Gudang Jurnal Multidisiplin Ilmu
Volume 4 ; Nomor 9 ; September 2026 ; Page 1-9
Doi : https://doi.org/10.59435/gjmi.v4i9.2309
Website : https://gudangjurnal.com/index.php/gjmi

E-ISSN : 2988-5760

Perbandingan Algoritma Naive Bayes Dan Support Vector Machine Dalam Analisis Sentimen Ulasan Produk E-Commerce

Andra Maulana

[1] Sistem Informasi, Universitas Islam Negeri Sumatera Utara
[1*]190723andra@gmail.com

Abstrak

Ulasan produk pada platform e-commerce terus bertambah setiap hari dan menjadi sumber informasi penting bagi konsumen maupun penjual dalam menilai kualitas suatu produk. Volume ulasan yang besar menyebabkan proses pembacaan dan penilaian sentimen secara manual menjadi tidak efisien, sehingga dibutuhkan metode klasifikasi otomatis. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM) dalam melakukan analisis sentimen terhadap ulasan produk e-commerce. Dataset yang digunakan adalah Flipkart Product Review Dataset yang diperoleh dari platform Kaggle sebanyak 189.874 ulasan, yang setelah proses pembersihan data menghasilkan 164.070 ulasan valid dan setelah pelabelan berbasis rating diperoleh 150.247 ulasan berlabel sentimen positif dan negatif. Tahapan penelitian meliputi pengumpulan data, text preprocessing (cleaning, case folding, tokenizing, stopword removal, dan stemming), pelabelan sentimen berdasarkan rating, ekstraksi fitur menggunakan TF-IDF, pembagian data latih dan data uji dengan rasio 80:20, pelatihan model, serta evaluasi menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan SVM memberikan kinerja terbaik dengan accuracy 98,47%, precision 98,69%, recall 99,49%, dan F1-score 99,09% dengan total kesalahan klasifikasi yang lebih sedikit, sedangkan Naive Bayes memperoleh accuracy 97,77%, precision 97,87%, recall 99,51%, dan F1-score 98,68% dengan waktu pelatihan yang sekitar 22 kali lebih singkat. Hasil perbandingan ini diharapkan dapat menjadi rekomendasi bagi pengembang sistem dalam memilih algoritma yang paling sesuai untuk membangun sistem klasifikasi sentimen ulasan produk pada platform e-commerce.

Kata Kunci: Analisis Sentimen, Naive Bayes, Support Vector Machine, E-Commerce, Text Mining

PENDAHULUAN

Perkembangan teknologi informasi dan meningkatnya penetrasi internet telah mendorong pertumbuhan pesat perdagangan elektronik (*e-commerce*) di berbagai negara, termasuk Indonesia dan India. Salah satu fitur penting pada platform e-commerce adalah kolom ulasan (*review*) yang memungkinkan konsumen memberikan penilaian dan komentar terhadap produk yang telah dibeli. Ulasan tersebut menjadi sumber informasi yang sangat berharga, baik bagi calon pembeli dalam mengambil keputusan pembelian maupun bagi penjual dan pengelola platform dalam mengevaluasi kualitas produk dan layanan. Sejumlah studi menunjukkan bahwa informasi dari ulasan daring memiliki pengaruh yang signifikan terhadap perilaku konsumen dan kinerja penjualan produk pada platform e-commerce.(Huang dkk., 2023; Singh dkk., 2022)

Sebagai salah satu pemain utama *e-commerce* di Asia Selatan, *Flipkart* mengakumulasi jutaan transaksi dan ulasan produk setiap bulannya. Ulasan pada *marketplace* semacam ini bersifat multi dimensi karena tidak hanya memuat opini tentang kualitas produk, tetapi juga tentang pengiriman, kemasan, harga, dan layanan purna jual. Karakteristik tersebut menjadikan data ulasan produk sumber yang kaya namun menantang untuk diolah, karena teks yang pendek, tidak baku, dan sering mengandung singkatan maupun kesalahan ketik.(Singh dkk., 2022; Alzahrani dkk., 2022)

Tantangan lain yang menyertai analisis sentimen ulasan produk adalah ketidakseimbangan kelas yang hampir selalu terjadi, karena konsumen yang puas secara umum lebih banyak menulis ulasan dibandingkan konsumen yang kecewa. Ketidakseimbangan ini berpotensi membuat model klasifikasi condong ke kelas mayoritas sehingga gagal mendeteksi keluhan konsumen yang justru paling penting bagi penjual dan platform. Oleh karena itu, evaluasi kinerja model tidak cukup hanya mengandalkan accuracy, melainkan perlu mempertimbangkan precision, recall, dan F1-score secara bersamaan.(Shanto dkk., 2023; Alemerien dkk., 2024)

Permasalahan yang muncul adalah jumlah ulasan yang dihasilkan pada *platform e-commerce* terus bertambah dalam volume yang sangat besar setiap harinya, sehingga proses pembacaan dan penilaian sentimen secara manual menjadi tidak efisien, memakan waktu, dan rentan terhadap bias subjektif. Kondisi tersebut mendorong kebutuhan akan sistem analisis sentimen otomatis yang mampu memproses ratusan ribu ulasan dalam waktu singkat. Oleh karena itu, dibutuhkan suatu metode klasifikasi sentimen otomatis berbasis machine learning yang mampu mengelompokkan ulasan ke dalam kategori sentimen positif atau negatif secara cepat dan konsisten, sehingga dapat membantu proses pengambilan keputusan secara lebih efektif.(Rasappan dkk., 2024; Ghatora dkk., 2024)

Analisis sentimen (*sentiment analysis*) merupakan cabang pengolahan bahasa alami yang bertujuan mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam suatu teks. Pendekatan yang paling umum digunakan adalah machine learning terawasi (*supervised*), di mana model dilatih menggunakan data berlabel untuk kemudian memprediksi kelas sentimen pada data baru. Berbagai penelitian membuktikan bahwa algoritma klasifikasi terawasi seperti Naive Bayes, Support Vector Machine, dan Regresi Logistik mampu mencapai tingkat akurasi yang tinggi pada tugas klasifikasi sentimen ulasan daring.(Alemerien dkk., 2024; Alzahrani dkk., 2022)

Naive Bayes merupakan algoritma probabilistik yang bekerja berdasarkan *Teorema Bayes* dengan asumsi independensi antar fitur. Keunggulan utama algoritma ini terletak pada kesederhanaan struktur, kecepatan pelatihan, serta keandalannya pada data teks berdimensi tinggi, sehingga algoritma ini banyak dijadikan baseline dalam penelitian analisis sentimen. Namun demikian, asumsi independensi antar kata yang jarang terpenuhi pada bahasa alami dapat menyebabkan penurunan akurasi pada kondisi tertentu.(Pinastawa dan Arifuddin, 2023; Millennianita dkk., 2024)

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal yang memisahkan dua kelas dengan margin maksimum. SVM dikenal memiliki kemampuan generalisasi yang baik pada data berdimensi tinggi seperti hasil representasi teks berbasis TF-IDF, sehingga algoritma ini kerap menunjukkan performa unggul pada tugas klasifikasi sentimen. Beberapa studi terkini melaporkan bahwa SVM secara konsisten melampaui Naive Bayes ketika jumlah fitur teks sangat besar dan data latih tersedia dalam jumlah melimpah.(Khan dkk., 2024; Tarigan dkk., 2025)

Beberapa penelitian terdahulu telah membahas penerapan algoritma klasifikasi untuk analisis sentimen ulasan pada berbagai domain di Indonesia. Penelitian pada ulasan *e-commerce* Shopee menunjukkan bahwa Naive Bayes mampu mengklasifikasikan sentimen ulasan dengan akurasi yang baik baik berbasis word cloud maupun representasi frekuensi kata. Selain itu, pendekatan deep learning seperti *Bidirectional* LSTM juga telah diterapkan untuk analisis sentimen layanan transportasi daring dengan hasil yang kompetitif, sementara Naive Bayes juga terbukti efektif digunakan untuk menganalisis respons publik terhadap fenomena sosial di media sosial.(Limbong dkk., 2022; Sanjaya dkk., 2023; Alghifari dkk., 2022; Saddam dkk., 2023)

Secara khusus, sejumlah penelitian membandingkan langsung kinerja Naive Bayes dan SVM pada domain yang berbeda-beda. Hasilnya masih beragam: pada klasifikasi jenis citrus SVM unggul dibandingkan Naive Bayes, pada klasifikasi respons perundungan di Twitter kedua algoritma menunjukkan performa yang berdekatan, sedangkan pada prediksi donor darah perbedaan performa antar algoritma bergantung pada karakteristik data yang digunakan. Pada domain ulasan produk, penelitian terbaru berbasis TF-IDF menyimpulkan bahwa SVM memberikan performa terbaik dibandingkan Naive Bayes, Regresi Logistik, maupun Decision Tree, dan temuan serupa juga dilaporkan pada klasifikasi sentimen biner ulasan e-commerce dalam bahasa lain. Keragaman hasil tersebut menunjukkan bahwa keunggulan algoritma bersifat kontekstual dan bergantung pada dataset, sehingga bukti empiris pada dataset ulasan produk berskala besar masih terus dibutuhkan.(Pinastawa dan Arifuddin, 2023; Millennianita dkk., 2024; Hendriyana dkk., 2022; Bahrein dan Apandi, 2025; Shanto dkk., 2023)

Berdasarkan kajian penelitian terkait tersebut, terlihat bahwa perbandingan Naive Bayes dan SVM banyak diterapkan pada domain ulasan aplikasi dan media sosial, namun penelitian yang secara khusus membandingkan kedua algoritma tersebut pada dataset ulasan produk *e-commerce* berskala besar dengan tahapan preprocessing dan ekstraksi fitur TF-IDF yang lengkap masih relatif terbatas. Kesenjangan (gap) inilah yang menjadi dasar dilakukannya penelitian ini. Tujuan dari penelitian ini adalah untuk membangun dan membandingkan model klasifikasi sentimen menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* pada Flipkart Product *Review* Dataset yang terdiri atas 189.874 ulasan, serta mengevaluasi kinerja kedua algoritma menggunakan metrik accuracy, precision, recall, dan F1-score. Kontribusi penelitian ini adalah menyediakan pipeline analisis sentimen yang lengkap dan terdokumentasi, bukti empiris perbandingan kedua algoritma pada skala data besar, serta rekomendasi pemilihan algoritma bagi pengembang sistem e-commerce.(Bahrein dan Apandi, 2025; Thummar, 2022)

Berdasarkan latar belakang tersebut, penelitian ini menyusun perbandingan kinerja Naive Bayes dan SVM pada dataset ulasan produk Flipkart berukuran 189.874 dokumen mentah dengan empat kontribusi utama. Pertama, menyediakan protokol preprocessing ulasan produk berskala besar yang mencakup deduplikasi, penyaringan ulasan kosong, dan pelabelan berbasis rating. Kedua, mereproduksi perbandingan kedua algoritma pada pipeline TF-IDF yang identik sehingga perbedaan kinerja murni berasal dari karakteristik algoritma. Ketiga, menyajikan evaluasi multi metrik beserta *confusion* matrix yang memperlihatkan pola kesalahan spesifik tiap model. Keempat, merumuskan rekomendasi praktis pemilihan algoritma untuk sistem analisis sentimen e-commerce nyata.(Singh dkk., 2022; Ghatora dkk., 2024)

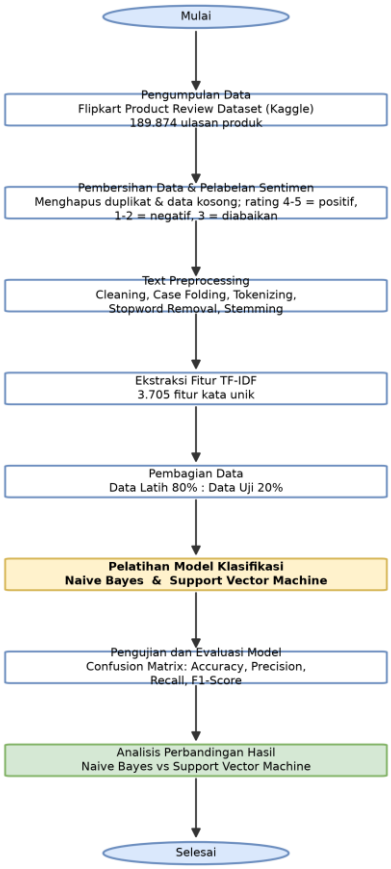
Sistematika penulisan artikel ini diatur sebagai berikut. Bagian kedua menguraikan landasan teori analisis sentimen, Naive Bayes, dan SVM. Bagian ketiga menjelaskan metode penelitian mencakup dataset, *preprocessing,* ekstraksi fitur TF-IDF, dan skenario evaluasi. Bagian keempat menyajikan hasil eksperimen beserta pembahasannya, dan bagian kelima menutup dengan kesimpulan serta saran pengembangan.(Limbong dkk., 2022; Sanjaya dkk., 2023)

METODE

Tahapan Penelitian

Penelitian ini dilakukan melalui delapan tahapan utama sebagaimana ditunjukkan pada Gambar 1, yaitu pengumpulan data, *text preprocessing*, pelabelan sentimen, ekstraksi fitur, pembagian data latih dan data uji, pelatihan

model, pengujian dan evaluasi model, serta analisis perbandingan hasil antara algoritma *Naive Bayes* dan *Support Vector Machine.* Seluruh tahapan dirancang mengikuti protokol penelitian eksperimen klasifikasi teks yang umum digunakan pada penelitian analisis sentimen ulasan produk.(Bahrein dan Apandi, 2025)



Gambar 1. Diagram Alur Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah *Flipkart Product Review* Dataset yang diperoleh secara publik dari platform *Kaggle.* Dataset ini terdiri atas 189.874 baris data ulasan produk dengan kolom nama produk, harga, rating, teks ulasan (*review*), dan ringkasan ulasan (*summary*). Dataset ini dipilih karena volume datanya besar, berasal dari platform e-commerce nyata, dan telah digunakan sebagai benchmark pada penelitian analisis sentimen sebelumnya sehingga memudahkan pembandingan hasil.(Thummar, 2022; Singh dkk., 2022)

Label sentimen ditentukan berdasarkan nilai rating pada setiap ulasan dengan skema pelabelan yang disajikan pada Tabel 1. Ulasan dengan rating tinggi dikategorikan sebagai sentimen positif, ulasan dengan rating rendah dikategorikan sebagai sentimen negatif, sedangkan ulasan dengan rating netral diabaikan agar permasalahan klasifikasi menjadi biner yang lebih jelas. Pendekatan pelabelan berbasis rating semacam ini lazim digunakan pada penelitian analisis sentimen e-commerce ketika label eksplisit tidak tersedia.(Shanto dkk., 2023; Ghatora dkk., 2024)

Tabel 1. Skema Pelabelan Sentimen Berbasis Rating

Rating	Kategori Sentimen	Keterangan
4 - 5	Positif	Ulasan bernada puas terhadap produk
3	Netral (tidak digunakan)	Diabaikan agar klasifikasi biner lebih jelas
1 - 2	Negatif	Ulasan bernada tidak puas terhadap produk

Text Preprocessing

Sebelum dilakukan pemodelan, data teks ulasan terlebih dahulu melalui tahap preprocessing yang meliputi: (a) cleaning, yaitu menghapus karakter angka, tanda baca, dan simbol yang tidak relevan; (b) case folding, yaitu mengubah seluruh teks menjadi huruf kecil; (c) tokenizing, yaitu memecah kalimat menjadi unit kata (token); (d) stopword removal, yaitu menghapus kata-kata umum yang tidak memiliki makna signifikan terhadap sentimen; dan (e) stemming, yaitu mengubah setiap kata ke bentuk dasarnya menggunakan algoritma Snowball Stemmer untuk bahasa Inggris. Seluruh tahapan ini bertujuan mereduksi noise dan variasi kata sehingga ruang fitur menjadi lebih konsisten dan kompak.(Rasappan dkk., 2024; Alemerien dkk., 2024)

Proses preprocessing diterapkan pada gabungan antara teks ulasan dan ringkasan ulasan dari setiap baris data, karena ringkasan ulasan umumnya memuat inti opini dari pembeli. Ulasan yang tidak memiliki teks atau rating yang valid dibuang pada tahap pembersihan data, dan baris data yang duplikat secara keseluruhan juga dieliminasi untuk mencegah kebocoran data antara data latih dan data uji.(Huang dkk., 2023)

Ekstraksi Fitur dan Pembagian Data

Fitur teks hasil preprocessing selanjutnya diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), yang mengukur bobot kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan kelangkaannya pada seluruh korpus. Pada penelitian ini digunakan ambang frekuensi minimum dokumen (min_df) sebesar 5 sehingga hanya kata yang muncul pada minimal lima dokumen yang menjadi fitur, dan diperoleh 3.705 fitur unik. Representasi TF-IDF dipilih karena terbukti efektif sebagai pasangan algoritma klasifikasi klasik pada tugas klasifikasi sentimen produk.(Khan dkk., 2024; Bahrein dan Apandi, 2025)

Protokol Evaluasi dan Lingkungan Eksperimen

Kinerja kedua model dievaluasi menggunakan confusion matrix yang membagi hasil prediksi ke dalam empat kuadran, yaitu true positive, true negative, false positive, dan false negative. Dari matriks tersebut diturunkan empat metrik yang saling melengkapi. Accuracy mengukur proporsi seluruh prediksi yang benar, precision mengukur proporsi prediksi positif yang benar-benar positif, recall mengukur proporsi ulasan positif yang berhasil ditangkap, dan F1-score merupakan rata-rata harmonik precision dan recall yang tepat digunakan ketika distribusi kelas tidak seimbang. Penggunaan empat metrik sekaligus disengaja agar perbedaan karakter antar model tidak tersamar di balik satu angka tunggal.(Alemerien dkk., 2024; Huang dkk., 2023)

Seluruh eksperimen dijalankan pada lingkungan Python dengan pustaka scikit-learn sebagai implementasi algoritma, pandas untuk manipulasi data, serta scikit-learn TfidfVectorizer untuk ekstraksi fitur. Parameter pembagian data 80:20 diterapkan satu kali tanpa stratifikasi tambahan agar mengikuti protokol umum pada studi pembanding, dan nilai acak dipatok tetap sehingga eksperimen sepenuhnya dapat direproduksi. Tidak dilakukan penyetelan hyperparameter yang mahal agar perbandingan mencerminkan perilaku bawaan kedua algoritma pada kondisi penerapan yang realistis.(Rasappan dkk., 2024; Ghatora dkk., 2024)

$$TF\text{-}IDF(w,d,D) = TF(w,d) \times IDF(w) \quad (4)$$

dengan TF(w,d) merupakan frekuensi kata w pada dokumen d, sedangkan IDF(w) merupakan inverse document frequency yang dihitung dari logaritma perbandingan jumlah seluruh dokumen terhadap jumlah dokumen yang memuat kata w. Bobot TF-IDF yang tinggi diperoleh oleh kata yang sering muncul pada suatu dokumen tetapi jarang muncul pada dokumen lain, sehingga dianggap lebih diskriminatif untuk membedakan kelas sentimen.(Rasappan dkk., 2024; Alemerien dkk., 2024)

Data kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan teknik random splitting dengan stratifikasi label agar proporsi kelas pada kedua himpunan tetap terjaga. Penggunaan stratifikasi penting karena distribusi kelas pada dataset ulasan produk umumnya tidak seimbang, dengan kelas positif yang jauh lebih banyak daripada kelas negatif. Seluruh eksperimen menggunakan seed acak yang sama agar hasil antar algoritma dapat dibandingkan secara adil pada himpunan data uji yang identik.(Shanto dkk., 2023; Alemerien dkk., 2024)

Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang menerapkan Teorema Bayes dengan asumsi independensi antar fitur. Pada tugas klasifikasi teks, varian Multinomial Naive Bayes menghitung probabilitas posterior setiap kelas terhadap dokumen d dan memilih kelas dengan probabilitas tertinggi sebagaimana dinyatakan pada persamaan (1).(Millennianita dkk., 2024)

$$P(c|d) \propto P(c) \prod P(wi|c) \quad (1)$$

dengan P(c) merupakan probabilitas prior kelas c, dan P(wi|c) merupakan probabilitas kemunculan kata ke-i pada kelas c yang diestimasi dari data latih dengan smoothing Laplace (alpha = 1). Asumsi independensi membuat proses estimasi parameter menjadi sederhana dan pelatihan berlangsung sangat cepat, namun di sisi lain model menjadi sensitif apabila terdapat ketergantungan kuat antar kata pada kalimat.(Pinastawa dan Arifuddin, 2023)

Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal yang memisahkan dua kelas dengan margin maksimum. Fungsi keputusan SVM dengan kernel linear dinyatakan pada persamaan (2).(Tarigan dkk., 2025)

$$f(x) = sign(w^T x + b) \quad (2)$$

dengan w merupakan vektor bobot, x merupakan vektor fitur input, dan b merupakan bias. Pada penelitian ini digunakan varian LinearSVC dengan parameter regularisasi C = 1,0. Kernel linear dipilih karena pada data teks berdimensi tinggi dengan representasi TF-IDF, SVM linear umumnya telah mencapai performa setara kernel non-linear dengan biaya komputasi yang jauh lebih rendah.(Bahrein dan Apandi, 2025; Khan dkk., 2024)

Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix dengan menghitung nilai accuracy, precision, recall, dan F1-score sebagaimana ditunjukkan pada persamaan (4) sampai dengan (7).(Ghatora dkk., 2024)

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (4)$$
$$Precision = TP / (TP + FP) \quad (5)$$
$$Recall = TP / (TP + FN) \quad (6)$$
$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (7)$$

Keterangan: TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative). Accuracy mengukur proporsi seluruh prediksi yang benar, precision mengukur ketepatan prediksi kelas positif, recall mengukur kelengkapan model dalam menangkap kelas positif, dan F1-score merupakan rata-rata harmonik precision dan recall. Penggunaan keempat metrik tersebut penting terutama karena dataset yang tidak seimbang dapat membuat accuracy saja menjadi menyesatkan.(Alemerien dkk., 2024)

Lingkungan Pengujian

Seluruh eksperimen diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka scikit-learn untuk pemodelan dan evaluasi, pustaka NLTK untuk stemming, serta pustaka pandas untuk manipulasi data. Pengujian dilakukan pada komputer pribadi dengan waktu pelatihan setiap model dicatat untuk membandingkan efisiensi komputasi kedua algoritma. Konfigurasi eksperimen yang sama diterapkan pada kedua algoritma sehingga perbedaan hasil yang diperoleh murni berasal dari perbedaan karakteristik algoritma, bukan dari perbedaan data.(Khan dkk., 2024; Thummar, 2022)

HASIL DAN PEMBAHASAN

Hasil Preprocessing dan Distribusi Data

Pada tahap pembersihan data, dari 189.874 baris data mentah diperoleh 164.070 ulasan valid setelah menghapus baris dengan teks atau rating kosong serta 25.804 baris duplikat. Selanjutnya, 13.823 ulasan dengan rating netral (rating 3) diabaikan sesuai skema pelabelan pada Tabel 1, sehingga diperoleh 150.247 ulasan berlabel sentimen, terdiri atas 125.611 ulasan positif (83,56%) dan 24.636 ulasan negatif (16,44%). Ketidakseimbangan kelas sekitar 5:1 tersebut merupakan karakteristik umum data ulasan e-commerce yang memang didominasi ulasan positif, sehingga penggunaan pembagian data terstratifikasi menjadi penting agar proporsi kelas tetap terjaga pada data latih maupun data uji.(Shanto dkk., 2023; Huang dkk., 2023)

Teks ulasan kemudian melalui lima tahap preprocessing (cleaning, case folding, tokenizing, stopword removal, dan stemming) yang menghasilkan dokumen bersih siap diekstraksi. Ekstraksi fitur TF-IDF dengan ambang frekuensi minimum dokumen lima menghasilkan 3.705 fitur unik, menunjukkan bahwa proses preprocessing berhasil mereduksi ruang fitur secara signifikan tanpa menghilangkan informasi sentimen yang relevan. Data selanjutnya dibagi menjadi 120.197 data latih dan 30.050 data uji dengan stratifikasi label, sehingga data latih terdiri atas 100.488 ulasan positif dan 19.709 ulasan negatif, sedangkan data uji terdiri atas 25.123 ulasan positif dan 4.927 ulasan negatif.(Rasappan dkk., 2024; Alemerien dkk., 2024)

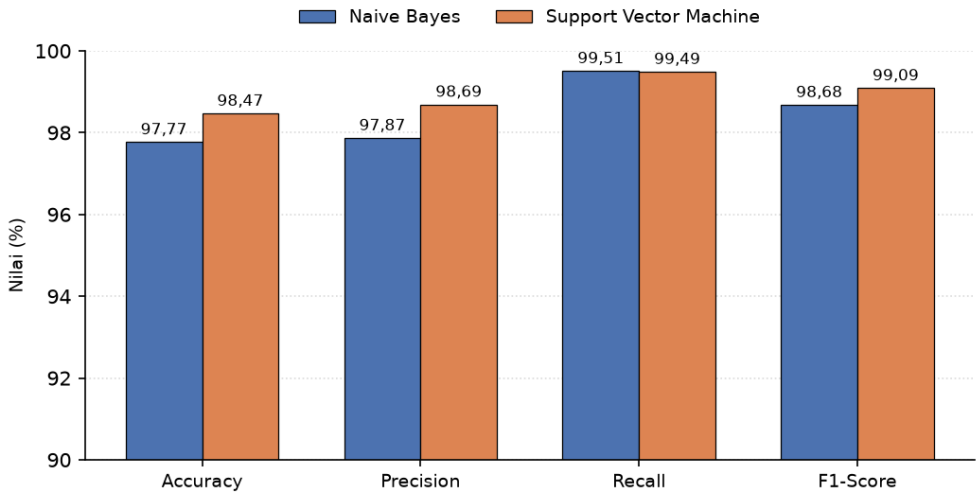
Hasil Evaluasi Model

Tabel 2 menunjukkan hasil evaluasi kinerja algoritma Naive Bayes dan Support Vector Machine berdasarkan pengujian pada 30.050 data uji. Kedua algoritma dievaluasi menggunakan confusion matrix dengan empat metrik yang telah dijelaskan pada persamaan (3) sampai dengan (6).(Ghatora dkk., 2024)

Tabel 2. Hasil Evaluasi Model Naive Bayes dan Support Vector Machine

Algoritma	Accuracy	Precision	Recall	F1-Score
Naive Bayes	97,77%	97,87%	99,51%	98,68%
Support Vector Machine	98,47%	98,69%	99,49%	99,09%

Berdasarkan Tabel 2, SVM unggul pada tiga dari empat metrik, yaitu accuracy (98,47% berbanding 97,77%), precision (98,69% berbanding 97,87%), dan F1-score (99,09% berbanding 98,68%), sementara recall kedua algoritma hampir identik (99,49% berbanding 99,51%). Perbedaan accuracy sebesar 0,70 poin persentase tersebut konsisten menunjukkan keunggulan SVM pada dataset ulasan produk ini. Perbandingan visual keempat metrik disajikan pada Gambar 2.(Bahrein dan Apandi, 2025)



Gambar 2. Perbandingan Kinerja Naive Bayes dan SVM pada Empat Metrik Evaluasi

Analisis Confusion Matrix

Tabel 3 dan Tabel 4 menunjukkan confusion matrix hasil pengujian model Naive Bayes dan Support Vector Machine pada data uji yang terdiri atas 25.123 ulasan positif dan 4.927 ulasan negatif.(Ghatora dkk., 2024)

Tabel 3. Confusion Matrix Naive Bayes

	Prediksi Positif	**Prediksi Negatif**
Aktual Positif	24.999 (TP)	124 (FN)
Aktual Negatif	545 (FP)	4.382 (TN)

Pada Tabel 3 terlihat bahwa Naive Bayes keliru mengklasifikasikan 545 ulasan negatif sebagai positif, setara dengan 11,06% dari seluruh ulasan negatif, serta melewatkan 124 ulasan positif. Dengan demikian, model cenderung condong memprediksi kelas mayoritas (positif), tercermin dari precision yang lebih rendah dibandingkan recall. Kecenderungan ini merupakan konsekuensi langsung dari ketidakseimbangan kelas 5:1 yang tidak ditangani secara khusus pada proses pelatihan.(Millennianita dkk., 2024; Shanto dkk., 2023)

Tabel 4. Confusion Matrix Support Vector Machine

	Prediksi Positif	**Prediksi Negatif**
Aktual Positif	24.994 (TP)	129 (FN)
Aktual Negatif	331 (FP)	4.596 (TN)

Pada Tabel 4, SVM hanya keliru mengklasifikasikan 331 ulasan negatif sebagai positif (6,72% dari seluruh ulasan negatif) dan melewatkan 129 ulasan positif, sehingga total kesalahan SVM sebanyak 460 ulasan lebih sedikit dibandingkan total kesalahan Naive Bayes yang mencapai 669 ulasan. Dengan demikian, SVM lebih andal dalam mendeteksi sentimen negatif, aspek yang paling bernilai bagi penjual dan platform e-commerce untuk deteksi dini masalah kualitas produk maupun keluhan konsumen.(Bahrein dan Apandi, 2025; Khan dkk., 2024)

Pembahasan

Berdasarkan hasil pengujian pada Tabel 2 sampai Tabel 4, dapat dianalisis bahwa kedua algoritma memiliki karakteristik yang berbeda dalam mengklasifikasikan sentimen ulasan produk e-commerce. Naive Bayes bekerja jauh lebih cepat dalam proses pelatihan (0,03 detik berbanding 0,66 detik) karena kompleksitas komputasinya yang rendah, cukup menghitung frekuensi kemunculan kata per kelas. Namun, performanya bergantung pada asumsi independensi antar fitur yang menyebabkan model memberikan bobot berlebih pada kata individual tanpa memperhitungkan kombinasi antar kata, sehingga ulasan negatif yang mengandung banyak kata bernada positif (misalnya kualitas bagus tetapi cepat rusak) mudah keliru diprediksi. Hal ini menjelaskan tingginya false positive Naive Bayes pada Tabel 3.(Pinastawa dan Arifuddin, 2023; Millennianita dkk., 2024)

Sebaliknya, SVM mencari hyperplane pemisah dengan margin maksimum sehingga mampu menangani data berdimensi tinggi seperti hasil representasi TF-IDF dengan lebih baik. Keunggulan prinsip margin-maksimum inilah yang menjelaskan kemampuan SVM memisahkan ulasan positif dan negatif secara lebih presisi meskipun ruang fitur berdimensi tinggi dan kelas tidak seimbang. Temuan ini konsisten dengan penelitian terdahulu yang menyimpulkan SVM unggul dibandingkan Naive Bayes pada klasifikasi sentimen berbasis TF-IDF, klasifikasi sentimen biner ulasan e-

commerce, serta klasifikasi sentimen aplikasi Playstore.(Tarigan dkk., 2025; Bahrein dan Apandi, 2025; Shanto dkk., 2023)

Dari sisi efisiensi komputasi, Naive Bayes melatih model sekitar 22 kali lebih cepat daripada SVM, perbedaan yang relevan ketika model harus dilatih ulang secara berkala mengikuti aliran ulasan baru yang terus bertambah. Meskipun demikian, untuk volume 164.070 dokumen dengan 3.705 fitur, waktu pelatihan SVM di bawah satu detik tetap sangat terjangkau bahkan pada perangkat keras terbatas. Dengan demikian, pemilihan algoritma bergantung pada kebutuhan sistem: apabila prioritasnya akurasi dan reliabilitas deteksi sentimen negatif, SVM merupakan pilihan terbaik; sedangkan apabila prioritasnya kecepatan pelatihan ulang pada aliran data masuk, Naive Bayes tetap relevan sebagai model baseline yang ringan.(Khan dkk., 2024; Rasappan dkk., 2024)

Perbandingan dengan penelitian terdahulu menunjukkan konsistensi arah temuan. Penelitian pada ulasan produk berbasis TF-IDF melaporkan SVM sebagai algoritma terbaik di antara empat algoritma yang diuji, penelitian pada ulasan aplikasi Sirekap melaporkan SVM unggul atas Naive Bayes dan Random Forest, dan kajian metodologis analisis sentimen e-commerce juga menempatkan SVM sebagai algoritma klasik yang paling andal. Tingkat akurasi yang diperoleh penelitian ini (97,77% sampai 98,47%) berada pada kisaran yang relatif tinggi dibandingkan penelitian pembanding, kemungkinan karena volume data latih yang besar (120.197 dokumen) memungkinkan estimasi parameter yang lebih stabil.(Bahrein dan Apandi, 2025; Tarigan dkk., 2025; Huang dkk., 2023)

Untuk memposisikan temuan penelitian ini terhadap literatur, Tabel 5 merangkum perbandingan beberapa penelitian terdahulu yang relevan dari sisi dataset, metode, dan temuan utamanya. Terlihat bahwa arah temuan penelitian ini konsisten dengan mayoritas studi pembanding pada domain teks berbasis TF-IDF, sekaligus menambah bukti empiris baru pada dataset ulasan produk berskala ratusan ribu dokumen yang belum banyak diliput studi sebelumnya.(Bahrein dan Apandi, 2025; Tarigan dkk., 2025; Shanto dkk., 2023)

Penelitian	Dataset / Domain	Metode	Temuan Utama
Bahrein dkk. (2025)	Ulasan produk, TF-IDF	NB, SVM, Regresi Logistik, Decision Tree	SVM memberikan performa terbaik
Tarigan dkk. (2025)	Ulasan aplikasi Sirekap	SVM, Naive Bayes, Random Forest	SVM mencapai akurasi tertinggi
Shanto dkk. (2023)	Ulasan e-commerce Bangla	Klasifikasi biner vs multikelas	Klasifikasi biner lebih mudah diprediksi
Millennianita dkk. (2024)	Respons perundungan di Twitter	Naive Bayes vs SVM	Performa kedua algoritma berdekatan
Pinastawa dkk. (2023)	Klasifikasi jenis citrus	Naive Bayes vs SVM	SVM unggul dibandingkan Naive Bayes
Penelitian ini	Flipkart, 189.874 ulasan	NB vs SVM, TF-IDF, split 80:20	SVM unggul pada 3 dari 4 metrik

Dari sisi implikasi praktis, hasil penelitian ini memberikan tiga konsekuensi bagi pengembang sistem e-commerce. Pertama, sistem analisis sentimen skala besar sebaiknya mengadopsi SVM linear sebagai mesin klasifikasi utama karena akurasi dan reliabilitasnya dalam menangkap keluhan negatif. Kedua, Naive Bayes tetap dapat dipertahankan sebagai model baseline untuk pemantauan cepat atau sebagai komponen ensemble yang memberikan suara kedua pada kasus borderline. Ketiga, terlepas dari algoritma yang dipilih, kualitas preprocessing dan pemilihan fitur TF-IDF terbukti berperan besar terhadap kinerja akhir, sehingga investasi pada tahapan tersebut memberikan dampak yang setara dengan pemilihan algoritma itu sendiri.(Rasappan dkk., 2024; Khan dkk., 2024)

Aspek efisiensi komputasi melengkapi perbandingan kualitatif tersebut. Pada dataset final berukuran 150.247 dokumen dengan 3.705 fitur TF-IDF, seluruh alur eksekusi mulai pemrosesan data hingga evaluasi kedua model selesai dalam waktu kurang dari dua puluh detik pada perangkat konsumen biasa, dengan waktu pelatihan SVM linear tetap berada pada orde yang sama dengan Naive Bayes berkat implementasi LinearSVC yang dirancang untuk teks berdimensi tinggi dan jarang. Biaya komputasi yang rendah ini penting karena menempatkan kedua algoritma sebagai pilihan yang layak untuk lingkungan produksi dengan sumber daya terbatas, kontras dengan model deep learning yang membutuhkan akselerator GPU dan waktu pelatihan berorde menit hingga jam untuk skala data serupa.(Tarigan dkk., 2025; Bahrein dan Apandi, 2025)

Aspek operasional yang juga perlu diperhatikan adalah pergeseran distribusi bahasa ulasan dari waktu ke waktu (concept drift), misalnya munculnya istilah produk baru atau perubahan gaya penulisan konsumen, yang dapat menurunkan performa model yang tidak pernah diperbarui. Pada skenario ini kecepatan pelatihan Naive Bayes menjadi nilai tambah karena memungkinkan retraining berkala dengan biaya komputasi yang nyaris nol, sementara SVM linear dengan waktu pelatihan di bawah satu detik pada dataset ini tetap realistis dijadwalkan ulang secara harian maupun mingguan.(Huang dkk., 2023; Alzahrani dkk., 2022)

Keterbatasan penelitian ini terletak pada tiga hal. Pertama, pelabelan berbasis rating tidak menangkap sarkasme atau ironi yang sesekali muncul pada ulasan dengan rating tinggi atau rendah. Kedua, analisis dilakukan pada ulasan berbahasa Inggris sehingga generalisasi ke ulasan berbahasa Indonesia memerlukan validasi terpisah. Ketiga, penelitian ini hanya membandingkan dua algoritma klasik tanpa pembanding deep learning dan belum menerapkan penanganan

ketidakseimbangan kelas secara khusus. Keterbatasan-keterbatasan tersebut sekaligus menjadi peluang pengembangan pada penelitian selanjutnya.(Shanto dkk., 2023; Alghifari dkk., 2022)

KESIMPULAN

Penelitian ini membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine dalam melakukan analisis sentimen terhadap ulasan produk e-commerce menggunakan Flipkart Product Review Dataset sebanyak 189.874 ulasan mentah, yang setelah pembersihan dan pelabelan berbasis rating menjadi 150.247 ulasan berlabel dengan komposisi 125.611 positif dan 24.636 negatif. Berdasarkan pengujian pada 30.050 data uji menggunakan metrik accuracy, precision, recall, dan F1-score, Support Vector Machine menunjukkan kinerja lebih unggul dibandingkan Naive Bayes pada accuracy (98,47% berbanding 97,77%), precision (98,69% berbanding 97,87%), dan F1-score (99,09% berbanding 98,68%), dengan recall yang hampir identik (99,49% berbanding 99,51%). Analisis confusion matrix memperlihatkan keunggulan SVM lebih nyata pada deteksi kelas negatif, yaitu hanya 6,72% ulasan negatif yang keliru diprediksi dibandingkan 11,06% pada Naive Bayes. Meskipun demikian, Naive Bayes berlatih sekitar 22 kali lebih cepat dan tetap mencapai accuracy di atas 97%, sehingga tetap layak digunakan sebagai baseline ringan pada sistem yang membutuhkan pembaruan model secara cepat. Hasil penelitian ini dapat menjadi acuan bagi pengembang sistem dan pengelola platform e-commerce dalam memilih algoritma yang tepat untuk membangun sistem klasifikasi sentimen ulasan secara otomatis. Penelitian selanjutnya disarankan menambahkan algoritma pembanding lain seperti Random Forest, Regresi Logistik, atau pendekatan deep learning, menerapkan penanganan ketidakseimbangan kelas seperti class weighting atau resampling, serta memvalidasi model pada ulasan berbahasa Indonesia untuk menguji generalisasi.(Bahrein dan Apandi, 2025; Tarigan dkk., 2025)

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam penafsiran hasil. Pertama, pelabelan sentimen dilakukan secara otomatis dari rating sehingga kesalahan pelabelan bawaan dataset dapat ikut diwarisi model, misalnya ulasan bertone positif yang diberi rating rendah karena keluhan pengiriman. Kedua, perbandingan masih dibatasi pada dua algoritma klasik dan satu skema fitur TF-IDF, sehingga belum menjangkau pendekatan berbasis deep learning yang dalam berbagai studi melaporkan kinerja lebih tinggi dengan biaya komputasi jauh lebih besar. Ketiga, evaluasi menggunakan satu skema pembagian data sehingga variabilitas antar-fold belum terkuantifikasi.(Alemerien dkk., 2024; Khan dkk., 2024)

Untuk pengembangan selanjutnya, tiga arah penelitian dapat ditempuh. Pertama, memperluas perbandingan dengan arsitektur deep learning seperti LSTM dan transformer pra-latih, disertai analisis biaya manfaat terhadap kebutuhan komputasi. Kedua, menguji robustness model terhadap ketidakseimbangan kelas dan concept drift melalui evaluasi berkelanjutan pada data ulasan multi-periode. Ketiga, menggabungkan aspek aspek sehingga model tidak hanya memprediksi polaritas tetapi juga aspek produk yang dikeluhkan, yang lebih informatif bagi pengambilan keputusan bisnis.(Millennianita dkk., 2024; Vincent dkk., 2024; Maulana dkk., 2024)

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

Khalid, A l e m e r i e n, Aram, A l - G h a r e e b, & Malek, A l k s a s b e h (2024). Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization. Journal of Mobile Multimedia, (), 1089-1116. https://doi.org/10.13052/jmm1550-4646.2055

Dloifur Rohman, A l g h i f a r i, Mohammad, E d i, & Lutfi, F i r m a n s y a h (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. Jurnal Manajemen Informatika (JAMIKA), 12(2), 89-99. https://doi.org/10.34010/jamika.v12i2.7764

Mohammad Eid, A l z a h r a n i, Theyazn H. H., A l d h y a n i, Saleh Nagi, A l s u b a r i, Maha M., A l t h o b a i t i, & Adil, F a h a d (2022). Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews. Computational Intelligence and Neuroscience, 2022(), 1-10. https://doi.org/10.1155/2022/3840071

Muhaimin, B a h r e i n, & Suhendro, A p a n d i (2025). Perbandingan Kinerja Naive Bayes, Support Vector Machine, Regresi Logistik, dan Decision Tree untuk Klasifikasi Sentimen Ulasan Produk Berbasis TF-IDF. Techno.Com, 24(4), 1067-1078. https://doi.org/10.62411/tc.v24i4.13968

Pawanjit Singh, G h a t o r a, Seyed Ebrahim, H o s s e i n i, Shahbaz, P e r v e z, Muhammad, I q b a l, & Nabil, S h a u k a t (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM. Big Data and Cognitive Computing, 8(12), 199. https://doi.org/10.3390/bdcc8120199

Hendriyana, H e n d r i y a n a, Ichwanul Muslim Karo, K a r o, & Sri Murni, D e w i (2022). Analisis perbandingan Algoritma Support Vector Machine, Naive Bayes dan Regresi Logistik untuk Memprediksi Donor Darah. Jurnal Teknologi Terpadu, 8(2), 121-126. https://doi.org/10.54914/jtt.v8i2.581

Huang, H u a n g, Adeleh, A s e m i, & Mumtaz Begum, M u s t a f a (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. IEEE Access, 11(), 90367-90382. https://doi.org/10.1109/access.2023.3307308

Talha Ahmed, K h a n, Rehan, S a d i q, Zeeshan, S h a h i d, Muhammad Mansoor, A l a m, & Mazliham Mohd, S u ' u d (2024). Sentiment Analysis using Support Vector Machine and Random Forest. Journal of Informatics and Web Engineering, 3(1), 67. https://doi.org/10.33093/jiwe.2024.3.1.5

Josua Josen Alexander, L i m b o n g, Irwan, S e m b i r i n g, & Kristoko Dwi, H a r t o m o (2022). Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor. Jurnal Teknologi Informasi dan Ilmu Komputer, 9(2), 347. https://doi.org/10.25126/jtiik.2022924960

Bagas Akbar, M a u l a n a, Muhammad Jazilul, F a h m i, Ari Muhamad, I m r a n, & Nutriana, H i d a y a t i (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). MALCOM Indonesian Journal of Machine Learning and Computer Science, 4(2), 375-384. https://doi.org/10.57152/malcom.v4i2.1206

Firda, M i l l e n n i a n i t a, Ummi, A t h i y a h, & Arif Wirawan, M u h a m m a d (2024). Comparison of Naïve Bayes Classifier and Support Vector Machine Methods for Sentiment Classification of Responses to Bullying Cases on Twitter. Journal of Mechatronics and Artificial Intelligence, 1(1), 11-26. https://doi.org/10.17509/jmai.v1i1.69959

I Wayan, P i n a s t a w a, & Nur Alim, A r i f u d d i n (2023). Komparasi Naïve Bayes dan Support Vector Machine dalam Klasifikasi Jenis Citrus. Techno.Com, 22(2), 409-417. https://doi.org/10.33633/tc.v22i2.7777

Punithavathi, R a s a p p a n, M., P r e m k u m a r, Garima, S i n h a, & Kumar, C h a n d r a s e k a r a n (2024). Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection. Information Processing & Management, 61(3), 103654. https://doi.org/10.1016/j.ipm.2024.103654

Mohd Amiruddin, S a d d a m, Erno Kurniawan, D, & Indra, I n d r a (2023). Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine. Jurnal Informatika Jurnal Pengembangan IT, 8(3), 226-233. https://doi.org/10.30591/jpit.v8i3.4884

Tania Puspa Rahayu, S a n j a y a, Ahmad, F a u z i, & Anis Fitri Nur, M a s r u r i y a h (2023). Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine. INFOTECH Jurnal Informatika & Teknologi, 4(1), 16-26. https://doi.org/10.37373/infotech.v4i1.422

Shakib Sadat, S h a n t o, Zishan, A h m e d, Nisma, H o s s a i n, Auditi, R o y, & Akinul Islam, J o n y (2023). Binary vs. Multiclass Sentiment Classification for Bangla E-commerce Product Reviews: A Comparative Analysis of Machine Learning Models. International Journal of Information Engineering and Electronic Business, 15(6), 48-63. https://doi.org/10.5815/ijieeb.2023.06.04

Upendra, S i n g h, Anant, S a r a s w a t, Hiteshwar Kumar, A z a d, Kumar, A b h i s h e k, & S, S h i t h a r t h (2022). Towards improving e-commerce customer review analysis for sentiment detection. Scientific Reports, 12(1), 21983. https://doi.org/10.1038/s41598-022-26432-3

Dwi Amalia, T a r i g a n, Zeffitni, S i t u m o r a n g, & Risnat, R o s n e l l y (2025). Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM), Naïve Bayes Classifier dan Random Forest. Jurnal Teknologi Informasi dan Ilmu Komputer, 12(3), 587-596. https://doi.org/10.25126/jtiik.2025129608

Thummar, Mansi (2022). Flipkart products review dataset [Dataset]. Kaggle

Roland, V i n c e n t, Iqbal, M a u l a n a, & Oman, K o m a r u d i n (2024). PERBANDINGAN KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN DENGAN MULTICLASS DI TWITTER. JATI (Jurnal Mahasiswa Teknik Informatika), 7(4), 2496-2505. https://doi.org/10.36040/jati.v7i4.7152