Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Perbandingan Naïve Bayes dan Support Vector Machine Berbasis Term Frequency−Inverse Document Frequency pada Analisis Sentimen Ulasan Produk Afiliasi Lintas Platform TikTok dan Shopee

Clara Indriani Putri*, Aditia Yudhistira

Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1,*]clara_indriani_putri@teknokrat.ac.id, [2]aditia_yudhistira@teknokrat.ac.id
Email Penulis Korespondensi: [1]clara_indriani_putri@teknokrat.ac.id
Submitted: **25/02/2026**; Accepted: **19/03/2026**; Published**: 19/03/2026**

Abstrak−Pertumbuhan pemasaran afiliasi pada platform digital, khususnya *TikTok* dan *Shopee*, mendorong meningkatnya volume ulasan konsumen yang dapat dimanfaatkan sebagai sumber wawasan bagi pelaku usaha. Namun, ulasan pada kedua platform memiliki karakteristik bahasa yang berbeda: *Shopee* cenderung lebih repetitif dan transaksional, sedangkan *TikTok* lebih informal, kaya *slang*, dan lebih *noisy*. Perbedaan ini menimbulkan *research gap* karena kinerja model klasifikasi sentimen dapat berubah lintas *platform*, sementara studi komparatif pada ulasan afiliasi lintas *platform* masih terbatas. Penelitian ini bertujuan menganalisis dan membandingkan kinerja *Multinomial Naïve Bayes* dan *Support Vector Machine* dalam mengidentifikasi polaritas sentimen positif dan negatif pada ulasan produk afiliasi *TikTok* dan *Shopee*. Data dikumpulkan melalui web scraping pada periode Desember 2025–Januari 2026 dan menghasilkan 5.502 ulasan mentah. Setelah prapemrosesan (*case folding*, pembersihan berbasis *regex*, normalisasi, penghapusan *stopword*, dan *stemming* dengan *Sastrawi*), diperoleh 4.593 ulasan bersih. Pelabelan otomatis berbasis *leksikon* dengan penanganan negasi menghasilkan dataset biner sebanyak 3.314 ulasan (2.729 positif dan 585 negatif), yang menunjukkan ketidakseimbangan kelas; pada penelitian ini tidak dilakukan penyeimbangan data, sehingga evaluasi tidak hanya menggunakan akurasi, tetapi juga presisi, recall, dan F1-score. Representasi fitur menggunakan *Term Frequency–Inverse Document Frequency*, kemudian data dibagi dengan skema *hold-out* 80:20 (2.651 data latih dan 663 data uji). Hasil pengujian menunjukkan *Support Vector Machine* memberikan performa lebih tinggi (akurasi 95,93%; F1 negatif 0,81) dibanding *Multinomial Naïve Bayes* (akurasi 89,14%; F1 negatif 0,12). Keunggulan *Support Vector Machine* berkaitan dengan kemampuannya membangun *hyperplane* bermargin maksimum pada ruang fitur *Term Frequency–Inverse Document Frequency* yang berdimensi tinggi dan *sparse*, sehingga lebih *robust* terhadap variasi bahasa dan *noise* dibanding pendekatan probabilistik *Naïve Bayes* yang sensitif pada dominasi kelas mayoritas.

Kata Kunci: Analisis Sentimen; Pemasaran Afiliasi; Term Frequency–Inverse Document Frequency; Multinomial Naïve Bayes; Support Vector Machine

Abstract−The growth of affiliate marketing on digital platforms, particularly TikTok and Shopee, has led to a rapid increase in consumer reviews that can be leveraged as actionable insights for businesses. However, reviews across platforms exhibit different linguistic characteristics: Shopee reviews tend to be more repetitive and transactional, whereas TikTok reviews are more informal, rich in slang, and noisier. This difference creates a research gap because sentiment classification performance may vary across platforms, while comparative studies on cross-platform affiliate reviews remain limited. This study aims to analyze and compare the performance of Multinomial Naïve Bayes and Support Vector Machine in identifying positive and negative sentiment polarity in TikTok and Shopee affiliate product reviews. Data were collected via web scraping during December 2025–January 2026, yielding 5,502 raw reviews. After text preprocessing (case folding, regex-based cleaning, normalization, stopword removal, and stemming using Sastrawi), 4,593 clean reviews were obtained. Lexicon-based automatic labeling with negation handling produced a binary dataset of 3,314 reviews (2,729 positive and 585 negative), indicating class imbalance; therefore, no data balancing was applied and evaluation emphasized precision, recall, and F1-score in addition to accuracy. Feature representation used Term Frequency–Inverse Document Frequency, and the dataset was split using an 80:20 hold-out scheme (2,651 training and 663 testing instances). Experimental results show that the Support Vector Machine achieved higher performance (95.93% accuracy; 0.81 negative-class F1) than Multinomial Naïve Bayes (89.14% accuracy; 0.12 negative-class F1). This superiority is related to the ability of Support Vector Machine to learn a maximum-margin hyperplane in the high-dimensional and sparse Term Frequency–Inverse Document Frequency feature space, making it more robust to linguistic variation and noise than the probabilistic Naïve Bayes approach, which is more sensitive to majority-class dominance.

Keywords: Sentiment Analysis; Affiliate Marketing; Term Frequency–Inverse Document Frequency; Multinomial Naïve Bayes; Support Vector Machine

1. PENDAHULUAN

Transformasi digital telah menggeser praktik perdagangan dari *e-commerce* konvensional menuju *social commerce*, yaitu aktivitas belanja daring yang semakin terintegrasi dengan interaksi sosial1 pembelian pada konteks *TikTok Shop*, karena konsumen memanfaatkan ulasan sebagai referensi kualitas sekaligus sebagai penguat keyakinan sebelum transaksi [2]. Temuan lain juga menunjukkan bahwa *affiliate marketing* berpengaruh positif dan signifikan terhadap minat beli, sehingga kualitas konten promosi afiliasi dan persepsi kredibilitas sumber informasi menjadi faktor yang relevan dalam proses keputusan pembelian [3]. Konsekuensinya, volume ulasan dan komentar terkait produk afiliasi meningkat cepat, tetapi kualitas informasi yang tersebar beragam mulai dari ulasan informatif hingga teks singkat yang ambigu dan sarat bahasa gaul.

Masalah yang muncul kemudian adalah (i) tingginya volume ulasan yang menyebabkan analisis manual tidak efisien, (ii) variasi gaya bahasa (singkatan, kata tidak baku, *emoji*), serta (iii) keterbatasan manusia dalam menjaga

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

konsistensi penilaian sentimen. Untuk menjawab tantangan tersebut, analisis sentimen digunakan sebagai pendekatan komputasional untuk mengidentifikasi polaritas opini dari teks dan mengelompokkannya ke kelas tertentu. Studi tinjauan menegaskan bahwa analisis sentimen dan deteksi emosi dari teks menjadi bidang yang berkembang pesat dengan beragam pendekatan pemodelan dan tantangan pada data teks yang *noisy* dan informal [4]. Tinjauan tersier atas berbagai studi sistematis di bidang analisis sentimen juga menunjukkan bahwa penelitian pada area ini sangat luas, namun hasil performa model sering kali sangat bergantung pada karakteristik domain, kualitas prapemrosesan, serta strategi pelabelan data [5].

Dalam ranah klasifikasi teks, perkembangan *deep learning* memberikan peningkatan performa pada banyak tugas, tetapi model klasik masih relevan untuk kondisi *dataset* terbatas, kebutuhan interpretabilitas, serta komputasi yang efisien. Kajian komprehensif tentang klasifikasi teks menekankan bahwa performa model sangat dipengaruhi oleh representasi fitur dan karakteristik data, sehingga pemilihan pendekatan perlu disesuaikan dengan konteks dan tujuan penelitian [6]. Pada banyak riset analisis sentimen berbasis ulasan, dua algoritma klasik yang sering digunakan untuk baseline pembanding adalah *Naïve Bayes* dan *Support Vector Machine* (SVM), terutama ketika teks direpresentasikan menggunakan fitur berbasis Bag-of-Words atau pembobotan seperti TF-IDF [7].

Secara empiris, perbandingan SVM dan *Naïve Bayes* pada data ulasan aplikasi *TikTok* di Indonesia menunjukkan bahwa SVM cenderung unggul pada metrik tertentu dan lebih stabil untuk pemisahan kelas pada ruang fitur berdimensi tinggi [8]. Studi lain yang membandingkan *Naïve Bayes* dan SVM pada data ulasan *TikTok* juga melaporkan bahwa SVM mencapai akurasi lebih tinggi dibandingkan Naïve *Bayes* pada dataset yang diuji [9]. Walaupun demikian, penelitian-penelitian tersebut umumnya berfokus pada ulasan aplikasi/isu tertentu dan belum secara spesifik menelaah ulasan produk yang dipromosikan melalui skema afiliasi lintas platform.

Meskipun perbandingan Naïve Bayes dan Support Vector Machine (SVM) telah banyak dilakukan, sebagian besar studi terdahulu berfokus pada ulasan aplikasi atau satu sumber data. Dalam konteks pemasaran afiliasi, karakter teks TikTok yang cenderung informal dan ekspresif berbeda dari Shopee yang lebih deskriptif dan transaksional, sehingga distribusi fitur dan performa model berpotensi tidak stabil lintas platform. Oleh karena itu, gap penelitian ini meliputi: (1) terbatasnya kajian yang mengevaluasi robustness model klasifikasi sentimen pada ulasan produk afiliasi lintas platform (TikTok vs Shopee) dalam bahasa Indonesia, serta (2) minimnya analisis kesalahan klasifikasi berbasis fenomena linguistik dan implikasinya terhadap deteksi keluhan (kelas Negatif) sebagai kelas minoritas bernilai operasional tinggi.

Dari sisi representasi fitur, *Term Frequency–Inverse Document Frequency* (TF-IDF) banyak dipakai karena sederhana, efektif, dan sesuai untuk data teks berdimensi tinggi. Pada konteks ulasan *marketplace* berbahasa Indonesia, evaluasi TF-IDF menunjukkan bahwa vektorisasi TF-IDF dapat memberikan performa kompetitif pada beberapa algoritma (termasuk Naïve Bayes dan SVM), meskipun masing-masing model memiliki sensitivitas berbeda terhadap parameter dan distribusi data [10]. Pada data ulasan produk *Shopee*, studi perbandingan TF-IDF dan TF-RF juga menunjukkan bahwa kombinasi TF-IDF dengan SVM dapat menghasilkan akurasi tinggi pada klasifikasi sentimen, sehingga pemilihan skema pembobotan fitur menjadi komponen penting pada pipeline analisis sentimen [11]. Selain itu, penelitian yang membandingkan teknik ekstraksi fitur (TF-IDF, BoW, Word2Vec) pada sistem analisis sentimen berbasis SVM memperlihatkan bahwa TF-IDF sering memberikan hasil yang paling konsisten pada data teks tertentu [12].

Kendala lain yang umum ditemui adalah ketersediaan data berlabel. Pelabelan manual membutuhkan biaya dan waktu, sementara data ulasan berkembang cepat. Karena itu, sebagian penelitian menggunakan pelabelan otomatis atau pendekatan *weak supervision* berbasis *leksikon* untuk mempercepat penyediaan label awal, lalu mengevaluasi kinerja model *machine learning* pada data tersebut. Contohnya, studi pelabelan lemah pada ulasan berbahasa Indonesia menggunakan *leksikon* positif–negatif dan menunjukkan bahwa *pipeline klasik* berbasis TF-IDF dapat menghasilkan indikator sentimen yang andal pada dataset yang terbatas[13]. Untuk Bahasa Indonesia, pendekatan berbasis leksikon seperti *InSet Lexicon* juga digunakan untuk menghitung skor polaritas dan menentukan kelas sentimen secara otomatis sebelum evaluasi menggunakan *confusion matrix* [14]. Namun, efektivitas pendekatan ini sangat dipengaruhi oleh kualitas prapemrosesan bahasa Indonesia, termasuk *stemming* untuk mengurangi variasi morfologi. Studi perbandingan algoritma *stemming* pada dokumen bahasa Indonesia menunjukkan bahwa *stemming* berbasis *Sastrawi* memiliki kinerja baik dalam mengembalikan kata berimbuhan ke bentuk dasar [15].

Selain variasi morfologi, fenomena linguistik seperti negasi (“tidak”, “bukan”, “ga”) dapat membalik polaritas sentimen dan menurunkan akurasi jika tidak ditangani dengan tepat. Penelitian mengenai pengaruh negasi menegaskan bahwa pemrosesan negasi yang tidak memadai dapat menyebabkan bias dan kesalahan klasifikasi pada deteksi polaritas [16]. Bahkan pada alat analisis sentimen berbasis *leksikon*, kemampuan mendeteksi negasi merupakan faktor krusial dan perlu diuji karena berdampak langsung pada kualitas pelabelan otomatis [17].

Berdasarkan gap tersebut, novelty penelitian ini terletak pada penyusunan dataset ulasan produk afiliasi lintas platform (TikTok dan Shopee) serta evaluasi komparatif dua baseline utama, yaitu Multinomial Naïve Bayes dan Support Vector Machine (kernel linear), dengan penekanan pada kinerja kelas Negatif sebagai kelas minoritas yang bernilai operasional tinggi. Selain itu, penelitian ini melengkapi evaluasi kuantitatif dengan analisis kesalahan berbasis fenomena bahasa (negasi, campuran sentimen, dan sarkasme) untuk menjelaskan sumber kegagalan model.

Penelitian ini bertujuan membandingkan kinerja kedua model dalam mengklasifikasikan sentimen ulasan produk pada konteks TikTok Affiliate dan Shopee Affiliate. Kontribusi penelitian meliputi: (i) penyusunan pipeline analisis sentimen yang replikatif untuk ulasan afiliasi lintas platform, (ii) evaluasi performa model menggunakan

This Journal is licensed under a Creative Commons Attribution 4.0 International License

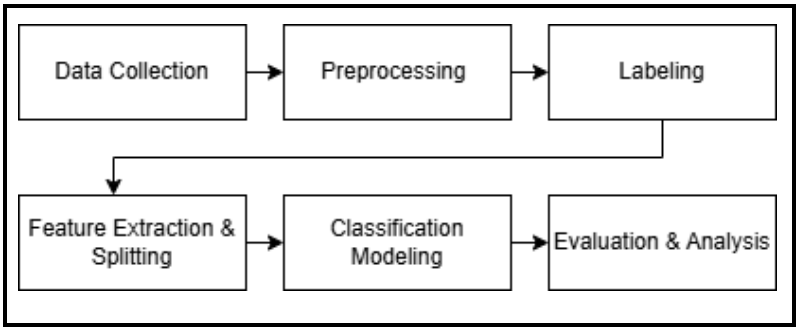
Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

metrik accuracy, precision, recall, dan *F1-score* (terutama pada kelas Negatif), dan (iii) rekomendasi pemodelan yang lebih sesuai untuk kebutuhan pemantauan keluhan pada ekosistem e-commerce di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan desain eksperimental untuk membandingkan performa dua model klasifikasi sentimen, yaitu *Multinomial Naïve Bayes* dan *Support Vector Machine* (SVM) *kernel linear*. Secara umum, alur kerja analisis sentimen meliputi pengumpulan data, prapemrosesan, pelabelan, ekstraksi fitur, pelatihan model, dan evaluasi. Alur ini sejalan dengan praktik umum pada penelitian analisis sentimen berbasis pembelajaran mesin, termasuk studi yang mengevaluasi ekstraksi fitur TF-IDF dan algoritma klasifikasi pada data berbahasa Indonesia [18] serta praktik implementasi proyek *machine learning end-to-end* yang menekankan pemisahan data latih/uji dan evaluasi berbasis metrik yang jelas [19]. Kerangka metodologi penelitian dirangkum pada Gambar 1, yang menunjukkan alur proses mulai dari pengambilan data hingga evaluasi model.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian diawali dengan pengumpulan data ulasan dari *TikTok* dan *Shopee*, dilanjutkan prapemrosesan untuk mengurangi *noise* dan menyeragamkan teks. Tahap berikutnya adalah pelabelan sentimen untuk membentuk dataset berlabel, kemudian ekstraksi fitur TF-IDF dan pembagian data latih/uji sebagai dasar pemodelan. Selanjutnya model *Multinomial Naïve Bayes* dan *Support Vector Machine* dilatih pada data latih, lalu dievaluasi pada data uji menggunakan *confusion matrix* dan metrik turunan untuk menilai performa secara komprehensif.

2.2 Pengumpulan Data *(Data Collection)*

Data yang digunakan berupa ulasan/komentar konsumen tentang produk yang dipromosikan melalui skema afiliasi pada dua platform, yaitu *TikTok* (*TikTok Affiliate*) dan *Shopee* (*Shopee Affiliate*). Mengingat adanya pembatasan akses dan ketatnya keamanan jaringan pada *Application Programming Interface* (API) kedua platform tersebut. Pengambilan data dilakukan melalui teknik *web scraping* menggunakan *ekstensi* peramban *Instant Data Scraper* untuk mengekstraksi elemen teks ulasan dan metadata dasar (misalnya nama akun, waktu, dan isi ulasan) ke format tabel (CSV/Excel). Pada platform *Shopee Affiliate*, data diekstraksi dari kolom ulasan produk yang ditautkan oleh akun afiliator. Sementara itu, pada platform *TikTok Affiliate*, pengumpulan data dilakukan melalui dua sumber spesifik, yaitu kolom komentar pada video konten afiliasi serta ulasan produk yang tercantum pada fitur "keranjang kuning" (*yellow basket*). Periode pengambilan data pada naskah adalah Desember 2025 sampai Januari 2026 dengan total data mentah sebanyak 5.502 ulasan.

Setelah proses *scraping*, dilakukan kontrol kualitas data untuk menjaga integritas evaluasi. Tahap ini mencakup penghapusan entri tidak valid (misalnya ulasan kosong/berisi simbol-emoji yang menjadi kosong setelah *cleaning*) serta deduplikasi berbasis teks bersih (clean_text) agar ulasan identik tidak dihitung berulang pada pelatihan maupun pengujian. Penelitian ini belum menerapkan deteksi spam berbasis pola yang lebih kompleks, sehingga masih dimungkinkan terdapat komentar non-informatif yang lolos apabila bentuknya tidak sama persis.

2.3 Prapemrosesan Data *(Preprocessing)*

Data mentah yang terkumpul dari kedua platform digabungkan menjadi satu dataset tunggal. Prapemrosesan dilakukan untuk mengurangi *noise*, menyeragamkan bentuk teks, dan meningkatkan kualitas fitur sebelum pemodelan. Tahapan yang diterapkan adalah:

a. *Case folding*, yaitu mengubah seluruh huruf menjadi huruf kecil agar konsisten.
b. *Cleaning* berbasis *regular expression* (*regex*) untuk menghapus tautan, *mention*, *hashtag*, angka yang tidak relevan, tanda baca berlebih, dan karakter non-alfabet tertentu.
c. *Normalization*, yaitu mengubah kata tidak baku, singkatan, atau slang menjadi bentuk baku bahasa Indonesia menggunakan kamus normalisasi (*norm_dict*) yang disusun dan dikurasi secara manual berdasarkan temuan pada data *TikTok* dan *Shopee* (*domain-specific*). Normalisasi dilakukan pada tingkat token untuk menyeragamkan

Copyright © 2026 **Author**, Page 2575
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

variasi penulisan, memperbaiki *typo*, serta menyatukan bentuk negasi informal agar tidak membentuk fitur yang berbeda pada proses ekstraksi TF-IDF. Contohnya, "gk/ga/gak" dinormalisasi menjadi "tidak", "brg" menjadi "barang", dan "bgt" menjadi "banget".

d. *Stopword removal*, yaitu menghapus kata umum yang tidak membawa muatan sentimen dominan.
e. *Stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar. Tahap ini penting pada bahasa Indonesia karena variasi imbuhan tinggi; studi evaluasi stemming bahasa Indonesia menunjukkan performa baik pada stemmer Sastrawi dalam mengembalikan bentuk kata dasar [15].

Setelah prapemrosesan, jumlah data bersih menjadi 4.593 ulasan.

2.4 Pelabelan Data *(Labeling)*

Karena data berjumlah besar, pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan berbasis *leksikon* (lexicon-based) dengan perhitungan skor polaritas. Pelabelan otomatis berbasis leksikon mempercepat penyediaan label pada data berukuran besar, namun berpotensi menghasilkan *noise* label terutama pada teks informal (slang), kalimat sarkastik, serta negasi implisit. Untuk memastikan bahwa model supervised tidak hanya "meniru" aturan *leksikon*, dilakukan audit kualitas label menggunakan anotasi manual pada subset data. Sejumlah 400 ulasan dipilih secara acak-*stratified* dari kelas Positif dan Negatif, kemudian dianotasi secara independen oleh dua anotator dengan pedoman label yang sama. Tingkat kesepakatan anotator dihitung menggunakan *Cohen's kappa* sebagai ukuran kesepakatan antar-penilai untuk data kategori nominal. Selain itu, label *leksikon* dibandingkan dengan label manual untuk mengestimasi tingkat kesalahan label (label *noise*). Hasil validasi ini digunakan untuk (1) melaporkan *reliabilitas* label, dan (2) menafsirkan hasil akurasi model dengan lebih hati-hati sebagai estimasi performa terhadap label yang tervalidasi. Secara ringkas, setiap ulasan dihitung skor totalnya berdasarkan bobot kata positif/negatif, lalu ditentukan label: Positif, Negatif, atau Netral. Praktik pelabelan otomatis berbasis *leksikon* untuk Bahasa Indonesia (misalnya menggunakan *InSet Lexicon*) telah digunakan pada penelitian analisis sentimen di *Twitter* dan dievaluasi menggunakan *confusion matrix* [14]. Untuk mengurangi kesalahan polaritas akibat negasi, algoritma pelabelan menerapkan aturan negation handling, misalnya membalik/menurunkan kontribusi kata positif ketika didahului kata negasi. Negasi diketahui berdampak signifikan pada deteksi polaritas dan dapat menurunkan akurasi jika tidak ditangani dengan tepat [16], serta menjadi faktor penting pada evaluasi alat analisis sentimen berbasis *leksikon* [17]. Pada tahap akhir pelabelan, kelas Netral dieliminasi untuk memfokuskan penelitian pada klasifikasi biner (Positif vs Negatif). Data akhir untuk klasifikasi biner berjumlah 3.314 ulasan.

2.5 Ekstraksi Fitur dan Pembagian Data *(Feature Extraction & Splitting)*

Data teks yang telah dipraproses ditransformasikan menjadi fitur numerik menggunakan pembobotan TF-IDF. Implementasi TF-IDF dilakukan menggunakan pustaka *scikit-learn* melalui *TfidfVectorizer*, yang mengubah koleksi dokumen teks menjadi matriks fitur TF-IDF. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) untuk mengevaluasi kemampuan generalisasi model.

2.6 Klasifikasi Model *(Classification Modeling)*

Pengembangan model klasifikasi sentimen menggunakan pendekatan pembelajaran mesin terawasi (*supervised learning*). Penelitian ini mengimplementasikan dan mengkomparasi kinerja dua algoritma, yaitu *Multinomial Naïve Bayes* dan *Support Vector Machine* (SVM), guna menentukan metode yang paling optimal dalam mengklasifikasikan ulasan produk.

a. *Multinomial Naïve Bayes*

Model ini cocok untuk data teks yang direpresentasikan sebagai fitur diskret seperti frekuensi kata (dan dalam praktik dapat bekerja juga pada fitur TF-IDF). Prinsipnya memilih kelas dengan probabilitas posterior terbesar. Secara umum, aturan keputusan dapat dituliskan sebagai:

$$\hat{y} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{i=1}^{n} P(x_i \mid c) \quad (1)$$

dengan ĉ adalah kelas prediksi, C adalah himpunan kelas, dan t_i adalah fitur kata ke-i [20].

b. Support Vector Machine (kernel linear)

SVM bertujuan mencari hyperplane pemisah terbaik pada ruang fitur berdimensi tinggi. Dengan kernel linear, fungsi keputusan dapat dinyatakan sebagai:

$$f(x) = w.x + b \quad (2)$$

dengan w adalah vektor bobot, x adalah vektor fitur TF-IDF, dan b adalah bias. Kelas ditentukan berdasarkan tanda f(x) [19].

2.7 Evaluasi dan Analisis *(Evaluation and Analysis)*

Kinerja kedua model dievaluasi menggunakan *confusion matrix*. Tabel 1 menampilkan struktur *confusion matrix* klasifikasi *biner* untuk memetakan label aktual dan prediksi, sehingga kesalahan model pada tiap kelas dapat diidentifikasi.

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Tabel 1. *Confusion matrix* untuk klasifikasi biner

Prediksi	Positif	Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

Berdasarkan Tabel 1, TP dan TN menunjukkan prediksi yang benar, sedangkan FP dan FN menunjukkan kesalahan prediksi. FP terjadi ketika data aktual Negatif diprediksi sebagai Positif, sedangkan FN terjadi ketika data aktual Positif diprediksi sebagai Negatif. Informasi ini menjadi dasar perhitungan metrik evaluasi untuk menilai kinerja model secara keseluruhan maupun per kelas.

Metrik accuracy, precision, recall, dan F1-score dihitung menggunakan Persamaan (3)–(6). Akurasi mengukur ketepatan prediksi secara keseluruhan, tetapi pada data tidak seimbang dapat tampak tinggi karena dominasi kelas mayoritas. Oleh karena itu, metrik per kelas—terutama recall dan F1-score pada kelas Negatif—dilaporkan untuk menilai kemampuan model mendeteksi keluhan secara lebih adil.

Rumus yang digunakan adalah:

a. Akurasi (*Accuracy*): Mengukur rasio ketepatan prediksi secara keseluruhan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

b. Presisi (*Precision*): Mengukur tingkat ketepatan prediksi untuk suatu kelas, yaitu proporsi prediksi kelas tersebut yang benar.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

c. *Recall*: Mengukur kemampuan model menemukan kembali seluruh data suatu kelas, yaitu proporsi data kelas tersebut yang berhasil dikenali dengan benar.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

d. *F1-Score*: Merupakan rata-rata harmonik (*harmonic mean*) antara presisi dan *recall*, yang menjadi acuan utama jika terdapat ketidakseimbangan jumlah data antar kelas.

$$F\text{-}1\ Score = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall} \quad (6)$$

Hasil metrik tersebut kemudian digunakan untuk membandingkan performa model dan menilai model yang paling sesuai untuk klasifikasi sentimen pada ulasan produk *TikTok* Affiliate dan *Shopee* Affiliate.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data *(Data Collection)*

Pengumpulan data dilakukan pada periode Desember 2025–Januari 2026 melalui teknik *web scraping* menggunakan *Instant Data Scraper*. Kolom waktu pada ulasan merepresentasikan tanggal publikasi oleh pengguna dan dapat berada di luar periode scraping, bergantung pada riwayat ulasan yang ditampilkan pada halaman produk. Proses ini menghasilkan 5.502 ulasan mentah (*raw data)* dari dua platform afiliasi, yaitu Shopee Affiliate terdiri dari 3.075 ulasan dan 2.427 ulasan dari TikTok Affiliate, yang disimpan dalam format .xlsx sebelum pemrosesan lanjutan.

Secara kualitatif, karakteristik ulasan kedua platform berbeda: ulasan Shopee cenderung lebih deskriptif dan transaksional namun masih mengandung singkatan/typo, sedangkan ulasan TikTok lebih informal, ekspresif, dan kaya slang serta emoji. Kondisi ini menunjukkan data mentah masih noisy sehingga memerlukan prapemrosesan sebelum pemodelan. Sebagai ilustrasi, Tabel 2 menyajikan contoh ulasan mentah dari Shopee Affiliate untuk menunjukkan variasi ejaan, singkatan, dan potensi typo yang muncul pada data sebelum prapemrosesan..

Tabel 2. Sampel data mentah *Shopee Affiliate*

User	*Time*	*Review (Teks Asli)*
u*****h	2025-06-20 20:07 | Variasi: DAY & NIGHT RENEW	Alhamdulillah aku meraca cocok dgn pruduk ini
y*****_	2025-06-26 07:57 | Variasi: NIGHT SET	Checkout jam 00.00 tgl 25. Jam 11an tanggal 25 udsudah sampai. Pengitiman dari Tangerang ke Jawa Timur cuma sekejap, pake ekspedisi apa? Salut bgtt, makasih

Berdasarkan Tabel 2, ulasan Shopee cenderung bersifat deskriptif dan transaksional (misalnya menyebut pengiriman, variasi produk, dan pengalaman pembelian), namun masih mengandung variasi ejaan dan singkatan seperti “dgn”, “udh”, serta typo seperti “pruduk/pengitiman”. Karakteristik ini menunjukkan kebutuhan normalisasi dan stemming agar kata yang semakna tidak terpecah menjadi fitur berbeda pada tahap TF-IDF.

Selanjutnya, Tabel 3 menampilkan sampel ulasan mentah dari TikTok Affiliate untuk memperlihatkan karakteristik bahasa yang lebih informal dan ekspresif pada platform tersebut.

Copyright © 2026 **Author**, Page 2577
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

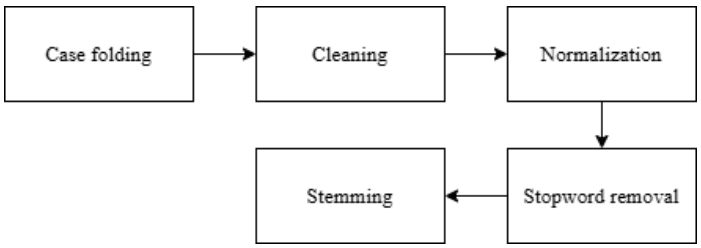
Tabel 3. Sampel data mentah *TikTok Affiliate*

User	*Time*	*Review (Teks Asli)*
aidaaa	05/12/2025 05:11:31	LUCUUUU BANGETTTT😭😭😭
mecca	08/11/2025 03:55:10	mending beli kipas biasa lah, klo rusak ga nyeseg nyeseg amattt, punyaku baru ada 1tahun udh rusak,mana waktu beli fomoo beli 350rb😭😭

Berdasarkan Tabel 3, ulasan TikTok lebih konversasional dan emosional, ditandai penggunaan slang, repetisi huruf (“BANGETTTT”), serta emoji secara intens. Variasi ini meningkatkan noise dan dapat menurunkan performa model berbasis frekuensi jika tidak ditangani melalui cleaning, normalisasi, dan pengurangan repetisi huruf pada tahap prapemrosesan.

3.2 Prapemrosesan Data *(Preprocessing)*

Prapemrosesan dilakukan untuk mengurangi *noise* pada ulasan e-commerce (misalnya emoji, URL, tag, simbol berlebih) serta menyeragamkan variasi penulisan (*slang/typo*) agar teks lebih stabil untuk diekstraksi menjadi fitur TF-IDF. Tahapan diterapkan secara berurutan sesuai metodologi, yaitu *case folding*, pembersihan berbasis *regex (cleaning),* normalisasi (*normalization*), *stopword removal*, dan *stemming* menggunakan *Sastrawi.* Untuk menjelaskan urutan tahap prapemrosesan, alur lengkap proses pembersihan dan penyeragaman teks disajikan pada Gambar 2.



Gambar 2. Alur Prapemrosesan Data

Berdasarkan Gambar 2, prapemrosesan dilakukan secara berurutan mulai dari *case folding* dan *cleaning* untuk mengurangi *noise* permukaan (emoji, URL, simbol), kemudian normalisasi untuk menyeragamkan *slang/typo*, dilanjutkan *stopword removal* dan *stemming* untuk mengurangi variasi morfologis. Urutan ini penting karena normalisasi lebih efektif diterapkan setelah teks dibersihkan, sedangkan stemming membantu menyatukan bentuk kata agar sparsity pada fitur TF-IDF menurun.

a. *Case folding* mengubah seluruh teks menjadi huruf kecil (*lowercase*) agar perbedaan kapitalisasi tidak menghasilkan token yang berbeda. Contoh: “JANGAN BELI DISINI!!” → “jangan beli disini”.
b. *Cleaning* berbasis *regex* menghapus komponen non-informatif dan menyisakan karakter alfabet. Proses ini mencakup penghapusan URL, mention *(@username*), *hashtag* (#tag), angka, simbol/emoji, reduksi huruf berulang (misalnya “baaaagus” → “bagus”), serta normalisasi spasi. Contoh: “Mantappp 😍😍 https://t.co/xxx #murah 100% ori!!!” → “mantap murah ori”.
c. Normalisasi (*normalization*) menyatukan slang, singkatan, dan typo menjadi bentuk baku menggunakan kamus normalisasi (norm_dict). Tahap ini penting untuk menyeragamkan variasi seperti “brg/bgs/bgt” dan negasi informal “gk/ga/gak/engga”. Contoh: “brg nya bgs bgt” → “barang nya bagus banget”, dan “g2g” → “glad2glow” (diperbaiki sebelum penghapusan angka agar token merek tidak hilang).
d. *Stopword removal* menghapus kata fungsi yang minim kontribusi terhadap sentimen berdasarkan daftar *stopword final* (final_stopwords), sehingga token yang tersisa lebih berfokus pada kata bermakna. Contoh: “barang nya bagus banget suka deh” → “barang bagus banget suka”.
e. *Stemming* menggunakan *Sastrawi* mengubah kata berimbuhan menjadi kata dasar untuk menyeragamkan variasi morfologis dan mengurangi sparsity pada fitur TF-IDF. Contoh: “pengiriman” → “kirim”, “membanggakan” → “bangga”, “mengecewakan” → “kecewa”.

Setelah seluruh tahapan prapemrosesan, dilakukan pembersihan data tingkat dokumen dengan menghapus entri yang menjadi kosong (misalnya ulasan yang hanya berisi emoji/simbol sehingga hilang setelah cleaning) serta menghapus duplikasi berdasarkan teks bersih. Dampaknya, jumlah data berkurang dari 5.502 menjadi 4.593 ulasan bersih, sehingga dataset lebih seragam dan siap untuk ekstraksi fitur serta pelatihan model. Untuk memperlihatkan hasil setiap tahap prapemrosesan terhadap perubahan bentuk teks, contoh transformasi input–output pada tiap tahapan dirangkum pada Tabel 4.

Tabel 4. Contoh transformasi teks pada tahap pra-pemrosesan

Tahapan	Tujuan/Aturan Utama	Contoh *input* → *output* (ringkas)
Data mentah (Raw)	Teks asli sebelum diproses	"Brg nya bgs bgt!!! Suka deh 😍👍" → (belum diproses)

Copyright © 2026 **Author**, Page 2578
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Tahapan	Tujuan/Aturan Utama	Contoh *input* → *output* (ringkas)
Case folding	Ubah huruf menjadi lowercase agar konsisten	"JANGAN BELI DISINI!! PALSU & RUSAK..." → "jangan beli disini!! palsu & rusak..."
Cleaning (regex)	Hapus URL, mention, hashtag; hapus angka/simbol/emoji; reduksi huruf berulang; rapikan spasi	"Mantappp 😍😍 https://t.co/xxx #murah 100% ori!!!" → "mantap murah ori"
Normalization (norm_dict)	Ubah slang/typo/singkatan ke bentuk baku; termasuk merek khusus	"brg nya bgs bgt" → "barang nya bagus banget"
Normalization (khusus token merek)	Perbaiki token merek sebelum angka dihapus	"g2g glow bagus" → "glad2glow glad2glow bagus" (contoh, sesuai mapping g2g/glow → glad2glow)
Stopword removal	Hapus kata fungsi/partikel yang tidak informatif berdasarkan final_stopwords	"barang nya bagus banget suka deh" → "barang bagus banget suka"
Stemming (Sastrawi)	Ubah kata berimbuhan ke kata dasar	"pengiriman membanggakan" → "kirim bangga"

Berdasarkan Tabel 4, *cleaning* menghilangkan komponen non-informatif (URL, simbol, emoji) sehingga teks menjadi lebih bersih, sedangkan normalisasi menyatukan variasi *slang/typo* ("brg/bgs/bgt") agar tidak membentuk fitur yang berbeda pada TF-IDF. *Stopword removal* mempertahankan kata bermakna yang lebih relevan terhadap sentimen, sementara *stemming* menurunkan variasi imbuhan sehingga fitur menjadi lebih kompak. Secara keseluruhan, transformasi pada Tabel 4 menunjukkan bahwa teks menjadi lebih seragam dan siap untuk vektorisasi.

3.3 Pelabelan Data *(Labeling)*

Karena data tidak memiliki label sentimen (*unlabeled*), pelabelan dilakukan secara otomatis menggunakan pendekatan berbasis leksikon dengan penanganan negasi (*negation handling*) [14], [16], [17]. Setiap ulasan ditokenisasi dan setiap token dipetakan ke skor sentimen positif/negatif sesuai kamus. Skor ulasan dihitung sebagai akumulasi skor token, lalu label ditentukan berdasarkan nilai skor total: skor > 0 dikategorikan Positif, skor < 0 dikategorikan Negatif, dan skor = 0 dikategorikan Netral. Penanganan negasi diterapkan untuk menjaga validitas konteks; misalnya pada frasa "tidak bagus", token "tidak" memodifikasi polaritas token "bagus" sehingga hasil akhir tidak salah dibaca sebagai sentimen positif. Untuk memperjelas mekanisme penentuan label berbasis skor *leksikon* dan dampak penanganan negasi, simulasi perhitungan skor polaritas disajikan pada Tabel 5.

Tabel 5. Simulasi perhitungan skor pelabelan (*lexicon scoring*)

No	Kalimat ulasan	Tokenisasi & bobot	Skor total	Label
1	alhamdulillah meraca cocok produk	alhamdulillah(+3), cocok(+4)	7	Positif
2	semoga cocok pakai ini trimakasih	semoga(+2), cocok(+4), trimakasih(+2)	8	Positif
3	bahan tipis panas	tipis(-3), panas(-5)	-9	Negatif

Berdasarkan Tabel 5, label sentimen ditentukan dari akumulasi bobot token: skor > 0 diklasifikasikan sebagai Positif dan skor < 0 sebagai Negatif. Simulasi ini menegaskan bahwa kualitas kamus leksikon dan hasil prapemrosesan token berpengaruh langsung terhadap akurasi label, sehingga diperlukan kontrol kualitas (audit label) untuk meminimalkan noise.

Hasil pelabelan pada 4.593 ulasan menghasilkan tiga kelas (Positif, Negatif, dan Netral). Sebanyak 1.279 ulasan dikategorikan Netral (skor 0) dan dieliminasi untuk memfokuskan penelitian pada klasifikasi biner (Positif vs Negatif). Dengan demikian, dataset biner berjumlah 3.314 ulasan (4.593 − 1.279), dengan distribusi 2.729 ulasan Positif dan 585 ulasan Negatif. Distribusi ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*), karena kelas Negatif jauh lebih sedikit dibanding kelas Positif. Oleh karena itu, evaluasi model tidak hanya mengandalkan akurasi, tetapi juga *precision*, *recall*, dan *F1-score*.

3.4 Ekstraksi Fitur dan Pembagian Data *(Feature Extraction and Splitting)*

Pada tahap ekstraksi fitur, teks ulasan direpresentasikan sebagai vektor numerik menggunakan TF-IDF melalui *TfidfVectorizer* pada *scikit-learn*. Proses ini menghasilkan ruang fitur dengan dimensi 4.325 term unik. Secara konseptual, TF-IDF meningkatkan bobot term yang spesifik terhadap dokumen, dan menurunkan bobot term yang terlalu umum pada korpus. Untuk mengilustrasikan kata-kata yang paling dominan setelah vektorisasi, beberapa term dengan bobot TF-IDF rata-rata tertinggi dirangkum pada Tabel 6.

Tabel 6. Contoh term dengan bobot TF-IDF tertinggi

Peringkat	*Term*	Bobot *TF-IDF*	*Interpretasi*
1	bagus	0.0552	kata sentimen (kualitas)
2	banget	0.0502	penguat intensitas

Copyright © 2026 **Author**, Page 2579
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

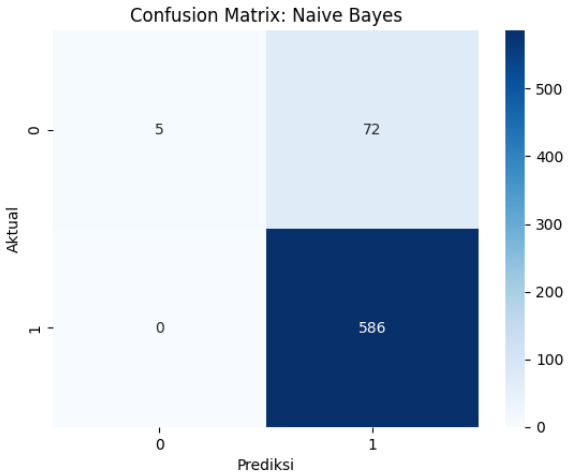
Peringkat	*Term*	Bobot *TF-IDF*	*Interpretasi*
3	cocok	0.0463	sentimen spesifik produk

Berdasarkan Tabel 6, term seperti "bagus", "banget", dan "cocok" memperoleh bobot tinggi karena sering muncul pada dokumen tertentu dan relatif informatif untuk membedakan sentimen. Temuan ini konsisten dengan karakter ulasan e-commerce, di mana kata kualitas dan penguat intensitas menjadi indikator penting polaritas. Dengan demikian, Tabel 6 menunjukkan bahwa TF-IDF menangkap sinyal leksikal yang relevan sebagai fitur masukan untuk model klasifikasi.

Untuk evaluasi, *dataset biner* dibagi menggunakan *skema hold-out* 80:20 menjadi data latih dan data uji. Dengan total 3.314 data, pembagian menghasilkan 2.651 data latih dan 663 data uji. Untuk mencegah data *leakage*, pembentukan kosakata dan pembelajaran bobot IDF pada TF-IDF hanya dilakukan pada data latih (*fit/fit_transform* pada data latih). Data uji tidak pernah digunakan saat proses *fitting* dan hanya ditransformasikan menggunakan parameter dari data latih (*transform*). Implementasi dilakukan menggunakan *Pipeline* agar seluruh tahapan prapemrosesan–ekstraksi fitur–klasifikasi diterapkan secara konsisten tanpa kebocoran informasi dari data uji.

3.5 Klasifikasi Naïve Bayes

Model pertama adalah *Multinomial Naïve Bayes* dengan parameter *default alpha* = 1.0 (*Laplace smoothing*), sesuai implementasi *MultinomialNB* pada *scikit-learn*. Hasil prediksi Multinomial Naïve Bayes pada data uji divisualisasikan melalui confusion matrix pada Gambar 3 untuk memudahkan pembacaan pola kesalahan.



Gambar 3. Confusion matrix Naïve Bayes (0=Negatif, 1=Positif)

Berdasarkan Gambar 3, sebagian besar kesalahan terjadi ketika ulasan Negatif diprediksi sebagai Positif (72 kasus), sehingga kemampuan deteksi keluhan menjadi lemah. Pola ini mengindikasikan bias terhadap kelas mayoritas, yang umum terjadi pada data tidak seimbang ketika model cenderung mengoptimalkan akurasi global. Ringkasan metrik evaluasi *Multinomial Naïve Bayes* disajikan pada Tabel 7 untuk menilai performa secara lebih komprehensif melalui *precision, recall, dan F1-score per kelas.*

Tabel 7. Ringkasan metrik klasifikasi Naïve Bayes (data uji=663)

Kelas	*Precision*	*Recall*	*F1-score*	*Support*
Negatif (0)	1.00	0.06	0.12	77
Positif (1)	0.89	1.00	0.94	586
Accuracy			0.89	663
Macro Avg	0.95	0.53	0.53	663
Weighted Avg	0.90	0.89	0.85	663

Berdasarkan Tabel 7, meskipun akurasi mencapai 89,14%, recall kelas Negatif sangat rendah (0,06) sehingga sebagian besar keluhan tidak terdeteksi. Perbedaan antara macro average dan weighted average menegaskan dampak class imbalance, sehingga evaluasi perlu menekankan metrik per kelas, khususnya recall dan F1-score pada kelas Negatif. Analisis per platform menunjukkan akurasi Naïve Bayes lebih tinggi pada Shopee (94,51%) dibanding TikTok (80,92%), yang mengindikasikan variasi bahasa TikTok yang lebih informal dan noisy menyulitkan model berbasis frekuensi term.

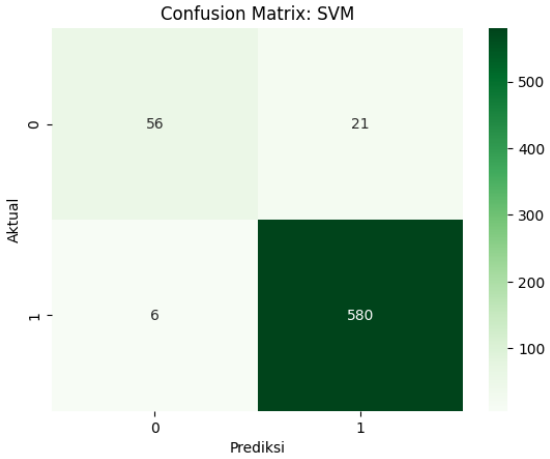
3.6 Klasifikasi SVM (*Linear Kernel*)

Model kedua adalah *Support Vector Machine* menggunakan *kernel linear*. Pemilihan *kernel linear* konsisten dengan penggunaan fitur TF-IDF yang bersifat berdimensi tinggi dan *sparse*, dan *kernel linear* merupakan pilihan yang sah

Copyright © 2026 **Author**, Page 2580
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

pada SVC di *scikit-learn*. Kinerja Support Vector Machine (kernel linear) pada data uji divisualisasikan melalui confusion matrix pada Gambar 4 untuk memperlihatkan distribusi kesalahan prediksi.



Gambar 4. Confusion matrix SVM (kernel linear) (0=Negatif, 1=Positif)

Berdasarkan Gambar 4, SVM mampu mendeteksi kelas Negatif dengan lebih baik (56 benar dari 77), dan jumlah kesalahan Negatif→Positif jauh lebih kecil dibanding *Naïve Bayes*. Hal ini menunjukkan bahwa SVM lebih seimbang dalam menangani data tidak seimbang dan tidak semata-mata "mengikuti" kelas mayoritas.

Ringkasan metrik evaluasi SVM ditampilkan pada Tabel 8 untuk menilai performa per kelas melalui *precision, recall, dan F1-score*.

Tabel 8. Ringkasan metrik klasifikasi SVM (data uji=663)

Kelas	*Precision*	*Recall*	*F1-score*	*Support*
Negatif (0)	0.90	0.73	0.81	77
Positif (1)	0.97	0.99	0.98	586
Accuracy			0.96	663
Macro Avg	0.93	0.86	0.89	663
Weighted Avg	0.96	0.96	0.96	663

Berdasarkan Tabel 8, SVM mencapai akurasi 0,96 dengan recall kelas Negatif 0,73 dan F1 Negatif 0,81, yang menunjukkan kemampuan deteksi keluhan jauh lebih baik dibanding Naïve Bayes. Nilai macro-F1 yang tinggi juga mengindikasikan keseimbangan performa antar kelas, sehingga model lebih layak digunakan pada skenario pemantauan keluhan pelanggan.

SVM mencapai akurasi 95.93% dan menunjukkan kemampuan deteksi kelas Negatif yang jauh lebih baik dibanding *Naïve Bayes* (*recall* Negatif 0.73; F1 Negatif 0.81). Jumlah salah klasifikasi pada SVM adalah 27 data (21 Negatif salah menjadi Positif dan 6 Positif salah menjadi Negatif), jauh lebih kecil dibanding *Naïve Bayes* yang salah pada 72 data. Akurasi per platform juga menunjukkan stabilitas SVM: 96.51% pada *Shopee* dan 95.04% pada *TikTok*. Perbedaan akurasi antar platform relatif kecil, yang menandakan SVM lebih *robust* terhadap variasi bahasa pada TikTok.

3.7 Pembahasan dan Analisis

Perbandingan Naïve Bayes dan SVM menunjukkan bahwa pemilihan model terbaik tidak dapat bertumpu pada akurasi semata, terutama pada kondisi data tidak seimbang. Naïve Bayes mencapai akurasi 89,14%, namun recall kelas Negatif hanya 0,06 sehingga sebagian besar ulasan Negatif gagal terdeteksi dan cenderung diprediksi sebagai Positif. Secara praktis, kondisi ini berisiko pada skenario pemantauan keluhan pelanggan karena keluhan dapat "tersembunyi" di antara dominasi ulasan Positif; hal tersebut juga tercermin dari macro F1 yang rendah (0,53).

Sebaliknya, SVM (kernel linear) mencapai akurasi 95,93% dan memberikan trade-off yang lebih seimbang antara precision dan recall, khususnya pada kelas Negatif (precision 0,90; recall 0,73; F1 0,81). Hasil ini konsisten dengan karakteristik fitur TF-IDF yang menghasilkan ruang fitur berdimensi tinggi dan sparse, di mana model linear berbasis margin umumnya lebih stabil. Sebaliknya, Naïve Bayes mengasumsikan independensi antar fitur dan lebih sensitif terhadap dominasi kelas mayoritas, sehingga performanya dapat menurun pada data yang noisy dan tidak seimbang [20].

Untuk menguji signifikansi perbedaan performa kedua model, dilakukan uji McNemar pada prediksi data uji. Berdasarkan tabel kesalahan silang diperoleh b = 51 (Naïve Bayes salah, SVM benar) dan c = 6 (Naïve Bayes benar, SVM salah). Statistik uji menghasilkan $\chi^2 = 33{,}96$ (df = 1; $p < 0{,}001$), sehingga perbedaan performa dinyatakan signifikan secara statistik. Uji McNemar dihitung menggunakan koreksi kontinuitas (Yates) pada pasangan prediksi yang tidak selaras.

Copyright © 2026 **Author**, Page 2581
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Dari sisi penelitian terkait, temuan ini konsisten dengan studi pada ulasan TikTok yang melaporkan SVM cenderung memberikan performa lebih tinggi dibanding Naïve Bayes pada variasi bahasa yang kuat [8], [9], serta studi ulasan marketplace yang menunjukkan kombinasi TF-IDF dan SVM memberikan kinerja kompetitif untuk klasifikasi sentimen [11], [12]. Oleh karena itu, hasil penelitian ini memperkuat bukti empiris bahwa SVM linear merupakan pilihan yang lebih aman ketika fitur yang digunakan adalah TF-IDF dan data mengandung variasi linguistik yang besar.

Analisis per platform berdasarkan kolom source menunjukkan bahwa akurasi pada Shopee lebih tinggi dibanding TikTok, terutama pada Naïve Bayes (94,51% vs 80,92%). Hal ini sejalan dengan karakteristik teks: Shopee relatif lebih terstruktur dan deskriptif, sedangkan TikTok lebih informal, kaya slang/singkatan, repetisi huruf, dan simbol yang bervariasi. Variasi bahasa yang meningkat membuat Naïve Bayes lebih rentan karena bergantung pada frekuensi term per kelas tanpa mekanisme margin, sementara SVM mempertahankan akurasi tinggi pada kedua platform (96,51% Shopee; 95,04% TikTok), yang mengindikasikan ketahanan yang lebih baik terhadap noise.

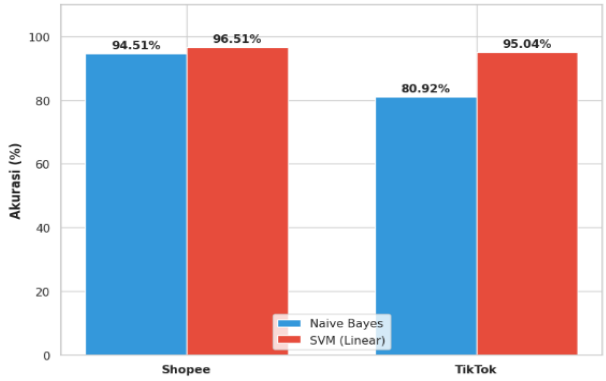
Pada subset data uji, prediksi SVM menunjukkan perbedaan proporsi sentimen antar platform. Ringkasan distribusi prediksi berdasarkan source disajikan pada Tabel 10, di mana Shopee didominasi sentimen Positif (96,01%) sedangkan TikTok memiliki proporsi Negatif lebih tinggi (17,56%). Temuan ini mendukung interpretasi bahwa ekspresi ketidakpuasan pada TikTok lebih eksplisit dibanding Shopee dalam konteks ulasan afiliasi. Namun, karena Tabel 10 dihitung dari prediksi data uji, analisis kecenderungan sentimen platform akan lebih kuat jika dihitung pada seluruh dataset biner menggunakan model terbaik. Untuk membandingkan performa kedua model secara ringkas, metrik utama pada data uji dirangkum pada Tabel 9.

Tabel 9. Ringkasan komparasi performa (data uji)

Model	*Accuracy* (%)	F1 Negatif	F1 Positif	*Macro-F1*	*Weighted-F1*
Naïve Bayes	89.14	0.12	0.94	0.53	0.85
SVM (linear)	95.93	0.81	0.98	0.89	0.96

Berdasarkan Tabel 9, SVM unggul pada seluruh metrik, terutama pada F1 kelas Negatif dan macro-F1, yang relevan pada kondisi class imbalance. Temuan ini menegaskan bahwa peningkatan performa SVM bukan hanya pada akurasi, tetapi juga pada kemampuan membedakan kelas minoritas.

Untuk memudahkan visualisasi perbedaan akurasi kedua model, perbandingan akurasi ditampilkan pada Gambar 5.



Gambar 5. Diagram Batang Perbandingan Akurasi Naïve Bayes dan SVM

Berdasarkan Gambar 5, akurasi SVM lebih tinggi dibanding Naïve Bayes. Visualisasi ini memperkuat kesimpulan bahwa model linear berbasis margin lebih stabil pada fitur TF-IDF yang berdimensi tinggi.

Secara teoritis, hasil ini dapat dijelaskan oleh karakteristik *representasi* fitur TF-IDF yang membentuk ruang fitur berdimensi tinggi dan *sparse*. Dalam konteks klasifikasi teks, *classifier linear* sering menjadi *baseline* yang sangat kuat. Dokumentasi resmi *scikit-learn* menyebutkan bahwa *linear* SVM secara luas dianggap sebagai salah satu algoritma terbaik untuk klasifikasi teks karena kemampuannya membangun *hyperplane* dengan margin maksimum pada ruang berdimensi tinggi. Sebaliknya, *Naïve Bayes* mengasumsikan independensi antar fitur dan lebih sensitif terhadap dominasi kelas mayoritas.

Evaluasi per platform menunjukkan bahwa performa model berbeda pada *Shopee* dan *TikTok*. *Naïve Bayes* mencatat akurasi 94.51% pada *Shopee* tetapi hanya 80.92% pada *TikTok*. Sementara itu, SVM mempertahankan akurasi tinggi pada kedua platform, yaitu 96.51% (*Shopee*) dan 95.04% (*TikTok*). Untuk melihat perbedaan proporsi prediksi sentimen antar platform pada data uji, distribusi prediksi SVM berdasarkan *source* dirangkum pada Tabel 10.

Tabel 10. Perbandingan distribusi prediksi sentimen SVM pada data uji berdasarkan source

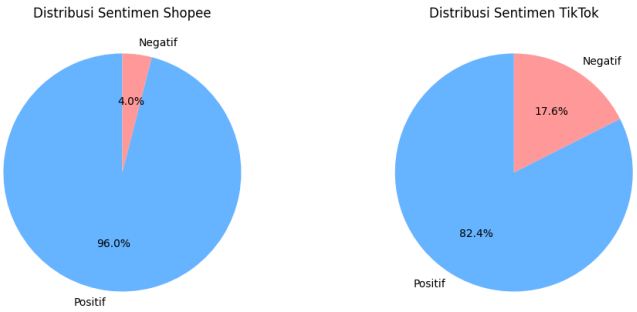
Source	Negatif	Positif	Total	% Positif	% Negatif
Shopee	16	385	401	96.01	3.99
Tiktok	46	216	262	82.44	17.56

Copyright © 2026 **Author**, Page 2582
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Berdasarkan Tabel 10, Shopee didominasi prediksi Positif, sedangkan TikTok menunjukkan proporsi Negatif yang lebih tinggi. Perbedaan ini konsisten dengan karakteristik teks TikTok yang lebih ekspresif dan kritis, meskipun interpretasi kecenderungan sentimen platform tetap bersifat indikatif karena dihitung pada subset data uji.

Untuk memperjelas perbandingan distribusi prediksi sentimen pada kedua platform, visualisasi distribusi ditampilkan pada Gambar 6.



Gambar 6. Perbandingan distribusi prediksi sentimen SVM pada data uji berdasarkan source

Berdasarkan Gambar 6, perbedaan proporsi Negatif pada TikTok terlihat lebih menonjol dibanding Shopee. Visualisasi ini mendukung pembahasan bahwa variasi bahasa dan cara ekspresi ketidakpuasan dapat berbeda antar platform dalam konteks ulasan afiliasi.

Data menunjukkan bahwa *TikTok* memiliki proporsi sentimen negatif lebih tinggi dibandingkan *Shopee*. Hal ini konsisten dengan karakteristik *linguistik TikTok* yang lebih *ekspresif*, informal, dan kritis. Sebaliknya, ulasan *Shopee* lebih terstruktur dan cenderung transaksional.

Untuk memahami karakteristik kata dominan pada masing-masing sentimen dan platform, digunakan *visualisasi Word Cloud* yang terlihat pada Gambar 7.

Gambar 7. Word Cloud Sentimen

Berdasarkan Gambar 7, Visualisasi menunjukkan dominasi kata seperti “bagus”, “cepat”, dan “cocok” pada sentimen positif, sedangkan sentimen negatif didominasi oleh kata seperti “rusak”, “kecewa”, dan “lama”. Perbedaan kosakata antar platform juga terlihat jelas, di mana *TikTok* lebih banyak mengandung *slang* dan pengulangan huruf.

Meskipun SVM (kernel linear) menunjukkan performa terbaik (akurasi 95,93%; macro F1-score 0,89), masih terdapat 27 data uji yang salah diklasifikasikan (14 ulasan *Shopee* dan 13 ulasan *TikTok*). Distribusi kesalahan yang relatif seimbang antar platform mengindikasikan bahwa kesalahan tidak semata-mata dipengaruhi sumber data, melainkan berkaitan dengan keterbatasan representasi Bag-of-Words/TF-IDF dalam menangkap konteks semantik. Mayoritas kesalahan berhubungan dengan tiga pola linguistik utama, yaitu negasi implisit/ganda, campuran sentimen dalam satu kalimat, serta sarkasme/ambiguitas. Pada kasus negasi, pembalikan makna (misalnya kata positif didahului “ga/gak”) tidak selalu tertangkap karena TF-IDF memperlakukan kata sebagai fitur independen. Kesalahan juga muncul pada kalimat dengan pujian dan kritik sekaligus, ketika bobot kata positif dan negatif saling menyeimbangkan sehingga dominasi sentimen sulit ditentukan. Selain itu, bahasa informal, repetisi huruf, dan ekspresi percakapan khas media sosial dapat membuat bobot term tidak sepenuhnya merefleksikan makna yang dimaksud. Ringkasan pola kesalahan tersebut disajikan pada Tabel 11.

Tabel 11. Kategori Penyebab Kesalahan Klasifikasi SVM

Kategori	Karakteristik Kesalahan
Negasi implisit/ganda	Kata bernilai positif dibalik oleh negasi
Campuran sentimen	Pujian dan kritik muncul bersamaan

Copyright © 2026 **Author**, Page 2583
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

Kategori	Karakteristik Kesalahan
Sarkasme/ambiguitas	Struktur kalimat tidak literal

Berdasarkan Tabel 11, kesalahan klasifikasi terutama dipicu oleh negasi implisit/ganda, campuran sentimen dalam satu kalimat, serta sarkasme/ambiguitas. Temuan ini menjelaskan keterbatasan representasi TF-IDF yang tidak memodelkan ketergantungan antar kata secara eksplisit.

Secara keseluruhan, analisis kesalahan ini menunjukkan bahwa keunggulan SVM dibandingkan *Naïve Bayes* tidak serta-merta berarti model telah sepenuhnya memahami kompleksitas linguistik teks media sosial. Peningkatan performa ke depan dapat dipertimbangkan melalui pendekatan yang memperhitungkan dependensi antar kata, seperti penggunaan *n-gram* yang lebih panjang atau model berbasis representasi kontekstual. Namun demikian, dalam konteks penelitian ini, SVM linear tetap menjadi model yang paling stabil dan seimbang untuk klasifikasi sentimen ulasan affiliate lintas platform.

4. KESIMPULAN

Penelitian ini menganalisis sentimen ulasan produk pada konteks *TikTok Affiliate* dan *Shopee Affiliate* dengan membandingkan kinerja dua algoritma klasifikasi berbasis teks, yaitu *Multinomial Naïve Bayes* dan *Support Vector Machin*e (SVM) *kernel linear*. Data diperoleh melalui *web scraping* pada periode Desember 2025–Januari 2026 dengan total 5.502 ulasan mentah, kemudian diprapemroses melalui tahapan *case folding*, *cleaning* berbasis *regex*, normalisasi, *stopword removal*, dan *stemming* sehingga tersisa 4.593 ulasan bersih. Pelabelan otomatis berbasis *leksikon* dengan penanganan negasi menghasilkan tiga kelas sentimen, dan setelah kelas netral dieliminasi diperoleh dataset biner berjumlah 3.314 ulasan dengan distribusi 2.729 ulasan positif dan 585 negatif. Ekstraksi fitur menggunakan TF-IDF menghasilkan 4.325 term unik, kemudian data dibagi dengan skema hold-out 80:20 menjadi 2.651 data latih dan 663 data uji. Hasil evaluasi menunjukkan bahwa SVM memberikan performa terbaik dengan akurasi 95.93%, serta kemampuan mendeteksi kelas negatif yang lebih seimbang (*precision* 0.90; *recall* 0.73; F1 0.81) *dibandingkan Naïve Bayes* yang meskipun mencapai akurasi 89.14%, memiliki *recall* kelas negatif sangat rendah (0.06) sehingga berisiko mengabaikan keluhan pelanggan. Analisis per platform juga menunjukkan bahwa *Shopee* cenderung didominasi sentimen positif, sedangkan *TikTok* memiliki proporsi sentimen negatif lebih tinggi, yang selaras dengan karakteristik bahasa TikTok yang lebih informal dan ekspresif. Keterbatasan penelitian ini terletak pada pelabelan otomatis berbasis *leksikon* yang masih berpotensi menghasilkan *noise label*, ketidakseimbangan kelas yang dapat memengaruhi metrik tertentu, serta keterbatasan TF-IDF dalam menangkap konteks semantik seperti negasi implisit, campuran sentimen, dan sarkasme yang menyebabkan sejumlah kesalahan klasifikasi. Penelitian selanjutnya disarankan untuk mengeksplorasi strategi penyeimbangan kelas, validasi label melalui anotasi manual sebagian data, serta penggunaan fitur berbasis *n-gram* yang lebih kaya atau representasi kontekstual untuk meningkatkan ketahanan model terhadap variasi bahasa media sosial.

REFERENCES

[1] N. I. Prestyasih and S. R. H. Hati, “The Role of Social Commerce Trust and Satisfaction on TikTok Consumer Purchasing Behavior,” *J. Ilm. Manaj. Kesatuan*, vol. 13, no. 4, pp. 2817–2826, Jul. 2025, doi: 10.37641/jimkes.v13i4.3455.

[2] S. Zuhri, N. Nawari, and M. A. Al Mubarok, “Pengaruh Online Customer Review, Affiliate Marketing Terhadap Keputusan Pembelian Di Tiktok Shop,” *J-MACC J. Manag. Account.*, vol. 6, no. 1, pp. 128–140, Apr. 2023, doi: 10.52166/j-macc.v6i1.9651.

[3] S. Brilianita and R. Sulistyowati, “Affiliate Marketing terhadap Minat Beli Mahasiswa di TikTok Shop,” *JPEKA J. Pendidik. Ekon. Manaj. dan Keuang.*, vol. 7, no. 2, 2023, doi: 10.26740/jpeka.v7n2.p157-167.

[4] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 81, Dec. 2021, doi: 10.1007/s13278-021-00776-6.

[5] A. Ligthart, C. Catal, and B. Tekinerdogan, “Systematic reviews in sentiment analysis: a tertiary study,” *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.

[6] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning--based Text Classification,” *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.

[7] A. Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, “Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review,” *Algorithms*, vol. 16, no. 5, p. 236, Apr. 2023, doi: 10.3390/a16050236.

[8] J. O. Leandro and M. I. Fianty, “Evaluation of Sentiment Analysis Methods for Social Media Applications: A Comparison of Support Vector Machines and Naïve Bayes,” *JOIV Int. J. Informatics Vis.*, vol. 9, no. 2, p. 796, Mar. 2025, doi: 10.62527/joiv.9.2.2905.

[9] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.

[10] H. Barus, I. N. Fajri, and Y. Pristyanto, “Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models”, JAIC, vol. 9, no. 5, pp. 2552–2561, Oct. 2025. doi: 10.30871/jaic.v9i5.10524

[11] K. P. Harmandini and K. M. L, “Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment,” *Sinkron*, vol. 8, no. 2, pp. 929–937, Mar. 2024, doi: 10.33395/sinkron.v8i2.13376.

[12] I. F. Rozi, I. Maulidia, M. Hani’ah, R. Arianto, D. R. Yunianto, and A. Y. Ananta, “Comparison of Feature Extraction in Support Vector Machine (SVM) Based Sentiment Analysis System,” *J. Ilm. Kursor*, vol. 13, no. 1, pp. 1–12, Jul. 2025, doi:

Copyright © 2026 **Author**, Page 2584
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2573−2585
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9454

10.21107/kursor.v13i1.417.

[13] Z. Rifa'i and B. P. Mukti, "Weakly Supervised Sentiment Analysis of Indonesian Rural Tourism Reviews: A TF-IDF Baseline for Melung Tourism Village," *Edu Komputika J.*, vol. 12, no. 1, pp. 48–60, 2025, doi: 10.15294/edukom.v12i1.31893.

[14] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, Mar. 2021, doi: 10.57152/malcom.v1i1.20.

[15] D. Mustikasari, I. Widaningrum, R. Arifin, and W. H. E. Putri, "Comparison of Effectiveness of Stemming Algorithms in Indonesian Documents," *in Proc. 2nd Borobudur Int. Symp. Sci. Technol. (BIS-STE 2020), Atlantis Press, 2021, pp. 154–158, doi: 10.2991/aer.k.210810.025.*

[16] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan, and R. Sharma, "Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection," *Procedia Comput. Sci.*, vol. 185, pp. 370–379, 2021, doi: 10.1016/j.procs.2021.05.038.

[17] M. Naldi and S. Petroni, "A Testset-Based Method to Analyse the Negation-Detection Performance of Lexicon-Based Sentiment Analysis Tools," *Computers*, vol. 12, no. 1, p. 18, Jan. 2023, doi: 10.3390/computers12010018.

[18] C. Apriansyah Hutagalung and V. Budi Lestari, "Data Mining Approach: K-Means Clustering and Naïve Bayes Classifier for Graduate Quality Analysis," *J-KOMA J. Ilmu Komput. dan Apl.*, vol. 8, no. 1, pp. 33–42, Jun. 2025, doi: 10.21009/j-koma.v8i1.05.

[19] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, Oct. 2022.

[20] Israt Jahan, Md Nakibul Islam, Md Mahadi Hasan, and Md Rafiuddin Siddiky, "Comparative analysis of machine learning algorithms for sentiment classification in social media text," *World J. Adv. Res. Rev.*, vol. 23, no. 3, pp. 2842–2852, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2983.

This Journal is licensed under a Creative Commons Attribution 4.0 International License