 17 Pro di kanal David Gadgetin. Penelitian menggunakan pendekatan Natural Language Processing (NLP) dengan memanfaatkan 1.382 komentar yang diperoleh melalui proses web scraping menggunakan YouTube Data API v3. Dataset diproses melalui tahapan text preprocessing yang meliputi case folding, normalization, tokenization, filtering (stopword removal), dan stemming, kemudian direpresentasikan menggunakan pembobotan Term Frequency–Inverse Document Frequency (TF-IDF). Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes dan SVM berbasis kernel linear dengan skema pembagian data 80:10:10, yaitu 80% data latih, 10% data validasi, dan 10% data uji. Kinerja kedua model dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, serta analisis confusion matrix. Hasil pengujian menunjukkan bahwa model SVM memperoleh akurasi sebesar 81,16% pada data validasi dan 84,17% pada data uji, sedangkan Naïve Bayes memperoleh akurasi sebesar 77i,54% pada data validasi dan 74,82% pada data uji. Pada kelas sentimen negatif, SVM memperoleh F1-score sebesar 0,86, sedangkan Naïve Bayes memperoleh 0,00. Berdasarkan hasil eksperimen pada dataset penelitian ini, SVM menunjukkan performa yang relatif lebih baik dibandingkan Naïve Bayes dalam klasifikasi sentimen komentar YouTube terhadap iPhone 17 Pro. Meskipun demikian, interpretasi terhadap hasil pada kelas minoritas tetap perlu dilakukan secara hati-hati karena jumlah data negatif relatif sedikit.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Support Vector Machine, Youtube, Iphone 17 Pro.

1. Pendahuluan

Perkembangan teknologi digital telah mengubah pola komunikasi dan pemasaran modern, di mana YouTube menjadi salah satu platform utama dalam penyebaran ulasan produk yang memengaruhi persepsi konsumen. Melalui konten kreator teknologi, opini publik terhadap produk seperti *smartphone* dapat terbentuk secara masif. Di sinilah *text mining* berperan penting sebagai alat yang kuat untuk menggali informasi berharga dari data teks yang tidak terstruktur (Prastyo et al., 2024).

Banyaknya opini yang dihasilkan pada kolom komentar YouTube membuat analisis manual menjadi kurang efektif, sehingga diperlukan pendekatan otomatis menggunakan *Natural Language Processing* (NLP). Salah satu implementasinya adalah analisis sentimen untuk mengklasifikasikan polaritas teks komentar ke dalam kategori positif, negatif, atau netral (Prastyo et al., 2024) Pendekatan ini tidak hanya membantu memberikan wawasan preferensi publik mengenai kualitas suatu produk, tetapi juga membantu perusahaan memahami kepuasan pengguna (Wijaya et al., 2025).

Beberapa penelitian terdahulu telah menerapkan metode ini pada platform YouTube. (Absharina & Nurqolbiah, 2025) menggunakan VADER dan TextBlob untuk menganalisis ulasan Starlink pada kanal Gadgetin dengan hasil dominan netral hingga positif. (Wijaya et al., 2025) menerapkan gabungan TextBlob dan *Support Vector Machine* (SVM) pada ulasan produk teknologi dengan akurasi 72%. Lebih lanjut, (Rahmawati & Aini, 2025) menggunakan algoritma *Long Short-Term Memory* (LSTM) dan memperoleh akurasi 90%, yang membuktikan ketangguhan metode *deep learning* dalam memahami bahasa informal media sosial. Di sisi lain, penggunaan algoritma SVM juga terbukti memiliki performa klasifikasi yang sangat baik meskipun dengan jumlah data yang relatif sedikit (Rahmawati & Aini, 2025)

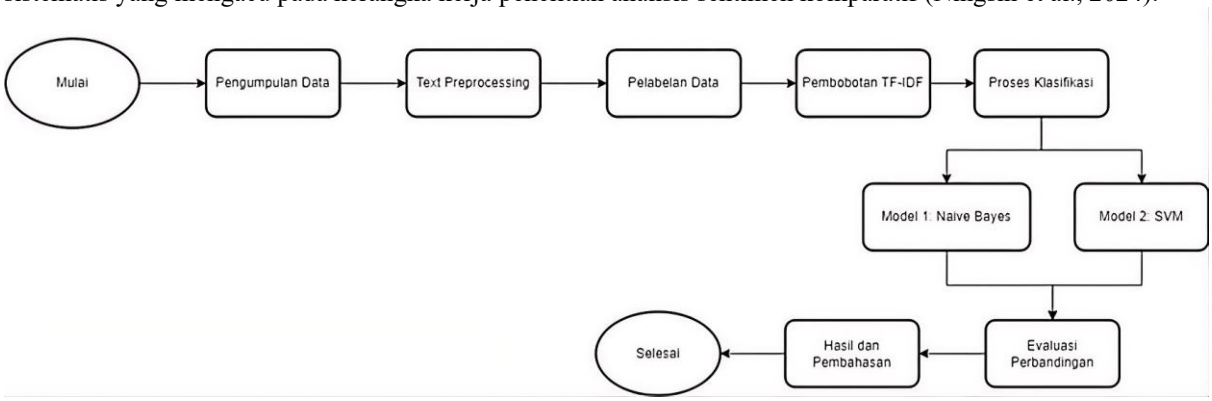
Meskipun kajian analisis sentimen pada YouTube sudah banyak dilakukan, sebagian besar studi terdahulu masih berfokus pada program edukasi, isu politik, produk kecantikan, atau produk teknologi secara umum (Clarisha, 2025) Penelitian yang secara khusus membedah respons publik terhadap penggunaan iPhone 17 Pro dalam konteks konten pemasaran digital di kanal David Gadgetin masih sangat terbatas. Sebagai *smartphone* premium yang

menarik perhatian tinggi, dinamika opini publik dan keluhan konsumen di media sosial terhadap perangkat ini berdampak langsung terhadap citra produk di pasaran (Ningsih et al., 2024). Kondisi ini menciptakan *research gap* karena belum tersedianya gambaran spesifik mengenai respons audiens terhadap penggunaan *smartphone* premium terbaru dalam aktivitas *digital marketing*.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube pada konten David Gadgetin mengenai penggunaan iPhone 17 Pro untuk pemasaran digital menggunakan NLP. Hasil dari penelitian ini diharapkan dapat memberikan informasi riil mengenai kecenderungan opini publik, sekaligus menjadi referensi strategis dalam membangun strategi pemasaran yang responsif dan terukur demi memperkuat posisi produk di pasar digital (Prastyo et al., 2024).

2. Metode Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode klasifikasi supervised learning untuk menganalisis sentimen masyarakat terhadap perilisan iPhone 17 Pro. Proses penelitian dilakukan melalui tahapan sistematis yang mengacu pada kerangka kerja penelitian analisis sentimen komparatif (Ningsih et al., 2024).



Gambar 1. Alur dan Skema Penelitian

2.1 Pengumpulan Data

Data penelitian berupa data primer yang bersumber dari komentar pengguna pada Youtube. Total data yang berhasil dihimpun adalah sebanyak 1.382 baris. Data ini dikumpulkan untuk merepresentasikan opini publik yang beragam mengenai topik penelitian.

2.2 Pelabelan Data (Labelling)

Proses pelabelan sentimen pada penelitian ini menggunakan metode *hybrid labeling*, yakni kombinasi antara pelabelan otomatis oleh AI dan verifikasi manual oleh manusia. Tahap awal dilakukan dengan memanfaatkan model AI untuk memberikan label awal (sentimen positif, negatif, atau netral) pada dataset. Selanjutnya, hasil pelabelan otomatis tersebut diperiksa kembali (*cross-check*) secara manual oleh 4 (empat) orang anotator untuk memastikan ketepatan klasifikasi.

Dalam menangani komentar yang bersifat ambigu atau mengandung sarkasme, anotator melakukan verifikasi mendalam dengan meninjau kembali konteks kalimat yang tidak tertangkap oleh model AI. Jika terdapat ketidaksesuaian antara label dari AI dengan hasil interpretasi manusia, maka anotator akan melakukan koreksi label secara langsung. Apabila terjadi perbedaan pendapat di antara para anotator saat proses verifikasi, penentuan label akhir dilakukan melalui mekanisme konsensus atau *majority voting* untuk mencapai validitas data yang objektif.

2.3 Pra-pemrosesan Data (Text Preprocessing)

Tahap ini dilakukan untuk menstandarisasi data teks agar memiliki format yang seragam. Merujuk pada prosedur (Clarisha, 2025) tahapan yang dilakukan meliputi:

- Case Folding: Mengubah semua huruf di dalam teks menjadi huruf kecil semua.
- Normalization: Mengubah kata-kata gaul, singkatan, atau *typo* menjadi kata baku.
- Tokenizing: Memotong-motong kalimat panjang menjadi potongan kata tunggal (token) sekaligus menghapus tanda baca dan angka.
- Filtering (Stopword): Membuang kata-kata umum yang sering muncul tapi tidak punya makna penting untuk analisis.
- Stemming: Mengubah kata berimbuhan menjadi kata dasarnya menggunakan bantuan *library* seperti Sastrawi.

2.4 Pembobotan TF-IDF

Pada tahap ini, komentar yang telah melalui proses text preprocessing dikonversi menjadi representasi vektor numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Implementasi ini

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

dilakukan melalui pustaka Scikit-Learn untuk memberikan bobot pada setiap kata agar fitur yang digunakan dalam klasifikasi memiliki signifikansi yang tepat. Proses pembobotan dilakukan melalui tiga tahapan matematis sebagai berikut:

Rumus menghitung TF :

$$TF\ (t,d) = \frac{Jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{Total\ jumlah\ kata\ dalam\ dokumen\ d}$$

Rumus menghitung IDF :

$$IDF\ (t)\ =\ log = \frac{Total\ jumlah\ dokumen\ dalam\ korpus}{Jumlah\ dokumen\ dalam\ korpus\ yang\ mengandung\ kata}$$

Rumus menghitung TF-IDF :

$$TF - IDF\ (t,d) = TF(t,d)\ x\ IDF\ (t)$$

Keterangan:

t = kata tertentu

d = dokumen tertentu

2.5. Klasifikasi Naïve Bayes

Model pertama yang digunakan adalah Naïve Bayes. Algoritma ini dipilih karena karakteristiknya yang berbasis pada teorema probabilitas Bayes, yang sangat efisien dalam menangani data teks berskala besar dengan dimensi fitur yang tinggi. Dalam penelitian ini, model Naïve Bayes berfungsi untuk memprediksi probabilitas kemunculan kata-kata dalam sebuah komentar terhadap kelas sentimen (positif, netral, atau negatif). Keunggulan algoritma ini terletak pada kecepatan komputasi yang tinggi dan kemampuannya dalam menghasilkan klasifikasi yang stabil meskipun dataset memiliki distribusi kata yang cukup kompleks.

2.6 Klasifikasi Support Vector Machine (SVM)

Model kedua yang digunakan sebagai pembanding adalah Support Vector Machine. Berbeda dengan MNB, SVM bekerja dengan pendekatan geometris, yaitu mencari hyperplane atau garis batas pemisah yang paling optimal antara kelas-kelas sentimen dalam ruang fitur berdimensi tinggi (Ningsih et al., 2024). SVM diimplementasikan untuk menentukan batas margin maksimum antara kelas positif, netral, dan negatif. Algoritma ini dipilih karena ketangguhannya dalam menangani data teks yang sparse (jarang) serta kemampuannya dalam memberikan akurasi klasifikasi yang lebih presisi pada dataset dengan fitur yang banyak.

2.7. Evaluasi Perbandingan

Setelah proses klasifikasi selesai, dilakukan evaluasi untuk membandingkan kinerja kedua model dengan membagi dataset ke dalam tiga bagian, yaitu 80% data latih (*training set*), 10% data validasi (*validation set*) untuk penyetelan parameter, serta 10% data uji (*testing set*) untuk pengujian performa akhir. Kinerja masing-masing model diukur menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score, yang kemudian dilengkapi dengan analisis *Confusion Matrix* untuk memvisualisasikan tingkat ketepatan prediksi pada setiap kelas sentimen. Seluruh hasil evaluasi tersebut menjadi acuan utama dalam menentukan algoritma mana yang paling unggul dan efektif untuk klasifikasi komentar iPhone 17 Pro.

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Proses pengumpulan data dalam penelitian ini dilakukan melalui teknik web scraping memanfaatkan YouTube Data API v3 untuk mengambil data opini publik secara terstruktur. Data komentar diambil secara menyeluruh dari kanal YouTube resmi David Gadgetin pada video spesifik yang berjudul "Review iPhone 17 Pro". Proses penarikan data (*scraping*) ini dilaksanakan pada 4 Juni 2026. Pendekatan pengumpulan data dalam penelitian ini dilakukan secara komprehensif dengan mengambil seluruh komponen ulasan, baik komentar utama (*top-level comments*) maupun komentar balasan (*replies/nested comments*) yang interaktif. Langkah ini diambil guna menangkap dinamika diskusi, interaksi antar-pengguna, serta konteks opini publik secara utuh dan mendalam. Dari proses penarikan awal pada tanggal tersebut, diperoleh total 1.413 baris data ulasan sebagai data komentar mentah. Setelah melalui serangkaian tahap seleksi data, pembersihan, serta eliminasi duplikasi atau teks kosong pada proses pembatasan masalah, ditetapkan jumlah sampel final sebanyak 1.382 data komentar. Dataset final inilah yang kemudian digunakan secara utuh sebagai basis data utama dalam tahap analisis dan pemodelan sentimen selanjutnya (Kamilah et al., 2025).

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

3.2 Pelabelan Data

Proses pelabelan data dalam penelitian ini menggunakan metode *hybrid labeling* yang mengombinasikan efisiensi model AI untuk label prediktif awal dengan verifikasi manual oleh 4 (empat) orang *human annotator*. Pemeriksaan ulang (*cross-check*) secara manual ini krusial untuk memastikan ketepatan klasifikasi, terutama dalam memahami aspek pragmatik, sarkasme, konteks utuh, serta bahasa gaul (*slang*) yang sering kali tidak mampu ditangkap secara akurat oleh pendekatan otomatis berbasis AI maupun kamus (Ningsih et al., 2024). Dalam menangani komentar ambigu atau sarkastik, anotator mengoreksi label secara langsung jika terdapat ketidaksesuaian dengan interpretasi manusia. Jika terjadi perbedaan pendapat, keputusan akhir diambil melalui mekanisme konsensus atau *majority voting* demi menjaga objektivitas dan validitas data hasil pelabelan (Ningsih et al., 2024).

Proses peninjauan ini mengklasifikasikan setiap komentar ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Seluruh kategori tersebut dipertahankan dan digunakan secara utuh dalam tahap analisis utama untuk memberikan gambaran distribusi opini publik yang komprehensif, serta untuk mengukur kemampuan algoritma klasifikasi (seperti Naïve Bayes) dalam membedakan polaritas sentimen yang beragam dari hasil pelabelan dasar ini.

Tabel 1. Contoh Komentar yang telah diberikan Label

Contoh Komentar	Label Data
Iphone gak pengen .desainnya jelek	Negatif
Alhamdulillah bisa beli iPhone 17	Positif
bang untuk konten mending Iphone 16 promax, atau 17 promax?	Netral

3.3 Hasil Preprocessing

Data mentah yang diperoleh dari hasil *crawling* masih bersifat tidak terstruktur dan mengandung banyak komponen acak (*noise*) seperti simbol, angka, singkatan, hingga penggunaan bahasa gaul (*slang*) (Dian Oktavianingsih et al., 2025). Oleh karena itu, tahapan *text preprocessing* mutlak dilakukan untuk membersihkan dan menyeimbangkan format data agar siap diolah secara optimal oleh algoritma *machine learning*. Tahapan ini diimplementasikan menggunakan Python pada platform Google Colab yang meliputi proses berurutan yaitu *case folding, normalization, tokenizing, filtering (stopword removal),* dan *stemming* (Kamilah et al., 2025).

Proses *preprocessing* diawali dengan tahap *case folding* untuk menyeragamkan seluruh huruf menjadi huruf kecil guna menghindari duplikasi makna akibat perbedaan kapitalisasi. Selanjutnya, dilakukan *normalization* untuk mengubah kata tidak baku atau bahasa *slang* menjadi bentuk standar agar sesuai dengan kaidah bahasa. Teks yang telah dinormalisasi kemudian masuk ke tahap *tokenization*, di mana kalimat dipecah menjadi potongan kata tunggal sekaligus mengeliminasi karakter yang tidak relevan. Berikutnya, kata-kata umum yang tidak membawa informasi penting dieliminasi melalui proses *filtering* atau *stopword removal*. Sebagai tahap akhir, diterapkan proses *stemming* untuk memotong imbuhan dan mengembalikan setiap kata ke dalam bentuk dasarnya (Kamilah et al., 2025).

Tabel 2. Contoh Komentar yang telah di Preprocessing

Tahapan	Hasil Transformasi Teks
Komentar Asli (*Raw Text*)	Terimakasih bang. Rencana bingung juga beli iphone 16 promax atau 17 promax. Akhirnya ambil keputusan 16 promax
1. *Case Folding*	terimakasih bang rencana bingung juga beli iphone promax atau promax akhirnya ambil keputusan promax
2. Normalization	terima kasih abang rencana bingung juga beli iphone promax atau promax akhirnya ambil keputusan promax
3. Tokenizing	['terima', 'kasih', 'abang', 'rencana', 'bingung', 'juga', 'beli', 'iphone', 'promax', 'atau', 'promax', 'akhirnya', 'ambil', 'keputusan', 'promax']
4. Filtering (*Stopword*)	['terima', 'kasih', 'abang', 'rencana', 'bingung', 'beli', 'iphone', 'promax', 'promax', 'ambil', 'keputusan', 'promax']
5. Stemming	['terima', 'kasih', 'abang', 'rencana', 'bingung', 'beli', 'iphone', 'promax', 'promax', 'ambil', 'putus', 'promax']

3.4 Hasil Perhitungan TF-IDF

TF-IDF memberikan bobot pada setiap istilah berdasarkan tingkat kemunculannya dalam sebuah dokumen maupun dalam keseluruhan kumpulan data. Teknik ini menilai tingkat kepentingan suatu kata dengan membandingkan frekuensi kemunculannya pada dokumen tertentu terhadap seluruh korpus, sehingga kata-kata yang memiliki relevansi tinggi akan lebih menonjol, sementara istilah umum dengan frekuensi tinggi akan memperoleh bobot yang lebih rendah. Berikut disajikan contoh hasil perhitungan TF-IDF pada salah satu komentar dalam dataset.

Tabel 3. Contoh Komentar yang telah diberikan Label

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

No.	Komentar (Hasil Preprocessing)
1.	beli, fitur, desain, oke
2.	banding, pro, layak, tahun

Tabel 4. Perhitungan TF-IDF

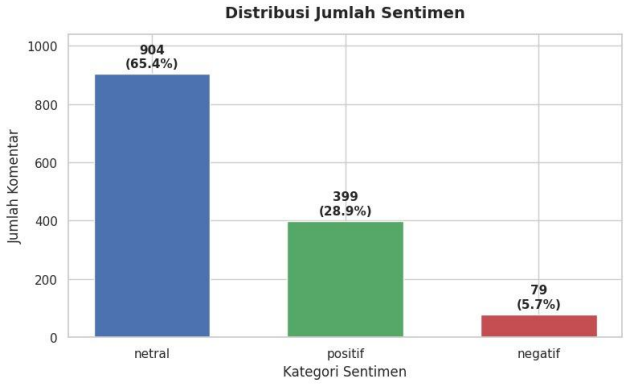
Kata	Kalimat 1 (TF)	Kalimat 2 (TF)	IDF	Kalimat 1 (TF-IDF)	Kalimat 2 (TF-IDF)
beli	1/4 = 0.25	0	log(2/1) = 0.3010	0.0753	0
fitur	1/4 = 0.25	0	log(2/1) = 0.3010	0.0753	0
desain	1/4 = 0.25	0	log(2/1) = 0.3010	0.0753	0
oke	1/4 = 0.25	0	log(2/1) = 0.3010	0.0753	0
banding	0	1/4 = 0.25	log(2/1) = 0.3010	0	0.0753
pro	0	1/4 = 0.25	log(2/1) = 0.3010	0	0.0753
layak	0	1/4 = 0.25	log(2/1) = 0.3010	0	0.0753
tahun	0	1/4 = 0.25	log(2/1) = 0.3010	0	0.0753

3.5 Analisis Hasil Model Klasifikasi Sentimen

Pada bagian ini dipaparkan analisis komprehensif mengenai hasil pengujian klasifikasi sentimen menggunakan dua algoritma *machine learning*, yaitu Naïve Bayes (MultinomialNB) dan Support Vector Machine (SVM). Proses eksperimen ini menerapkan skema pembagian data (*data splitting*) berlapis demi menjamin objektivitas model. Proporsi data dialokasikan sebesar 80% sebagai data latih (*training data*) untuk membangun kecerdasan model. Sisa data kemudian dibagi sama rata secara proporsional, yaitu 10% dialokasikan sebagai data validasi (*validation data*) untuk menyetel parameter awal model, dan 10% sisanya digunakan sebagai data uji (*testing data*) final untuk mengukur performa prediksi riil. Penggunaan model komparasi dengan skema ini sejalan dengan beberapa penelitian terdahulu dalam menguji ketangguhan algoritma di platform digital (Nurmadewi et al., 2026). Setelah melewati tahapan *preprocessing*, total dataset akhir sebanyak 1.382 baris ulasan terbagi secara presisi menjadi 1.105 data latih, 138 data validasi, dan 139 data uji (Pramudyantoro et al., 2025).

3.5.1 Distribusi Data Sentimen Dataset

Analisis awal dilakukan terhadap sebaran kelas sentimen pada keseluruhan korpus data untuk memahami karakteristik awal dataset. Hasil perhitungan menunjukkan bahwa dataset mengalami ketimpangan kelas (*imbalanced dataset*) yang cukup signifikan. Kelas sentimen netral mendominasi objek penelitian secara mutlak dengan total mencapai 904 komentar (65,4%), diikuti oleh sentimen positif sebanyak 399 komentar (28,9%), dan sentimen negatif sebagai kelas minoritas dengan representasi hanya sebesar 79 komentar (5,7%). Fenomena ketimpangan data seperti ini jamak ditemukan dalam analisis opini di media sosial, sehingga proses pembagian data sebelumnya wajib menggunakan parameter *stratify* agar persentase ketiga kelas tetap terjaga secara proporsional di setiap bagian data (Huwaida et al., 2024).



Gambar 2. Diagram Batang Distribusi Data Sentimen

3.5.2 Visualisasi Word Cloud Berdasarkan Kategori Sentimen

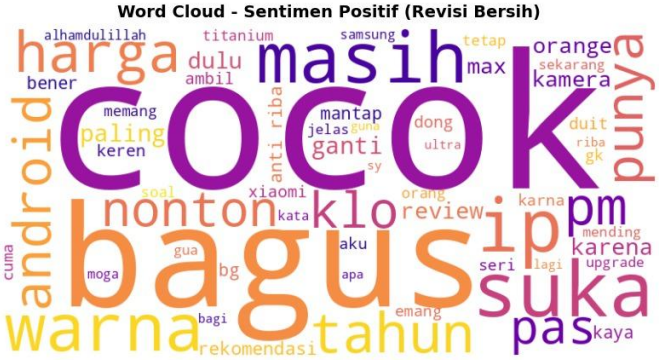
Ekstraksi kata kunci yang merepresentasikan topik utama pada masing-masing kategori sentimen divisualisasikan menggunakan fitur *Word Cloud*. Proses pembentukan awan kata ini sangat bergantung pada efektivitas tahapan pembersihan teks sebelumnya, terutama minimalisasi kata dasar lewat proses *stemming* agar kata-kata yang muncul lebih fokus dan bermakna (Albab et al., 2023). Ukuran huruf yang terbentuk dalam grafis tersebut berbanding lurus dengan tingkat frekuensi kemunculan kata di dalam dataset.

3.6 Word Cloud Sentimen Positif

Visualisasi Word Cloud pada sentimen positif digunakan sebagai langkah eksploratif untuk memetakan kata yang paling sering muncul secara intuitif, bukan sebagai dasar utama penentuan label sentimen. Berdasarkan visualisasi tersebut, kata-kata seperti "cocok", "bagus", "suka", "warna", dan "pas" tampil dengan ukuran yang relatif dominan. Kemunculan kata-kata ini mengindikasikan kecenderungan topik eksplorasi di mana audiens yang

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

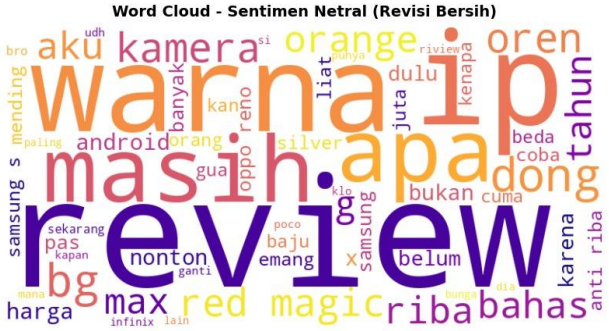
memberikan respons positif banyak menyoroti kesesuaian fitur (cocok/pas) serta kepuasan visual terhadap perangkat yang diulas, sekaligus memberikan gambaran awal mengenai preferensi utama penonton di kanal YouTube tersebut. (Tarigan, 2025).



Gambar 3. Word Cloud Sentimen Positif

3.7 Word Cloud Sentimen Netral

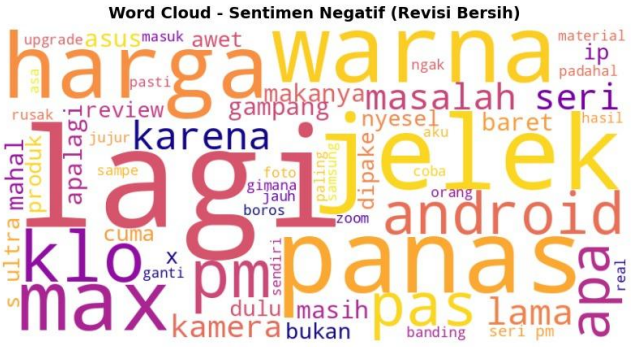
Dalam kluster sentimen netral, Word Cloud berfungsi untuk menangkap kecenderungan istilah umum yang kerap diutarakan publik guna memperkaya analisis deskriptif. Hasil visualisasi menunjukkan dominasi kata seperti "review", "warna", "ip", dan "apa", yang mengonfirmasi bahwa kelompok data ini bersifat eksploratif di sekitar pertanyaan, diskusi spesifikasi teknis, atau perbandingan objektif. Narasi visual ini mempertegas posisi kluster netral sebagai ruang diskusi logis penonton yang berfokus pada konten ulasan gadget, tanpa melibatkan bias emosional yang mengarah pada preferensi positif maupun negatif. (Nurpadhilah & Saputera, 2026).



Gambar 4. Word Cloud Sentimen Netral

3.8 Word Cloud Sentimen Negatif

Pada kategori sentimen negatif, Word Cloud dimanfaatkan sebagai instrumen pendukung untuk mengidentifikasi sinyal keluhan publik yang paling menonjol secara frekuensi. Kata-kata seperti "panas", "harga", "jelek", "warna", dan "lagi" muncul dengan ukuran yang cukup signifikan dalam visualisasi ini. Sesuai fungsinya sebagai visualisasi eksploratif awal, dominasi istilah-istilah tersebut memberikan petunjuk kuat bahwa topik keberatan atau kritik dari netizen sebagian besar berpusat pada isu suhu performa perangkat (panas), pertimbangan nilai ekonomis (harga), serta aspek peningkatan fitur yang dinilai belum memenuhi ekspektasi pengguna (Huwaida et al., 2024).



Gambar 5. Word Cloud Sentimen Negatif

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

3.8.1 Evaluasi Performa Model Klasifikasi Sentimen

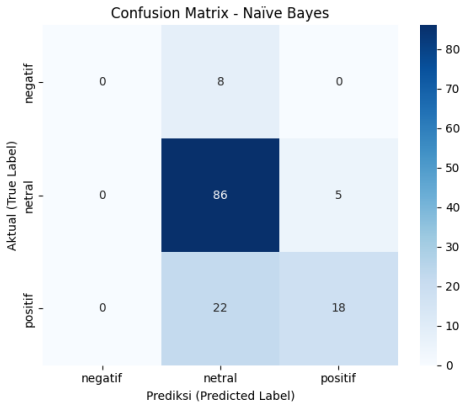
Berdasarkan hasil pengujian pada Tabel 5, model Support Vector Machine (SVM) memperoleh akurasi sebesar 81,16% pada tahap validasi dan 84,17% pada data uji final. Sementara itu, model Naïve Bayes memperoleh akurasi sebesar 77,54% pada data validasi dan 74,82% pada data uji akhir. Perbedaan performa kedua model terutama terlihat pada kemampuan dalam mengklasifikasikan kelas sentimen negatif. Pada penelitian ini, Naïve Bayes memperoleh nilai F1-score sebesar 0,00 pada kedua tahap pengujian, sedangkan SVM memperoleh F1-score sebesar 0,55 pada data validasi dan 0,86 pada data uji final. Berdasarkan hasil pengujian pada dataset penelitian ini, SVM menunjukkan performa yang relatif lebih baik dibandingkan Naïve Bayes, khususnya dalam mengklasifikasikan kelas sentimen negative.

Tabel 5. Akurasi Naïve Bayes &Support Vector Machine (SVM)

Tahap Pengujian	Algoritma	Kelas Sentimen	Precision	Recall	F1-Score	Akurasi Global
Data Validasi (10%)	Naïve Bayes	Negatif	0.00	0.00	0.00	77.54%
		Netral	0.75	0.98	0.85	
		Positif	0.90	0.47	0.62	
	SVM (Linear)	Negatif	1.00	0.38	0.55	81.16%
		Netral	0.80	0.94	0.87	
		Positif	0.83	0.60	0.70	
Data Uji Final (10%)	Naïve Bayes	Negatif	0.00	0.00	0.00	74.82%
		Netral	0.74	0.95	0.83	
		Positif	0.78	0.45	0.57	
	SVM (Linear)	Negatif	1.00	0.75	0.86	84.17%
		Netral	0.83	0.96	0.89	
		Positif	0.86	0.60	0.71	

3.8.2 Analisis Confusion Matrix Kedua Algoritma

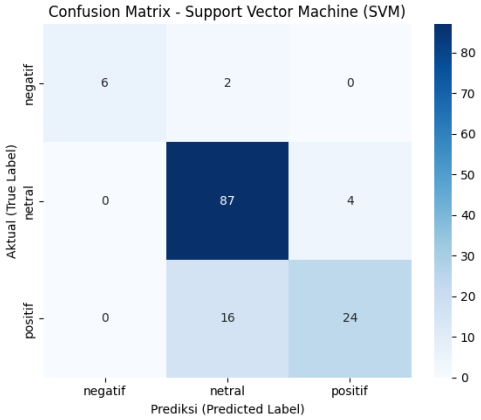
Analisis menggunakan instrumen *Confusion Matrix* dilakukan pada data uji final untuk mengidentifikasi pola kesalahan klasifikasi (*misclassification*) serta mengukur akurasi pemetaan riil dari masing-masing model. Pengujian ini memanfaatkan seluruh sampel data uji yang berjumlah 139 ulasan untuk membandingkan label aktual hasil anotasi manual dengan label prediksi yang dikeluarkan oleh sistem setelah fase pembobotan nilai TF-IDF (Honainah, 2026) Melalui visualisasi matriks ini, dapat diketahui secara detail pada kelas sentimen apa saja algoritma cenderung mengalami bias keputusan atau salah prediksi (Absharina & Nurqolbiah, 2025).



Gambar 6. Confusion matrix Naïve Bayes

Berdasarkan grafik pengujian pada model Naïve Bayes dengan total 139 data uji, ditemukan bahwa algoritma ini mengalami kegagalan lokal total dalam mengenali dokumen ulasan bersentimen negatif. Dari total 8 data uji aktual berkategori negatif, tidak ada satu pun data yang berhasil diprediksi dengan benar oleh Naïve Bayes karena seluruhnya (8 data) salah dikelompokkan ke dalam kelas netral, sehingga menghasilkan nilai *F1-score* sebesar 0,00. Selain itu, model ini juga salah mengklasifikasikan 22 data ulasan positif ke dalam kelas netral dan hanya berhasil menebak 18 data positif serta 86 data netral secara tepat. Kegagalan lokal Naïve Bayes pada kelas minoritas ini membuktikan keterbatasan teoritisnya yang mengasumsikan independensi antar variabel, sebuah kendala yang kerap ditemukan ketika diuji pada teks media sosial yang memiliki struktur kalimat tidak baku (Pramudyantoro et al., 2025).

DOI: https://doi.org/10.69693/ijmst.v4i2.11812



Gambar 7. Confusion Matrix SVM

Grafik *Confusion Matrix* pada model *Support Vector Machine* (SVM) memberikan gambaran mengenai visualisasi pemisahan kelas pada data uji. Berdasarkan hasil eksperimen pada dataset penelitian ini, SVM menunjukkan performa relatif lebih baik dibandingkan Naïve Bayes, terutama pada kelas negatif. Dari 8 data aktual pada kelas minoritas (sentimen negatif), model sukses memprediksi secara akurat sebanyak 6 data dan mengklasifikasikan 2 data ke dalam kelas netral. Namun, interpretasi terhadap hasil tersebut perlu dilakukan secara hati-hati karena jumlah data negatif pada data uji relatif kecil. Di sisi lain, pada kelas dengan volume lebih besar, sistem berhasil memprediksi 87 data netral dan 24 data positif secara presisi, dengan eror prediksi berupa 16 data positif yang meleset ke kelas netral dan 4 data netral yang meleset ke kelas positif. Hasil ini mengindikasikan karakteristik pendekatan geometris linear SVM dalam meminimalkan margin eror di ruang fitur dimensi tinggi untuk dataset ini (Nurmadewi et al., 2026).

3.8.3 Pembahasan Komparatif Hasil Klasifikasi Performa

Berdasarkan pengujian dengan skema pembagian data 80:10:10, yaitu 80% sebagai data latih (*training set*), 10% sebagai data validasi (*validation set*), dan 10% sebagai data uji (*testing set*), model Support Vector Machine (SVM) berbasis *kernel linear* memperoleh akurasi akhir sebesar 84,17%, sedangkan model Naïve Bayes memperoleh akurasi sebesar 74,82%. Penggunaan skema pembagian data tersebut bertujuan agar model dilatih menggunakan sebagian besar data, kemudian divalidasi sebelum dievaluasi pada data uji yang tidak pernah digunakan selama proses pelatihan. Pada dataset penelitian ini, hasil tersebut menunjukkan bahwa SVM memiliki performa klasifikasi yang relatif lebih baik dibandingkan Naïve Bayes. Temuan yang serupa juga terlihat pada tahap data validasi, di mana SVM memperoleh akurasi sebesar 81,16%, sedangkan Naïve Bayes memperoleh 77,54%.

Perbedaan performa kedua model pada penelitian ini dapat dipengaruhi oleh karakteristik masing-masing algoritma dalam menangani dataset yang memiliki ketimpangan kelas (*imbalanced dataset*) (Pramudyantoro et al., 2025). Pada dataset penelitian ini, SVM menunjukkan performa yang relatif lebih baik dibandingkan Naïve Bayes, terutama dalam mengklasifikasikan kelas sentimen negatif. Namun, interpretasi terhadap hasil tersebut perlu dilakukan secara hati-hati karena jumlah data negatif pada data uji relatif kecil. Hasil ini sejalan dengan karakteristik SVM yang membentuk *hyperplane* sebagai batas pemisah antar kelas pada ruang fitur berdimensi tinggi (Nurmadewi et al., 2026).

4. Kesimpulan

Berdasarkan seluruh rangkaian tahapan eksperimen yang telah dilakukan pada total dataset sebanyak 1.382 data dengan skema pembagian data terstratifikasi (80% training, 10% validation, dan 10% testing), dapat ditarik beberapa kesimpulan penting. Hasil pengujian komparatif pada dataset penelitian ini menunjukkan bahwa model Support Vector Machine (SVM) berbasis kernel linear menghasilkan performa relatif lebih baik dibandingkan algoritma Naïve Bayes, khususnya pada data uji yang memiliki ketimpangan kelas (imbalanced dataset). SVM mencatatkan tingkat akurasi akhir sebesar 84,17% dengan capaian F1-score sentimen negatif sebesar 0,86, sementara Naïve Bayes mencatatkan akurasi sebesar 74,82% dan mengalami keterbatasan dalam mengenali kelas negatif (F1-score 0,00) akibat sifat probabilistiknya yang cenderung sensitif terhadap kelas mayoritas. Meskipun demikian, interpretasi terhadap kelas negatif ini tetap perlu dilakukan secara hati-hati karena jumlah sampel minoritas dalam eksperimen ini relatif kecil. Secara keseluruhan, berdasarkan hasil pengujian pada penelitian ini, algoritma SVM linear menunjukkan performa yang relatif lebih baik dibandingkan Naïve Bayes dalam klasifikasi sentimen komentar YouTube terhadap iPhone 17 Pro.

DOI: https://doi.org/10.69693/ijmst.v4i2.11812

Reference

Absharina, E. D., & Nurqolbiah, F. (2025). Analisis Sentimen Komentar Video Youtube Review Starlink Indonesia Oleh Gadgetin Menggunakan VADER Dan Textblob. *The Indonesian Journal Of Computer Science*, *14*(3), 4383–4389. Https://Doi.Org/Https://Doi.Org/10.33022/Ijcs.V14i3.4686

Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization Of The Stemming Technique On Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, *20*(2), 1–12. Https://Doi.Org/10.26623/Transformatika.V20i2.5374

Clarisha, W. (2025). Analisis Sentimen Sunscreen Lokal Skintific , Somethinc , Dan Avoskin Dengan Naive Bayes Dan SVM Sentiment Analysis Of Local Sunscreen Skintific , Somethinc , And Avoskin With Naive Bayes And SVM. *Jambura Journal Of Electrical And Electronics Engineering*, *7*, 264–271.

Dian Oktavianingsih, A., Fadlil Fauzi, A., Immanuel, J., Dsimatupang, O., Amsury, F., & Fahlapi, R. (2025). Analisis Sentimen Komentar Youtube Terhadap Pengangkatan Menkeu Purbaya Menggunakan Algoritma SVM Dan Naïve Bayes. *Jurnal Cendekia Ilmiah*, *5*(1), 3702–3713.

Honainah. (2026). Penerapan Algoritma Naive Bayes Dengan Pembobotan TF-IDF Dan Pendekatan Lexicon Untuk Analisis Sentimen Opini Mahasiswa Terhadap Fasilitas Kampus. *Journal Scientific And Applied Informatics*, *09*(1), 71–80.

Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). Analisis Sentimen Komentar Youtube Terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes. *Jambura Journal Of Informatics*, *6*(1), 26–39. Https://Doi.Org/10.37905/Jji.V6i1.24718

Kamilah, F. I., Rochman, M. Z., & Rosyid, H. Al. (2025). ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP PERILISAN IPHONE 17 SERIES MENGGUNAKAN ALGORITMA NAIVE BAYES (Sentiment Analysis Of Youtube Comments On The Release Of The Iphone 17 Series Using The Naive Bayes Algorithm). *Journal Of Information System, Informatics And Computing*, *9*(2), 418–429. Https://Doi.Org/10.52362/Jisicom.V9i2.2198

Ningsih, W., Alfianda, B., & Wulandari, D. (2024). Comparison Of Naive Bayes And SVM Algorithms In Twitter Sentiment Analysis On Electric Car Use In Indonesia Perbandingan Algoritma SVM Dan Naïve Bayes Dalam Analisis Sentimen Twitter Pada Penggunaan Mobil Listrik Di Indonesia. *Indonesian Journal Of Machine Learning And Computer Science*, *4*(April), 556–562.

Nurmadewi, D., Jailani, Z. F., & Rafi, H. (2026). Comparison Of Support Vector Machine And Naïve Bayes For E-Commerce Review Sentiment Classification Perbandingan Support Vector Machine Dan Naïve Bayes Untuk Klasifikasi Sentimen Ulasan E-Commerce. *Indonesian Journal Of Machine Learning And Computer Science*, *6*(April), 923–933.

Nurpadhilah, N. A., & Saputera, S. A. (2026). Sentiment Analysis And Characteristics Of Youtube User Opinions Toward Samsung And Iphone Brands Using TF-IDF With Naive Bayes And KNN Comparison And Mcnemar. *Media Computer Science*, *5*(2), 987–998.

Pramudyantoro, A., Altiarika, E., Wahyuzi, Z., Pratama, Y. B., & Permana, T. D. (2025). Perbandingan Performa Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Komentar Youtube Terhadap Industri Esports Di Indonesia. *KAMPUS AKADEMIK PUBLISING Jurnal Ilmiah Nusantara (JINU)*, *2*(6), 1391–1399. Https://Doi.Org/10.61722/Jinu.V2i6.6753

Prastyo, D., Mursyidin, I. H., & Irawan, D. (2024). Klasifikasi Sentimen Komentar Youtube Dengan NLP Pada Debat Pilkada Banten. *Bit-Tech (Binary Digital - Technology)*, *7*(2). Https://Doi.Org/10.32877/Bt.V7i2.1833

Rahmawati, I., & Aini, N. (2025). Analisis Sentimen Komentar Youtube Pada Program Clash Of Champions Ruangguru Menggunakan Deep Learning Berbasis. *Indonesian Journal Of Learning And Technological Innovation*, *04*(01), 96–106.

Tarigan, D. A. (2025). Review Produk Iphone Dengan Analasis Sentimen Menggunakan Algoritma Text Mining TF-IDF. *Jurnal Sains Dan Teknologi Informasi*, *4*(2), 46–54. Https://Doi.Org/10.47065/Jussi.V4i2.7799

Wijaya, F., Ramadhani, R. A., & Setiawan, A. B. (2025). *Analisis Sentimen Komentar Video Youtube Dengan Leksikon Textblob Menggunakan Algoritma Svm Berdasarkan Rasio Data Uji*. *9*, 1894–1902.

DOI: https://doi.org/10.69693/ijmst.v4i2.11812