

# Sentiment Analysis of Netizen Opinions on TikTok Towards iPhone Using Naïve Bayes Algorithm and Support Vector Machine (SVM)

Sela Pebriana<sup>1</sup>, Castaka Agus Sugianto<sup>1</sup>

<sup>1</sup> Teknik Informatika, Politeknik TEDC Bandung, Kota Cimahi, 40513, Indonesia

## Informasi Artikel

Diterima : 24 Mei 2025  
Revisi : 7 Juli 2025  
Publikasi : 30 September 2025

## Kata Kunci:

Naïve Bayes  
Support Vector Machine  
Media Sosial  
Analisis Sentimen

## ABSTRAK

Penelitian ini bertujuan menganalisis sentimen pengguna TikTok terhadap produk iPhone dengan memanfaatkan komentar TikTok sebagai data primer. TikTok dipilih karena tingginya interaksi pengguna dan kemudahan akses terhadap opini publik secara spontan. Sebanyak 964 komentar dikumpulkan dan diproses melalui tahapan pembersihan data. Sentimen diklasifikasikan menjadi positif dan negatif menggunakan dua algoritma pembelajaran mesin populer: Naïve Bayes dan Support Vector Machine (SVM). Komparasi ini dilakukan untuk menilai efektivitas masing-masing algoritma dalam konteks data media sosial lokal yang bersifat singkat dan tidak terstruktur. Hasil menunjukkan akurasi Naïve Bayes sebesar 74% dan SVM sebesar 71%. Temuan ini menegaskan bahwa Naïve Bayes lebih unggul untuk analisis cepat opini publik berbasis teks pendek, serta berpotensi diterapkan dalam pemantauan persepsi konsumen dan strategi pemasaran digital secara efisien.

## ABSTRACT

This study aims to analyze TikTok users' sentiment toward the iPhone by utilizing TikTok comments as the primary data source. TikTok was chosen due to its high user engagement and ease of access to spontaneous public opinions. A total of 964 comments were collected and processed through a data cleaning stage. The sentiments were classified into positive and negative categories using two popular machine learning algorithms: Naïve Bayes and Support Vector Machine (SVM). This comparison was conducted to evaluate the effectiveness of each algorithm in handling local social media data, which is typically brief and unstructured. The results show that Naïve Bayes achieved an accuracy of 74%, while SVM reached 71%. These findings indicate that Naïve Bayes performs better in fast sentiment analysis of short-text public opinions and has practical potential for monitoring consumer perception and supporting efficient digital marketing strategies.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



## \*Penulis Koresponden

Email: [sela.pebriana20@gmail.com](mailto:sela.pebriana20@gmail.com)

Cara sitasi IEEE:

S. Pebriana, C.A. Sugianto, "Sentiment Analysis of Netizen Opinions on TikTok Towards iPhone Using Naïve Bayes Algorithm and Support Vector Machine (SVM)" *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 3, pp. 901–911, September 2025, doi: 10.30811/jaise.v5i3.7011

## 1. PENDAHULUAN

Produk iPhone menjadi salah satu topik yang ramai diperbincangkan di TikTok sepanjang tahun 2024. Data menunjukkan bahwa tingkat loyalitas pengguna iPhone mencapai 94%, lebih tinggi dibandingkan pengguna Android yang berada di angka 80% [1]. Tingginya loyalitas ini dipengaruhi oleh citra eksklusif yang melekat pada iPhone, terutama karena penggunaannya yang umum di kalangan selebritas, tokoh publik, dan konten kreator [2]. Fenomena ini menimbulkan berbagai opini masyarakat Indonesia mengenai kualitas, performa, dan nilai sosial dari produk iPhone.

TikTok dipilih sebagai objek penelitian karena memiliki tingkat interaksi yang tinggi serta kemudahan akses terhadap opini publik yang bersifat spontan dan organik. Komentar pengguna TikTok mencerminkan *respons* langsung terhadap suatu isu atau produk, sehingga menjadi sumber data primer yang relevan untuk analisis opini masyarakat secara *real-time*. Dengan memanfaatkan komentar pengguna, penelitian ini bertujuan untuk memahami lebih dalam persepsi publik terhadap iPhone melalui pendekatan analisis sentimen.

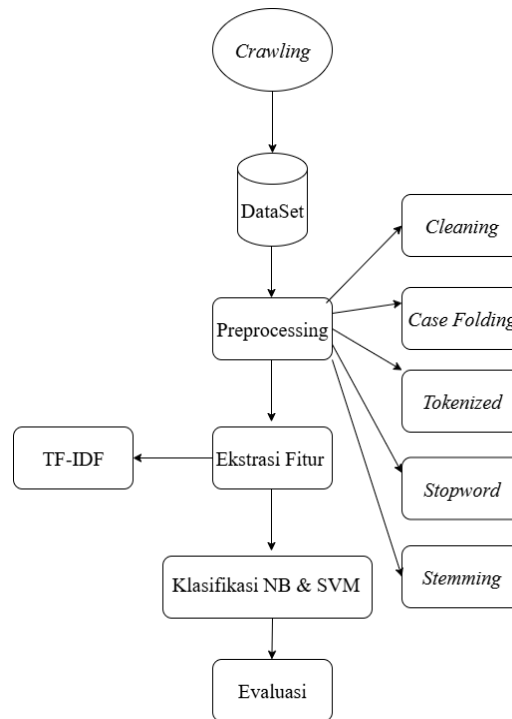
Analisis sentimen, atau *opinion mining*, merupakan metode untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif, negatif, atau netral [3]. Dalam penelitian ini, komentar pengguna TikTok dianalisis dan diklasifikasikan ke dalam dua kategori utama: sentimen positif dan sentimen negatif. Untuk mencapai klasifikasi yang akurat dan efisien, digunakan dua algoritma pembelajaran mesin populer, yakni *Naïve Bayes* dan *Support Vector Machine* (SVM). *Naïve Bayes* dikenal efektif dalam konteks data terbatas, memiliki kecepatan komputasi tinggi, serta mudah diterapkan [4][5]. Sementara itu, SVM dikenal handal dalam mengenali pola kompleks dan efisien dalam proses klasifikasi [6][7].

Penelitian sebelumnya telah membahas penerapan algoritma *Naïve Bayes* dalam menganalisis opini masyarakat, seperti pada studi oleh Feisy Kartika Ilmi, Syamsa Shahira Julyinda, Eva Mahdyta Kiswana dan M. Yoka Fathoni [8] Mengenai sentimen masyarakat terhadap kebijakan pemerintah Indonesia yang melarang penjualan iPhone 16 karena tidak memenuhi Persyaratan Tingkat Komponen Dalam Negeri (TKDN). Data yang diperoleh 164 data tweet dengan menghasilkan akurasi sebesar 71,82%, *precision* 74,90%, *recall* 71,82%, dan *F1-score* 71,47%. Adapun Penelitian lain menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* yang dilakukan oleh Friska Aditia Indriyani, Ahmad Fauzi, Sutan Faisal [9] Mengenai ulasan aplikasi TikTok pada Google Play Store. Kemudian 2000 data digunakan dalam proses klasifikasi data, 20% sebagai data uji dan 80% sebagai data latih. Kemudian pada tahap pengujian dari penelitian ini menggunakan *confusion matrix* dan akurasi yang didapatkan sebesar 79%, *precision* 86%, *recall* 8,5% dengan menggunakan metode *Naïve Bayes*. Lalu untuk metode SVM nilai akurasi yang didapatkan sebesar 84%, *precision* 77%, *recall* 73%. Jika dilihat dari nilai akurasi yang telah didapatkan dapat disimpulkan bahwa dalam penelitian ini pada metode SVM lebih baik daripada metode *Naïve Bayes*. Penelitian dengan menggunakan algoritma *Support Vector Machine* (SVM) yang dilakukan oleh Wasim Bourequat dan Hassan Mourad [10], penelitian ini bertujuan untuk menganalisis sentimen publik terhadap rilis iPhone menggunakan metode *Support Vector Machine* (SVM). Penelitian ini menggunakan total 1002 data, yang dibagi menjadi 801 data latih dan 201 data uji. Hasil evaluasi model menunjukkan bahwa SVM memberikan performa yang sangat baik dalam analisis sentimen, dengan nilai akurasi 89,21%, *precision* 92,43%, *recall* 95,53%, dan skor F1 sebesar 93,95%.

Namun demikian, masih terdapat *gap* penelitian dalam konteks pemanfaatan komentar TikTok sebagai data primer yang bersumber dari media sosial lokal Indonesia. Sebagian besar penelitian sebelumnya lebih banyak menggunakan data dari *platform* seperti Twitter atau Google Play Store, yang karakternya berbeda dalam hal struktur bahasa dan konteks budaya. Oleh karena itu, penelitian ini hadir untuk mengisi kekosongan tersebut dengan menganalisis komentar TikTok secara khusus, sekaligus membandingkan performa dua algoritma *Naïve Bayes* dan SVM dalam mengklasifikasikan sentimen terhadap iPhone berdasarkan karakteristik data media sosial yang tidak terstruktur.

## 2. METODE

Metode penelitian yang digunakan terdiri dari 6 langkah, yaitu *crawling*, *dataset*, *preprocessing*, ekstrasi fitur TF-IDF, klasifikasi *Naïve Bayes* & SVM dan evaluasi. Alur proses penelitian tersebut dapat ditunjukkan dalam bentuk visual pada gambar 1.



Gambar 1. Alur proses penelitian

Tahap pada metode *crawling* data merupakan teknik untuk mengumpulkan data dari situs web secara otomatis menggunakan perangkat lunak [11]. Data penelitian dikumpulkan dari tahun 2024 dengan menggunakan *APIfy*, peneliti dapat menggunakannya untuk proses *crawling* data *TikTok*. Hasil *crawling* data ditampilkan dalam gambar 2.

	id_akun	nama_akun	waktu \
0	7435111671259120439	aprilizya_1	2024-11-09T03:10:52.000Z
1	7434861843020137224	satriaagro	2024-11-08T11:01:15.000Z
2	7434829765243175736	radoo.02	2024-11-08T08:57:00.000Z
3	7438941918174626615	deadperson2024	2024-11-19T10:54:23.000Z
4	7434877127902233399	wammalokk	2024-11-08T12:00:39.000Z

	url_video \
0	<a href="https://www.tiktok.com/@anakkos_mlg/video/7434...">https://www.tiktok.com/@anakkos_mlg/video/7434...</a>
1	<a href="https://www.tiktok.com/@anakkos_mlg/video/7434...">https://www.tiktok.com/@anakkos_mlg/video/7434...</a>
2	<a href="https://www.tiktok.com/@anakkos_mlg/video/7434...">https://www.tiktok.com/@anakkos_mlg/video/7434...</a>
3	<a href="https://www.tiktok.com/@anakkos_mlg/video/7434...">https://www.tiktok.com/@anakkos_mlg/video/7434...</a>
4	<a href="https://www.tiktok.com/@anakkos_mlg/video/7434...">https://www.tiktok.com/@anakkos_mlg/video/7434...</a>

	komentar	sentimen
0	jangan mainin sambil ngecas	positif
1	susah banget cari charger buat ip	negatif
2	jangan beli yang 64 pokonya	negatif
3	beli please , layak	positif
4	pengen cameranya doang cuy	positif

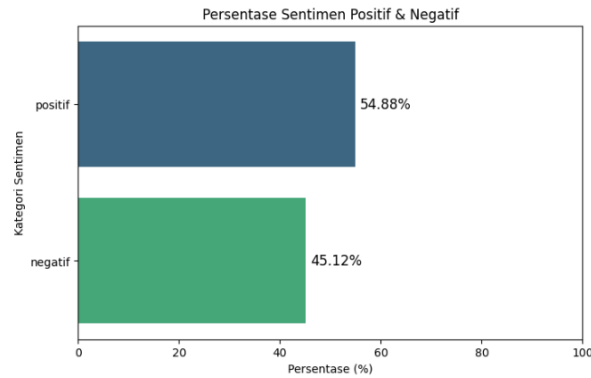
Gambar 1. *Crawling* data

Tahap Selanjutnya, dataset pada penelitian ini adalah kumpulan data atau informasi analisis sentimen *TikTok* terhadap iPhone yang didapat dari komentar konten tersebut. setelah itu, data dilakukan *labelling* secara manual dengan kategori sentimen positif dan negatif. Sebanyak 964 data yang disimpan didalam file csv. Kumpulan data ini dapat dilihat pada gambar 2.

Pada tahap *preprocessing* dilakukan pembersihan data untuk menghilangkan kesalahan dan faktor yang mengurangi kualitas data, termasuk penghapusan kata seperti *RT*, *URL*, *hashtag*, *mention*, *simbol*, serta duplikasi data data [12]. Hasil Tahapan ini dapat dilihat pada (Tabel 2-5).

*Ekstraksi fitur* adalah proses dalam data *mining*, *machine learning*, dan analisis data untuk mengubah data mentah menjadi representasi yang lebih informatif dan relevan guna meningkatkan kinerja model atau algoritma. Proses ini bertujuan untuk mengambil karakteristik penting atau atribut tertentu dari data, sehingga data tersebut dapat lebih mudah diproses oleh algoritma. Proses ini bertujuan untuk mengambil karakteristik penting atau atribut tertentu dari data, sehingga data tersebut dapat lebih mudah diproses oleh algoritma.

Klasifikasi data dilakukan menggunakan teknik *classification*, yang berfungsi untuk mengelompokkan data sesuai dengan kategori tertentu.



Gambar 2. *Persentase kelas positif dan negatif*

Pada gambar 3 diatas dapat dilihat, proses ini dilakukan dengan perhitungan probabilitas setiap kata yang telah dipisah supaya bisa mendapatkan hasil prediksi dari setiap kelas atau kategori dari dataset sentimen yang dimasukan, setiap data diprediksi berdasarkan 2 kelas, yaitu positif dan negatif. Setelah proses dilakukan maka dapat menghitung *accuracy*, *performance* dari kedua algoritma yang digunakan.

Tahap Terakhir, evaluasi model yang dilakukan dengan mengukur metrik akurasi, presisi, *recall*, dan *F1-score* guna memastikan kinerja yang optimal. Hasil akhir analisis ini diharapkan dapat memberikan wawasan bagi pengembang dalam meningkatkan kualitas produk iPhone.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. *Crawling*

Metode ini merupakan teknik untuk mengumpulkan data dari situs web secara otomatis menggunakan perangkat lunak [11]. Data yang digunakan terdiri dari 964 sampel yang dikumpulkan dari *platform* TikTok dalam bahasa Indonesia yang diambil menggunakan *Apify* dengan fitur *TikTok Comments Scraper*. Sampel ini dimanfaatkan untuk berbagai keperluan penelitian pada beberapa kategori. Gambar 4 berikut menampilkan hasil pengumpulan data.

#### 3.2. *Preprocessing*

Tahap *preprocessing* merupakan proses penting dalam pengolahan data yang bertujuan untuk membersihkan, menyiapkan, dan mengatur data mentah yang diperoleh dari *TikTok* umumnya masih belum teratur, agar siap digunakan dalam analisis lebih lanjut. Pada tahap ini, data dari file CSV dibaca dan diperiksa strukturnya untuk memastikan bahwa semua kolom dan isinya terbaca dengan benar. Proses ini memastikan bahwa data yang akan digunakan memiliki kualitas dan struktur yang baik, sehingga mendukung hasil analisis yang akurat dan efisien. *Preprocessing* mencakup beberapa tahapan, seperti *Cleaning*, *Tokenized*, *Stopword*, dan *Stemming*. Hasil dari tahapan ini dapat dilihat pada (Tabel 1-5).

Tabel 1. Hasil *cleaning*

Sebelum	Sesudah
beliii cepet! gw dulu orang yang percaya semua hp sama aja, but trust me, dibalik harganya yang mahal kamu dapat yang sepadan. jangan lupa pake carger mfi yahh!	beliii cepet gw dulu orang yang percaya semua hp sama aja but trust me dibalik harganya yang mahal kamu dapat yang sepadan jangan lupa pake carger mfi yahh

Dari Tabel 1 dapat dilihat bahwa proses cleaning berhasil menghapus tanda baca yang tidak diperlukan seperti tanda seru dan titik, serta menghilangkan karakter huruf yang berulang secara berlebihan. Proses ini bertujuan untuk menyederhanakan teks agar lebih mudah diproses pada tahap selanjutnya dalam analisis teks.

Tabel 2. Hasil *case folding*

Sebelum	Sesudah
karna lu udh kemakan berita kalo IP TUH BAGUS BANGET DAN LU HARUS BELI IP, tapi disisi lain android juga ga kalah bagus nya	karna lu udh kemakan berita kalo ip tuh bagus banget dan lu harus beli ip, tapi disisi lain android juga ga kalah bagus nya

Dari Tabel 2 dapat dilihat bahwa proses *case folding* berhasil mengubah seluruh huruf kapital menjadi huruf kecil [13]. Hal ini penting untuk menyeragamkan teks sehingga kata-kata seperti "IP" dan "ip" dianggap sama oleh sistem, yang pada akhirnya membantu meningkatkan akurasi dalam proses analisis dan klasifikasi teks.

Tabel 3. Hasil *tokenized*

Sebelum	Sesudah
aku beli hp gak mandang kamera, kualitasnya buat keseharian	['aku', 'beli', 'hp', 'gak', 'ma ndang', 'kamera', 'kualitasnya', 'buat , 'keseharian']

Dari Tabel 3 dapat dilihat bahwa proses *tokenizing* berhasil memisahkan kalimat menjadi kata-kata individu dalam bentuk list, yang memudahkan proses analisis selanjutnya.

Tabel 4. Hasil *stopword*

Sebelum	Sesudah
karna lu udh kemakan berita kalo ip tuh bagus banget dan lu harus beli ip, tapi disisi lain android juga ga kalah bagus nya	['karna', 'lu', 'udh', 'kemakan', 'berita', 'kalo', 'ip', 'tuh', 'bagus', 'banget', 'lu', 'beli', 'ip', 'disisi', 'android', 'ga', 'kalah', 'bagus', 'nya']

Dari Tabel 4 dapat dilihat bahwa proses *stopword* berhasil menghapus kata-kata umum yang kurang bermakna dalam analisis, seperti "dan", "yang", atau "itu" [14], sehingga hanya menyisakan kata-kata penting saja.

Tabel 5. Hasil *stemming*

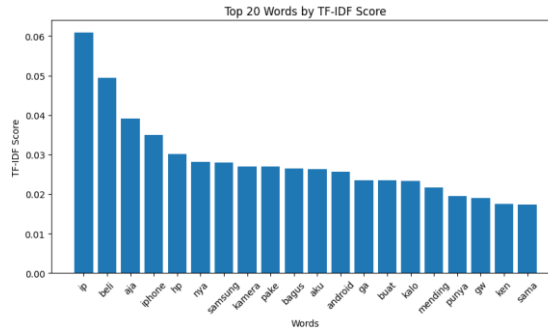
Sebelum	Sesudah
Kalo kenyamanan keknta tergantung orangnya si aku pake xr nyaman banget	['kalo', 'nyaman', 'keknta', 'gantung', 'orang', 'si', 'pake', 'xr', 'nyaman', 'banget']

Dari Tabel 5 dapat dilihat bahwa proses stemming berhasil mengubah kata-kata ke bentuk dasarnya, seperti "kenyamanan" menjadi "nyaman" dan "banget" tetap menjadi "banget", untuk mengurangi variasi kata yang tidak diperlukan dalam analisis. Proses ini bertujuan untuk menghilangkan imbuhan seperti awalan dan akhiran pada sebuah kata sehingga hanya menyisakan kata dasarnya [15].

3.3. Evaluasi

Berikut beberapa hasil evaluasi dalam penelitian ini;

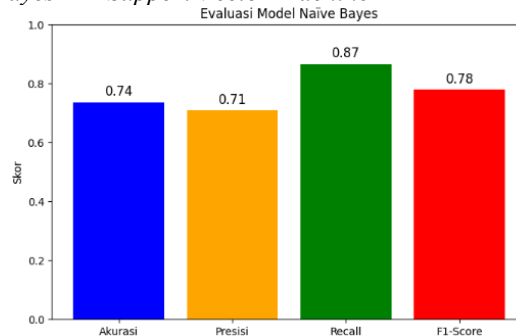
a. TF-IDF



Gambar 3. Visualisasi TF-IDF

Grafik batang vertikal ini menampilkan 20 kata teratas berdasarkan skor TF-IDF (*Term Frequency-Inverse Document Frequency*), yang menunjukkan kata-kata paling penting atau sering muncul dalam kumpulan data yang dianalisis. Kata dengan skor tertinggi adalah "ip", diikuti oleh "beli", "aja", "iphone", dan seterusnya. Grafik ini membantu mengidentifikasi kata kunci utama yang sering muncul dan dianggap relevan dalam analisis teks yang dilakukan.

b. Hasil *Performance Naive Bayes* dan *Support Vector Machine*



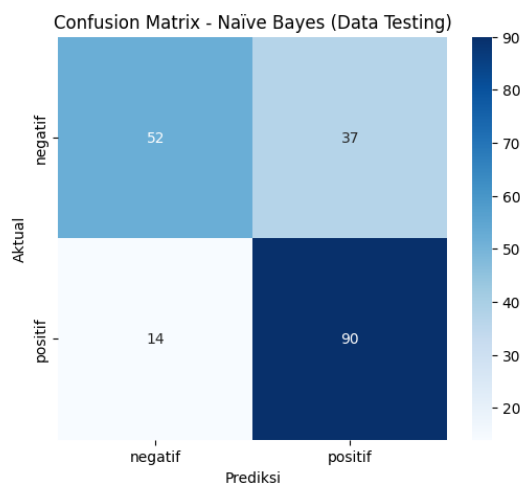
Gambar 4. Hasil *performance naive bayes*

Grafik di atas memperlihatkan hasil evaluasi kinerja model klasifikasi *Naive Bayes* berdasarkan empat metrik utama: Akurasi, Presisi, *Recall*, dan *F1-Score*. Secara umum, *recall* memperoleh nilai tertinggi (0,87), yang menunjukkan bahwa model memiliki kemampuan cukup baik dalam menangkap mayoritas data yang seharusnya diklasifikasikan sebagai positif. Hal ini berarti model cukup sensitif, dan cenderung tidak melewatkan data yang mengandung sentimen positif terhadap iPhone.

Namun, presisi yang hanya mencapai 0,71 menunjukkan bahwa tidak semua prediksi positif benar-benar valid, alias masih banyak data yang dikira positif tetapi nyatanya tidak demikian. Kondisi ini menunjukkan bahwa model lebih condong pada pendekatan "lebih baik salah menyertakan daripada melewatkan", yang bisa berdampak pada validitas analisis jika dipakai untuk pengambilan keputusan.

Nilai *F1-score* sebesar 0,78, yang merupakan rata-rata harmonis antara presisi dan *recall*, menegaskan adanya ketidakseimbangan tersebut. *F1-score* yang tidak terlalu tinggi mengindikasikan bahwa kinerja model masih bisa ditingkatkan, terutama jika digunakan dalam konteks yang memerlukan presisi tinggi, seperti analisis opini publik yang sensitif terhadap persepsi negatif.

Sementara itu, akurasi model berada di angka 0,74, yang secara umum masih tergolong cukup baik. Tetapi perlu dicatat bahwa akurasi bisa menyesatkan jika jumlah data positif dan negatif tidak seimbang. Oleh sebab itu, metrik lain seperti *recall* dan *F1-score* menjadi lebih penting untuk dijadikan acuan dalam menilai kualitas model dalam konteks data sosial media seperti komentar TikTok yang bersifat dinamis dan tidak terstruktur.



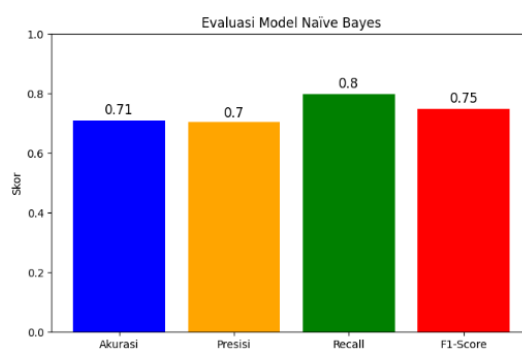
Gambar 5. *Confusion matrix naive bayes*

*Confusion matrix* pada gambar di atas menunjukkan performa model *Naïve Bayes* dalam mengklasifikasikan data sentimen pada tahap pengujian. Terdapat 52 data aktual negatif yang berhasil diprediksi dengan benar sebagai negatif, sedangkan 37 data negatif 89 data aktual negatif, model hanya mampu mengklasifikasikan 52 di antaranya secara tepat, sedangkan sisanya sebanyak 37 kasus salah diklasifikasikan sebagai positif. Hal ini menunjukkan adanya kelemahan model dalam mengenali sentimen negatif, yang dapat menyebabkan bias terhadap kelas positif.

Sebaliknya, dari 104 data aktual positif, model berhasil mengidentifikasi 90 dengan benar, dan hanya 14 yang salah diklasifikasikan sebagai negatif. Ini mengindikasikan bahwa model memiliki kecenderungan kuat untuk mendeteksi kelas positif, yang juga tercermin dalam nilai *recall* yang tinggi pada evaluasi sebelumnya.

Kondisi ini menunjukkan bahwa meskipun model cukup andal dalam menangkap data yang bersentimen positif, ia cenderung kurang akurat ketika menghadapi data negatif, yang ditunjukkan oleh tingginya *false positive*. Implikasi dari hal ini bisa signifikan tergantung pada konteks aplikasi misalnya, dalam sistem deteksi opini negatif, kesalahan ini berpotensi mengaburkan informasi penting.

Dengan demikian, model *Naïve Bayes* ini menunjukkan performa yang asimetris antara dua kelas, dan ke depan dapat dipertimbangkan strategi seperti penyeimbangan data atau tuning parameter untuk meningkatkan kemampuan klasifikasi pada kelas negatif.



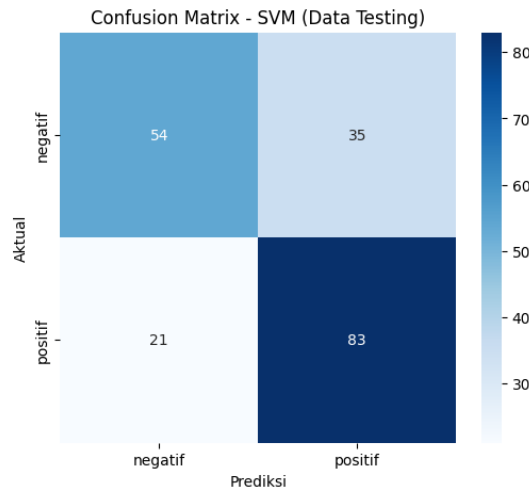
Gambar 6. Hasil *performance support vector machine*

Grafik batang ini memperlihatkan hasil evaluasi model *Support Vector Machine* (SVM) menggunakan empat metrik utama: akurasi, presisi, *recall*, dan *F1-score*. Nilai *recall* yang mencapai 0,80 menandakan bahwa model cukup efektif dalam mengidentifikasi sebagian besar data positif yang sebenarnya, yang berarti model cenderung lebih sensitif dan minim dalam melewatkan data penting. Ini menjadi nilai tambah jika tujuan utama adalah menangkap sebanyak mungkin sentimen positif secara menyeluruh.

Namun demikian, presisi model hanya mencapai 0,70, yang menunjukkan bahwa sebagian hasil prediksi positif sebenarnya tidak akurat. Artinya, ada proporsi kesalahan dalam mengklasifikasikan data sebagai positif padahal sebenarnya negatif. Ini dapat menjadi masalah jika kita memerlukan keakuratan tinggi dalam menyatakan bahwa suatu komentar benar-benar positif.

Dengan *F1-score* sebesar 0,75, model menunjukkan keseimbangan moderat antara presisi dan *recall*. Nilai ini cukup menggambarkan bahwa meskipun model mampu mengenali banyak data positif, kemampuannya dalam menghindari *false positive* masih bisa ditingkatkan.

Sementara itu, akurasi sebesar 0,71 memperlihatkan bahwa sekitar 71% dari keseluruhan prediksi model sesuai dengan label aslinya. Meskipun terlihat cukup baik, akurasi saja tidak cukup mewakili kinerja model jika distribusi data tidak seimbang. Oleh karena itu, penekanan terhadap metrik *recall* dan *F1-score* menjadi penting dalam menilai efektivitas model *Naïve Bayes* ini secara lebih komprehensif.



Gambar 7. *Confusion matrix support vector machine*

*Confusion matrix* ini menggambarkan hasil klasifikasi model SVM pada data uji. Dari 89 data aktual negatif, sebanyak 54 berhasil diidentifikasi dengan benar sebagai negatif, sementara 35 lainnya salah diklasifikasikan sebagai positif. Di sisi lain, dari 104 data aktual positif, model mampu mengenali 83 dengan tepat, sedangkan 21 kasus positif salah diprediksi sebagai negatif.

Secara keseluruhan, model SVM menunjukkan kecenderungan performa yang lebih seimbang dibandingkan model sebelumnya, namun masih terdapat tantangan dalam mengklasifikasikan data negatif. Tingginya jumlah *false positive* (35 kasus) menandakan bahwa model cenderung mengklasifikasikan beberapa data negatif sebagai positif, yang bisa menimbulkan kesalahan interpretasi, terutama dalam konteks deteksi opini atau sentimen negatif.

Meskipun begitu, kemampuan model untuk mengenali data positif tetap cukup kuat, sebagaimana dibuktikan dengan jumlah *true positive* yang relatif tinggi. Keseimbangan ini dapat menjadi dasar pertimbangan untuk optimasi lebih lanjut, seperti penyesuaian *threshold* atau penerapan teknik penyeimbangan data guna menekan tingkat kesalahan prediksi, khususnya pada kelas negatif.

Berdasarkan hasil evaluasi yang telah ditampilkan, model *Naïve Bayes* menunjukkan nilai *recall* yang lebih tinggi (0.80) dibandingkan model SVM, sementara nilai presisinya justru lebih rendah (0.70). Sebaliknya, model SVM memiliki presisi yang lebih baik, namun dengan nilai *recall* yang sedikit lebih rendah. Perbedaan ini berkaitan erat dengan karakteristik masing-masing algoritma. *Naïve Bayes*, yang mengasumsikan independensi antar fitur, cenderung lebih permisif dalam mengklasifikasikan data sebagai positif, sehingga mampu menangkap sebagian besar data positif (*high recall*), namun dengan konsekuensi munculnya lebih banyak *false positive*, yang menurunkan presisi. Di sisi lain, SVM cenderung lebih konservatif karena berfokus pada pencarian margin optimal antar kelas, sehingga lebih selektif dalam memberikan prediksi positif, menghasilkan presisi tinggi, tetapi berisiko melewatkan beberapa data positif (*lower recall*).

Dalam konteks klasifikasi komentar, perbedaan ini memiliki implikasi yang penting. Jika tujuan utama adalah memastikan semua komentar bermuatan positif atau berpotensi sensitif dapat terdeteksi (misalnya dalam moderasi konten), maka model dengan *recall* tinggi seperti *Naïve Bayes* lebih sesuai, karena kesalahan melewatkan komentar relevan (*false negative*) lebih berbahaya. Namun, jika kesalahan dalam menandai komentar netral sebagai komentar bermasalah perlu diminimalkan, maka model dengan presisi tinggi seperti SVM akan lebih tepat. Oleh karena itu, pemilihan model harus disesuaikan dengan prioritas sistem: apakah





#### 4. KESIMPULAN

Berdasarkan temuan dari penelitian ini, algoritma *Naive Bayes* menunjukkan performa yang lebih baik dibandingkan dengan algoritma *Support Vector Machine* (SVM) dalam menganalisis sentimen terhadap iPhone. *Naive Bayes* berhasil memperoleh akurasi tertinggi sebesar 74%, sedangkan akurasi maksimal yang dicapai oleh SVM adalah 71%.

Data yang telah dianalisis, mayoritas komentar yang diberikan oleh pengguna TikTok bersifat positif terhadap iPhone, menunjukkan citra produk yang kuat di mata publik. Hasil ini menunjukkan bahwa pendekatan analisis sentimen dapat memberikan gambaran umum terhadap persepsi Masyarakat.

#### REFERENSI

- [1] Anubhav, "Phone Users Exhibit Strong Brand Loyalty," Gizmochina. Accessed: Dec. 04, 2024. [Online]. Available: <https://www.gizmochina.com/2023/03/22/iphone-users-brand-loyalty/>
- [2] L. P. Febrian, "iPhone di Kalangan Milenial dan Gen Z," Kumparan. Accessed: Dec. 04, 2024. [Online]. Available: <https://kumparan.com/lidiapratamafebrian/iphone-di-kalangan-milenial-dan-gen-z-20SsC2P9aA1>
- [3] W. H. Silitonga and J. I. Sihotang, "Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naïve Bayesian Classification," *J. Teknol. Inf. dan Komun.*, vol. 9, no. 02, p. 116, 2019, doi: 10.36342/teika.v9i02.2199.
- [4] M. T. Hidayat, N. Suarna, and N. Rahaningsih, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Pt. Dilmoni Citra Mebel Indonesia," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, p. 694, 2023, doi: 10.36040/jati.v7i1.6310.
- [5] T. Juniardi and C. A. Sugianto, "Analisis Sentimen Tim Nasional Sepak Bola Indonesia di Turnamen Piala dunia U-17 Indonesia Pada Twitter (X) Menggunakan Algoritma Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 3715–3724, 2024, doi: <http://dx.doi.org/10.23960/jitet.v12i3S1.5188>.
- [6] M. Jain *et al.*, "Speech Emotion Recognition using Support," *arXiv (Cornell Univ.*, 2020, doi: <https://doi.org/10.48550/arXiv.2002.07590>.
- [7] H. Eldo, Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, "Penggunaan Algoritma Support Vector Machine ( SVM ) Untuk Deteksi Penipuan pada Transaksi Online," *J. Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, 2024, doi: <https://doi.org/10.33395/jmp.v13i2.14186>.
- [8] F. K. Ilmi, S. S. Julyinda, E. M. Kiswana, and M. Y. Fathoni, "Analisis Sentimen Emosi Publik Terhadap Kebijakan iPhone 16 Menggunakan Algoritma Naive Bayes," *Proc. Natl. Conf. Electr. Eng. Informatics, Ind. Technol. Creat. Media (CENTIVE 2024)*, vol. 4, no. 1, pp. 1141–1153, 2024, [Online]. Available: <https://centive.ittelkom-pwt.ac.id/index.php/centive/article/view/396>
- [9] F. A. Indriyani, A. Fauz, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 177–183, 2023, doi: 10.37373/tekno.v10i2.419.
- [10] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 36–44, 2021, doi: 10.25008/ijadis.v2i1.1216.
- [11] S. A. Rismawan and Y. Syahidin, "Implementasi Website Berita Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, pp. 167–178, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [12] W. Ningsih, B. Alfianda, Rahmadden, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 557–562, 2024, doi: 10.57152/malcom.v4i2.1253.
- [13] A. P. Astuti, S. Alam, and I. Jaelani, "Komparasi Algoritma Support Vector Machine dengan Naive Bayes Untuk Analisis Sentimen Pada Aplikasi BRImo," *J. Bangkit Indones.*, vol. 11, no. 2, pp. 1–7, 2022, doi: 10.52771/bangkitindonesia.v11i2.196.
- [14] A. Saepulrohman, S. Saepudin, and D. Gustian, "Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *is Best Account. Inf. Syst. Inf. Technol. Bus. Enterp. this is link OJS usf@.*, vol. 6, no. 2, pp. 91–105, 2019, doi: 10.34010/aisthebest.v6i2.4919.
- [15] R. Novaneliza, F. Handayani, R. J. Suhandar, H. Suroño, N. S. Azzahra, and D. Nadilla, "Perbandingan Algoritma Untuk Analisis Sentimen Pada Twitter Transportasi Umum Commuterline," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 13–21, 2023.