Journal Software, Hardware and Information Technology
p-ISSN: 2808-3385 | e-ISSN: 2776-8961
Volume 6 Issue 2, June 2026

Klasifikasi Sentimen Ulasan Buku Pada Platform Shopee: Studi Komparatif Naive Bayes dan SVM Dengan Pembobotan TF-IDF pada Dataset Seimbang

Sentiment Classification of Book Reviews on the Shopee Platform: A Comparative Study of Naive Bayes and SVM Using TF-IDF Weighting on a Balanced Dataset

Arief Dharmadi, Dwi Hartanti, Anisatul Farida
Jurusan Teknik Informatika, Universitas Duta Bangsa, Surakarta, Indonesia
dharmadi.arief@gmail.com, dwihartanti@udb.ac.id, anisatul_farida@udb.ac.id

Notes

Submitted: 30-04-2026
Revised: 18-05-2026
Accepted: 20-05-2026
Published: 30-06-2026

How to cite: A. Dharmadi, D. Hartanti, and A. Farida, "Klasifikasi Sentimen Ulasan Buku Pada Platform Shopee: Studi Komparatif Naive Bayes dan SVM Dengan Pembobotan TF-IDF pada Dataset Seimbang", Journal Software, Hardware and Information Technology (SHIFT), vol. 6, no. 2, pp. 103–116, Jun. 2026, doi: 10.24252/shift.v6i2.283.

DOI: 10.24252/shift.v6i2.283

Copyright 2026© the Author(s)

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

ABSTRACT

The surge in informal and unstructured product review data in e-commerce has become a major obstacle in manual business evaluation. This study aims to conduct a comparative study of book product review sentiment classification on the Shopee platform using the Multinomial Naive Bayes algorithm and Support Vector Machine (SVM) with a linear kernel. The dataset used consists of 1,101 reviews obtained through web scraping techniques. The sentiment class labeling procedure was carried out manually through cross-validation by three annotators based on Shopee star rating curation and textual content analysis to ensure data validity. After going through the text pre-processing (Sastrawi) stage and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction, the model was tested and its strength was evaluated using the Stratified 10-Fold Cross Validation technique. The test results showed that the SVM algorithm produced slightly superior performance with an accuracy rate of 86.47%, precision of 88.37%, and recall of 84.04%. Meanwhile, Multinomial Naive Bayes achieved 86.29% accuracy, 87.16% precision, and 85.12% recall. In conclusion, although both models demonstrated similar robustness on a balanced dataset, SVM offered a higher level of precision in addressing the specific characteristics of the book review domain, which is characterized by lengthy and informal descriptions in the digital ecosystem.

Keywords: *Book Reviews, E-commerce, Naïve Bayes, Sentiment Analysis, Support Vector Machine, TF-IDF.*

ABSTRAK

Lonjakan data ulasan produk informal dan tidak terstruktur di e-commerce telah menjadi hambatan utama dalam evaluasi bisnis manual. Penelitian ini bertujuan untuk melakukan studi komparatif klasifikasi sentimen ulasan produk buku di platform Shopee menggunakan algoritma Multinomial Naive Bayes dan Support Vector Machine (SVM) dengan kernel linier. Dataset yang digunakan terdiri dari 1.101 ulasan yang diperoleh melalui teknik web scraping. Prosedur pelabelan kelas sentimen dilakukan secara manual melalui validasi silang oleh tiga annotator berdasarkan kurasi peringkat bintang Shopee dan analisis konten tekstual untuk memastikan validitas data. Setelah melalui tahap pra-pemrosesan teks (Sastrawi) dan ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF), model diuji dan kekuatannya dievaluasi menggunakan teknik Stratified 10-Fold Cross Validation. Hasil pengujian menunjukkan bahwa algoritma SVM menghasilkan kinerja yang sedikit lebih unggul dengan tingkat akurasi 86,47%, presisi 88,37%, dan recall 84,04%. Sementara itu, Multinomial Naive Bayes mencapai akurasi 86,29%, presisi 87,16%, dan recall 85,12%. Kesimpulannya, meskipun kedua model menunjukkan kekokohan yang serupa pada dataset yang seimbang, SVM menawarkan tingkat presisi yang lebih tinggi dalam menangani karakteristik spesifik domain ulasan buku, yang berkarakteristik deskripsi yang panjang dan informal dalam ekosistem digital.

Kata Kunci: Analisis Sentimen, E-commerce, Naïve Bayes, Support Vector Machine, TF-IDF, Ulasan Buku.

1. PENDAHULUAN

Transformasi perilaku belanja ke ranah digital menjadikan Shopee sebagai salah satu marketplace dengan volume transaksi terbesar di Indonesia [1], [2]. Dalam ekosistem ini, ulasan pelanggan sangat penting karena berfungsi sebagai referensi pembeli sekaligus umpan balik bagi penjual untuk mengevaluasi kualitas layanan [3]. Berbeda dengan kategori produk elektronik atau pakaian yang cenderung berfokus pada fungsi praktis dan spesifikasi teknis, ulasan pada produk buku memiliki karakteristik linguistik yang unik. Berdasarkan studi literatur terkait analisis sentiment [4], [5], ulasan buku menuntut evaluasi konsumen yang lebih kompleks, di mana pembeli tidak hanya menilai kondisi fisik logistik seperti kualitas cetak dan pengemasan, tetapi juga cenderung menuliskan opini deskriptif yang lebih panjang terkait kualitas isi bacaan.

Karakteristik teks yang panjang dan subjektif ini menjadikan evaluasi kepuasan pelanggan secara manual menjadi tidak efisien.

Seiring dengan meningkatnya volume transaksi, akumulasi data ulasan menjadi sangat masif dan tidak terstruktur [6]. Selain jumlahnya yang besar, tantangan utama dalam memproses teks ulasan e-commerce di Indonesia adalah tingginya penggunaan bahasa tidak baku, istilah slang, singkatan, dan kesalahan penulisan [7]. Tanpa sistem analisis otomatis, ribuan ulasan berharga ini akan terabaikan dan penjual kesulitan memetakan sentimen konsumen [8]. Oleh karena itu, pendekatan Natural Language Processing (NLP) sangat krusial untuk membersihkan noise teks informal tersebut melalui preprocessing, lalu mengklasifikasikan opini ke dalam polaritas positif maupun negatif secara komputasional [9].

Untuk membangun sistem klasifikasi yang akurat, Naive Bayes dan Support Vector Machine (SVM) adalah dua algoritma machine learning yang paling sering diuji. Naive Bayes dikenal efisien karena arsitektur probabilistiknya, namun sering terkendala oleh asumsi independensi fitur yang kaku [10]. Sebaliknya, SVM dinilai unggul dalam mencari garis pemisah pada data teks berdimensi tinggi [11]. Berbagai studi telah membandingkan kedua metode ini pada platform seperti Lazada [12] dan TikTok Shop [13]. Sebelumnya, komparasi SVM dan Naive Bayes juga pernah diuji pada data ulasan pengguna media sosial X (Twitter) [14]. Meskipun algoritma tersebut berkinerja baik pada Twitter yang berkarakteristik teks pendek dan ringkas [15], relevansi performanya belum tentu berlaku sama jika diterapkan pada ulasan produk buku e-commerce yang memiliki struktur kalimat jauh lebih panjang dan deskriptif [16].

Kelemahan dari beberapa penelitian analisis sentimen terdahulu adalah evaluasi algoritma yang masih menggunakan dataset tidak seimbang dan hanya berfokus pada kategori produk belanja umum [6], [8], [9]. Oleh karena itu, kebaruan dari penelitian ini dipertegas pada analisis spesifik domain ulasan buku untuk menguji performa klasifikasi algoritma pada teks berkarakteristik deskriptif panjang, yang membedakannya dari karakteristik platform media sosial atau produk retail biasa [11], [14]. Selain itu, riset ini mengimplementasikan teknik preprocessing bahasa Indonesia yang dioptimasi untuk menangani struktur bahasa tidak baku di e-commerce. Penelitian ini bertujuan untuk membandingkan secara komprehensif ketangguhan Naive Bayes dan SVM dalam menghadapi kompleksitas tersebut, sehingga dapat menghasilkan model paling akurat yang dapat direkomendasikan bagi pelaku bisnis literatur digital.

2. METODE PENELITIAN

Metodologi penelitian ini dirancang secara sistematis untuk memastikan validitas dan kemampuan replikasi eksperimen klasifikasi sentimen [17]. Tahapan utamanya meliputi pengumpulan data, pelabelan, pra-pemrosesan teks, ekstraksi fitur, pemodelan, dan validasi.

2.1. Pengumpulan Data

Prosedur pelabelan data sentimen dalam penelitian ini tidak sepenuhnya bergantung pada sistem penilaian otomatis untuk menjaga validitas kontekstual bahasa. Pelabelan dilakukan secara manual dan divalidasi silang oleh tiga annotator, yang terdiri dari peneliti utama Arief Dharmadi, Dwi Hartanti, dan Anisatul Farida, untuk mencapai kesepakatan yang objektif. Klasifikasi menggabungkan skema peringkat bintang Shopee dan analisis tekstual [18], dengan kriteria berikut: ulasan bintang 1 dan 2 dikategorikan sebagai negatif, sedangkan ulasan bintang 4 dan 5 dikategorikan sebagai positif. Khusus untuk ulasan bintang 3, yang seringkali berpotensi ambigu, ketiga pemberi anotasi melakukan analisis konten secara manual. Jika teks ulasan terutama berisi keluhan tentang logistik atau kondisi pencetakan, ulasan tersebut ditempatkan

dalam kategori negatif, dan sebaliknya. Pendekatan ini menghasilkan kumpulan data yang sangat seimbang, dengan 551 ulasan positif dan 550 ulasan negatif.

2.2. Pra-pemrosesan Teks (Prepocessing) dan Ekstraksi Fitur

Tahapan pra-pemrosesan dilakukan untuk membersihkan data ulasan mentah dari anomali teks dan format bahasa informal. Proses ini mencakup penyeragaman huruf kecil (case folding), penghapusan karakter non-alfabet dan tanda baca menggunakan Regular Expression (filtering), pemotongan kalimat menjadi kata (tokenizing), penghapusan kata hubung tidak bermakna (stopword removal), dan pengembalian kata ke bentuk dasar (stemming). Implementasi stopword removal dan stemming dilakukan menggunakan pustaka Sastrawi berbasis Python. Data tekstual yang telah bersih kemudian dikonversi menjadi representasi vektor numerik menggunakan teknik pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) untuk menyoroti kata-kata dengan daya diskriminatif tinggi [19], [20].

2.3. Spesifikasi Pemodelan dan Evaluasi

Untuk membandingkan performa secara akurat, penelitian ini mengimplementasikan dua algoritma klasifikasi dengan spesifikasi parameter teknis yang didefinisikan secara eksplisit sebagai berikut:

Naive Bayes

Menggunakan varian Multinomial Naive Bayes (diinisialisasi dengan fungsi MultinomialNB()) karena karakteristik varian ini terbukti paling optimal dalam mengelola fitur diskrit dan representasi teks berbasis frekuensi kata hasil ekstraksi TF-IDF [21], [22].

***Support Vector Machine* (SVM)**

Diimplementasikan menggunakan fungsi Support Vector Classifier (SVC) dengan pengaturan parameter spesifik pada tipe kernel linier (kernel='linear'). Pemilihan kernel linier didasarkan pada asumsi bahwa matriks fitur teks berdimensi tinggi yang dihasilkan oleh TF-IDF umumnya dapat dipisahkan secara linier [23], [24].

Untuk mengatasi bias pengujian dan memastikan generalisasi model yang tinggi, proses evaluasi melampaui sekadar menggunakan rasio pembagian data pelatihan dan pengujian, dan sebagai gantinya menggunakan Validasi Silang 10-Fold Bertingkat. Teknik ini membagi dataset menjadi 10 bagian yang sama sambil mempertahankan proporsi kelas positif dan negatif di setiap bagian. Evaluasi akhir diukur menggunakan confusion matrix untuk mendapatkan skor rata-rata aktual untuk Akurasi, Presisi, dan Recall.

2.4. Implementasi Algoritma

Setelah data ulasan direpresentasikan dalam bentuk vektor numerik melalui pembobotan TF-IDF, matriks fitur tersebut dimasukkan ke dalam dua model klasifikasi untuk mempelajari pola sentimen:

Varian *Multinomial Naive Bayes*

Algoritma ini memprediksi probabilitas kemunculan kata kunci ulasan buku berdasarkan teorema Bayes. Model ini dihitung dengan mengasumsikan frekuensi kata dalam sebuah dokumen mengikuti distribusi multinomial, menggunakan rumus dasar:

$$P(c|d) \propto P(c) \prod_{i=1}^{n} P(t_i|c) \quad (1)$$

Dimana $P(c|d)$ adalah probabilitas kelas sentimen c (positif/negatif) terhadap dokumen ulasan d, dan $P(t_i|c)$, adalah probabilitas kemunculan *term* t_i pada kelas c.

Support Vector Machine* (SVM) *Linear

Model ini bekerja dengan membangun *hyperplane* optimal untuk memisahkan vektor ulasan positif dan negatif dengan margin maksimum. Dengan menggunakan tipe *kernel* linier, fungsi keputusan klasifikasi dirumuskan sebagai:

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

Di mana w mewakili *weight vector*, x adalah input vektor TF-IDF ulasan buku, dan b adalah bias.

2.5. Evaluasi Performa (*Confusion Matrix*)

Mekanisme pengujian kinerja untuk kedua model menggunakan kerangka kerja Validasi Silang Multilevel 10-*Fold*, bukan pengujian tunggal. Teknik ini secara acak membagi seluruh *dataset* (1.101 ulasan) menjadi 10 bagian (*fold*) dengan proporsi kelas yang seimbang. Proses pelatihan dan pengujian diulang 10 kali, dengan setiap iterasi menggunakan 9-*fold* sebagai data pelatihan dan 1-*fold* sisanya sebagai data pengujian.

Untuk setiap lipatan pengujian, kecocokan antara label asli yang dikurasi oleh ketiga annotator dan label yang diprediksi model dipetakan ke dalam *confusion matrix*. Persentase rata-rata di semua lipatan dihitung untuk menghasilkan metrik evaluasi akhir sebagai berikut:

Akurasi

Mengukur rasio keseluruhan prediksi sentimen yang ditebak secara benar oleh model terhadap total seluruh data uji.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100\% \quad (3)$$

Presisi

Mengukur akurasi model dalam memprediksi label positif dari total data yang diduga sebagai kelas positif oleh sistem.

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

Recall

Mengukur kemampuan model dalam menemukan kembali seluruh data yang aslinya berlabel positif dari total keseluruhan populasi data positif yang ada.

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

Keterangan:

TP = *True Positive,*
TN = *True Negative*,
FP = *False Positive*,
FN = *False Negative*.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengujian dan evaluasi eksperimen klasifikasi sentimen menggunakan teknik *Stratified 10-Fold Cross Validation*. Pembahasan berfokus pada hasil

pembersihan teks, ekstraksi fitur, dan perbandingan kinerja antara algoritma *Naive Bayes Multinomial* dan *Support Vector Machine* (SVM). Evaluasi mendalam dilakukan melalui analisis kesalahan dan pengujian signifikansi untuk menginterpretasikan ketahanan model dalam menangani karakteristik informal dan deskriptif dari ulasan buku.

3.1. Hasil Preprocessing Teks

Tahap pra-pemrosesan berhasil mengubah 1.101 data ulasan mentah yang tidak terstruktur menjadi serangkaian token kata yang terstandarisasi. Karakteristik teks ulasan buku di platform Shopee sangat didominasi oleh *noise* berupa bahasa informal, singkatan ekspresif (seperti "bgt", "ajeee"), dan kesalahan ejaan. Penggunaan pustaka Sastrawi dalam proses *filtering* dan *stemming* terbukti sangat penting untuk menormalisasi struktur kata tanpa kehilangan esensi polaritas sentimen pelanggan. Pengurangan dimensi ruang fitur teks ini menyederhanakan kompleksitas komputasi tahap ekstraksi fitur. Implementasi teknis dari serangkaian prosedur pembersihan data disajikan dalam Kode Sumber 1.

```
# Kode Sumber 1. Implementasi Text Preprocessing Menggunakan
Sastrawi
import re
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory

# 1. Inisialisasi Stemmer dan Stopword Remover
stemmer = StemmerFactory().create_stemmer()
stopword = StopWordRemoverFactory().create_stop_word_remover()

# 2. Fungsi Pembersihan Teks
def bersihkan_teks(teks):
    # Case Folding
    teks = str(teks).lower()

    # Filtering (Menghapus karakter selain alfabet)
    teks = re.sub(r'[^a-z]+', ' ', teks)

    # Stopword Removal (Sastrawi)
    teks = stopword.remove(teks)

    # Tokenizing dan Stemming
    tokens = teks.split()
    teks_final = stemmer.stem(" ".join(tokens))

    return teks_final.strip()

# 3. Eksekusi transformasi pada dataset ulasan
df['teks_bersih'] = df['review_comment'].apply(bersihkan_teks)
```

Kode Sumber 1. Implementasi *Text Preprocessing* Menggunakan Sastrawi

3.2. Hasil Pembobotan Kata (TF-IDF)

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk mengukur signifikansi dan bobot kepentingan suatu kata dalam keseluruhan *dataset* korpus. Representasi vektor numerik yang dihasilkan oleh TF-IDF memungkinkan model untuk menyoroti kata-kata dengan daya diskriminatif tinggi. Berdasarkan hasil ekstraksi fitur, diperoleh sepuluh istilah dominan dengan nilai bobot tertinggi, yang bertindak sebagai karakteristik utama atau pengidentifikasi fitur untuk setiap kategori sentimen, seperti yang disajikan pada Tabel 1.

Tabel 1. 10 Kata Kunci Dominan

No	Sentimen Positif	Sentimen Negatif
1	bagus	kurang
2	cepet	kecewa
3	cepat	bagi
4	keren	buruk
5	aman	tidak
6	suka	lama
7	lumayan	halaman
8	rapi	depan
9	baguss	kok
10	thanks	tekuk

Istilah-istilah seperti 'bagus,' 'cepat,' dan 'aman' secara konsisten mendominasi kategori sentimen positif, yang mencerminkan kepuasan pengguna terhadap kualitas cetak buku dan keamanan logistik kemasannya. Sebaliknya, kategori sentimen negatif didominasi oleh kata-kata yang mewakili ketidakpuasan fungsional atau layanan, seperti 'rusak,' 'usang,' dan 'kecewa.' Pembagian *dataset* pelatihan dan pengujian tidak lagi menggunakan *split* statis, tetapi dieksekusi secara dinamis di dalam *Stratified 10-Fold Cross Validation* untuk memastikan bahwa semua token kata menerima proporsi evaluasi yang adil dan representatif.

3.3. Performa Algoritma (*Confusion Matrix*)

Evaluasi kinerja klasifikasi dilakukan secara komprehensif menggunakan teknik *Stratified 10-Fold Cross-Validation* pada total 1.101 *dataset* ulasan. Skema ini memungkinkan pengujian untuk mengandalkan lebih dari satu klaster statis data uji, tetapi sebaliknya mengakumulasi hasil prediksi dari 10 lipatan eksperimental yang berbeda secara berurutan. Hasil pemetaan prediksi dari total lipatan diakumulasikan ke dalam *confusion matrix* untuk mengevaluasi distribusi *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN) secara *real-time*, seperti yang disajikan pada Tabel 2.

Tabel 2. Akumulasi *Confusion Matrix* dan Metrik Performa Akhir

Algoritma	*Actual Class*	*Predicted Negative*	*Predicted Positive*	Rata-rata Akurasi	Rata-rata Presisi	Rata-rata *Recall*
Support Vector Machine (SVM)	*Actual Negative*	489 TN	61 FP	86,47%	88,37%	84,04%
	Actual Positive	88 FN	463 TP			
Multinomial Naïve Bayes	*Actual Negative*	481 TN	69 FP	86,29%	87,16%	85,12%
	Actual Positive	82 FN	469 TP			

Berdasarkan data yang dikumpulkan pada Tabel 2, pengujian menggunakan *Stratified 10-Fold Cross Validation* menghasilkan angka kinerja yang sangat kompetitif antara kedua algoritma. Model SVM dengan *kernel* linier berhasil memprediksi 952 ulasan (489 TN dan 463 TP).

Sementara itu, algoritma *Multinomial Naive Bayes* mencatat jumlah prediksi benar yang hampir sama, yaitu 950 ulasan (481 TN dan 469 TP).

Mengingat selisih rata-rata akurasi akhir antara SVM (86,47%) dan *Multinomial Naive Bayes* (86,29%) hanya sebesar 0,18% (setara dengan selisih 2 ulasan dari total 1.101 data), analisis dilanjutkan dengan melakukan uji signifikansi statistik menggunakan *McNemar's Test*. Uji ini dieksekusi berdasarkan matriks kontingensi untuk menilai apakah perbedaan performa kedua model bersifat signifikan atau hanya terjadi secara kebetulan. Hasil perhitungan *McNemar's Test* menghasilkan nilai-p ($p - value$) sebesar 0,815 ($p > 0{,}05$). Secara statistik, nilai ini menunjukkan tidak adanya perbedaan yang signifikan secara meyakinkan antara akurasi global SVM dan *Naive Bayes* pada *dataset* ulasan buku Shopee yang seimbang ini.

Meskipun akurasi globalnya secara statistik setara, kedua algoritma tersebut menunjukkan karakteristik diskriminatif yang berbeda dalam setiap metrik polaritas. SVM terbukti lebih unggul dalam mengurangi jumlah *False Positives* menjadi hanya 61 kejadian, dibandingkan dengan *Naive Bayes* yang melewatkan hingga 69 kejadian. Karakteristik ini secara langsung berdampak pada keunggulan nilai presisi SVM yang mencapai 88,37%, melampaui *Naive Bayes* yang mencapai 87,16%. Sebaliknya, *Multinomial Naive Bayes* menunjukkan kinerja yang lebih sensitif dalam mengenali kelas positif dengan tingkat *recall* 85,12%, berbeda dengan SVM yang mencatat 84,04% karena tingginya jumlah *False Negatives* (88 ulasan).

Prosedur komputasional dari keseluruhan fase pelatihan, implementasi *Stratified 10-Fold Cross Validation*, hingga kalkulasi metrik evaluasi final dari kedua model tersebut dirinci pada Kode Sumber 2.

```
# Kode Sumber 4. Implementasi Pelatihan dan Evaluasi Model K-Fold
from sklearn.model_selection import StratifiedKFold, cross_vali-
date, cross_val_predict
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC

# 1. Inisialisasi model dengan parameter spesifik
nb_model = MultinomialNB()
svm_model = SVC(kernel='linear')

# 2. Konfigurasi Stratified 10-Fold Cross Validation
kfold = StratifiedKFold(n_splits=10, shuffle=True, ran-
dom_state=42)
scoring_metrics = ['accuracy', 'precision', 'recall']

# 3. Pelatihan dan ekstraksi metrik performa (Rata-rata 10-Fold)
scores_nb = cross_validate(nb_model, X, y, cv=kfold, scor-
ing=scoring_metrics)
scores_svm = cross_validate(svm_model, X, y, cv=kfold, scor-
ing=scoring_metrics)

# 4. Ekstraksi hasil prediksi gabungan untuk pembuatan Confusion
Matrix
y_pred_nb = cross_val_predict(nb_model, X, y, cv=kfold)
y_pred_svm = cross_val_predict(svm_model, X, y, cv=kfold)
```

Kode Sumber 2. Implementasi Pelatihan dan Evaluasi Model.

3.4. Analisis Komparatif dan Pembahasan

Meskipun pengujian *Stratified 10-Fold Cross Validation* Bertingkat tidak menunjukkan perbedaan yang signifikan secara statistik dalam akurasi global kedua algoritma, analisis kesalahan pada tingkat pengamatan individu mengungkapkan perbedaan mendasar dalam cara kedua model menangani teks deskriptif yang kompleks. Dalam kelas sentimen negatif, ditemukan

beberapa kasus di mana *Multinomial Naive Bayes* salah mengklasifikasikan (terdeteksi sebagai ulasan positif), sementara algoritma SVM berhasil memprediksinya secara akurat.

Salah satu kasus tersebut terlihat pada ulasan pertama: "sumpah lama bgt dipackingnya ada 5 harii astagfirullah bikin gue ngamok ajeee, cetakannya jg ga rapii, dahlah terkahir beli". Kasus lainnya ditemukan pada ulasan kedua: "Pengajuan pembatalan, sebelum pesanan dikirim tidak disetujui. Padahal mau ganti pesanan dan sudah konfirmasi sebelum dikirim. Sangat tidak ramah".

Kegagalan *Naive Bayes* dalam mengklasifikasikan kedua ulasan tersebut bersumber dari kelemahan mendasarnya, yakni asumsi independensi fitur. Pada contoh kedua, algoritma ini terkecoh oleh kemunculan kata-kata yang secara leksikal berkonotasi positif seperti 'disetujui', 'konfirmasi', dan 'ramah'. Karena *Naive Bayes* memecah teks dan menghitung probabilitas tiap kata secara terpisah, model ini gagal mengasosiasikan kata-kata tersebut dengan kata negasi 'tidak' yang mendahuluinya. Pada contoh pertama, tingginya intensitas bahasa *slang* dan ekspresi informal yang memanjangkan teks ('bgt', 'ajeee', 'ngamok') menciptakan *noise* yang mengaburkan distribusi probabilitas kelas negatif.

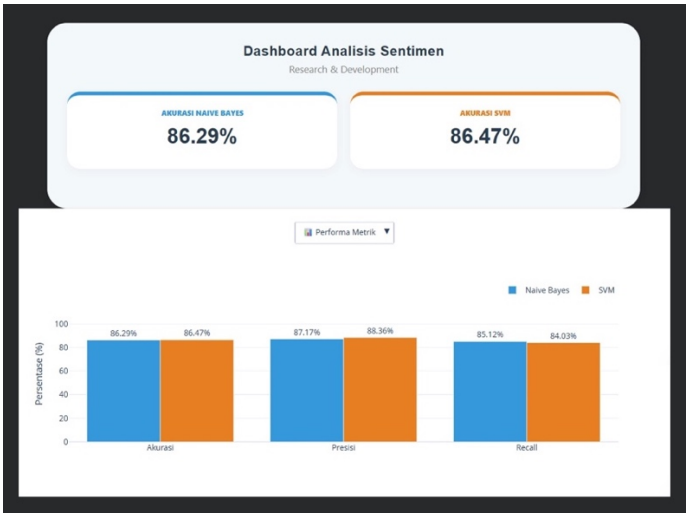
Sebaliknya, keunggulan SVM tidak hanya terletak pada kemampuannya untuk secara teoritis menarik garis pemisah, tetapi juga pada keandalannya dalam mengevaluasi bobot vektor kalimat secara holistik. Dengan menggunakan *kernel* linier, SVM memberikan bobot vektor pendukung negatif yang kuat pada kemunculan fitur-fitur spesifik seperti 'tidak', 'lama', dan 'rapii' (dari ga rapii). SVM mampu mengenali bahwa kombinasi fitur-fitur ini dalam ruang satu dimensi sebagian besar mengarah pada kekecewaan, bahkan ketika diselingi dengan kata-kata positif atau interupsi bahasa informal. Hal ini menegaskan bahwa untuk karakteristik teks ulasan buku yang cenderung deskriptif panjang dan penuh dengan struktur kalimat informal, SVM terbukti lebih tepat daripada pendekatan yang murni probabilistik.

Temuan observasional ini memberikan perspektif yang lebih mendalam apabila dikomparasikan dengan literatur terdahulu yang membandingka Naive Nayes dan SVM pada berbagai domain ulsan [25], 26]. Riset [8] sebelumnya mencatatkan akurasi SVM yang sangat tinggi, yakni 91,48%, pada *dataset* Shopee. Angka tersebut lebih tinggi dari capaian riset ini (86,47%). Namun secara metodologis, perbedaan ini sangat logis; riset [8] menggunakan data ulasan produk umum yang cenderung pendek dan langsung, serta dievaluasi menggunakan data *split* statis. Capaian performa dalam penelitian ini dinilai jauh lebih *robust* dan realistis karena diuji pada domain spesifik buku yang narasinya kompleks, serta telah divalidasi secara ketat menggunakan *10-Fold Cross Validation* untuk menghindari *overfitting*.

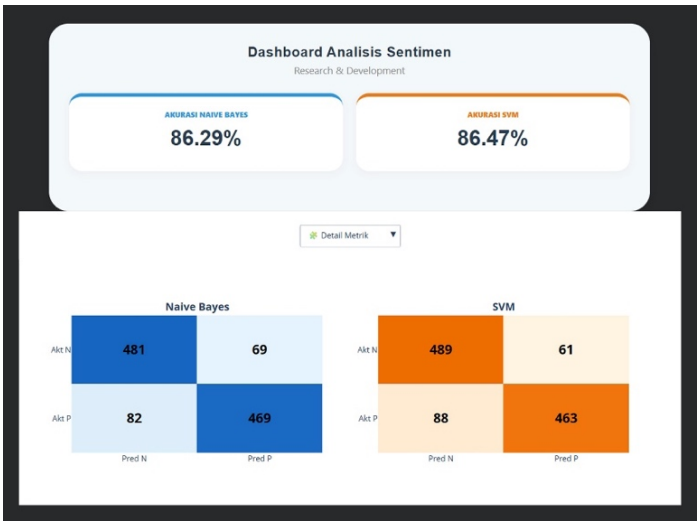
Di sisi lain, performa SVM dalam riset ini berhasil melampaui hasil studi [11] yang hanya mencapai akurasi 75% pada data platform Lazada. Lonjakan performa sebesar 11,47% ini dapat diatribusikan pada dua pembaruan desain eksperimen. Pertama, riset ini menggunakan kurasi *dataset* yang sangat seimbang (550:551) dan divalidasi silang oleh tiga orang anotator, sehingga menghindarkan model dari bias kelas mayoritas—sebuah aspek yang kerap diabaikan pada pengumpulan data rating otomatis. Kedua, implementasi tahapan *stopword removal* dan *stemming* yang dioptimasi khusus untuk menyaring tingginya bahasa informal terbukti ampuh mereduksi fitur yang tidak relevan. Komparasi ini membuktikan bahwa validitas pelabelan dan kualitas pra-pemrosesan data memberikan impak yang jauh lebih krusial terhadap reliabilitas model analitik di *e-commerce* dibandingkan sekadar pemilihan varian algoritma.

3.5. Prototipe Sistem Analisis Sentimen

Sebagai kontribusi terapan dari penelitian ini, model klasifikasi sentimen yang dievaluasi diintegrasikan ke dalam prototipe dasbor analitik interaktif. Prototipe ini dikembangkan menggunakan pustaka *Plotly* di Python untuk secara komprehensif meringkas dan memvisualisasikan metrik kinerja hasil uji *Stratified 10-Fold Cross Validation*. Implementasi antarmuka ini dirancang untuk memungkinkan penjual buku di platform Shopee untuk secara dinamis memantau perbandingan akurasi model dan matriks kebingungan sentimen pelanggan, tanpa perlu memahami kompleksitas komputasi yang mendasari perbandingan tersebut.



Gambar 1. Dashboard Analisis Sentimen Grafik Batang.



Gambar 2. Dashboard Analisis Sentimen Confusion Matrix.

Representasi visual pada Gambar 1 menyajikan komparasi performa metrik (Akurasi, Presisi, dan *Recall*) dari algoritma *Naive Bayes* dan SVM secara berdampingan. Sementara itu, pemetaan distribusi kesalahan prediksi (*False Positive* dan *False Negative*) divisualisasikan melalui *heatmap* matriks konfusi pada Gambar 2. Logika fungsionalitas interaktif yang menopang pembaruan tampilan grafis tersebut disajikan pada Kode Sumber 3. Implementasi kode ini difokuskan pada mekanisme *state management*, yang memungkinkan transisi visual secara instan dan efisien antara mode tampilan grafik performa dengan mode detail analisis *heatmap*.

```
# Kode Sumber 5. Logika Navigasi Dashboard Interaktif
import plotly.graph_objects as go
from plotly.subplots import make_subplots
# 1. Inisialisasi arsitektur kanvas (Subplots)
fig = make_subplots(rows=2, cols=2, specs=[[{"colspan": 2}, None], [{}, {}]])

# [Penambahan objek grafik batang dan heatmap diproses di sini]
# 2. Implementasi mekanisme State Management (Navigasi Interaktif)
fig.update_layout(
    updatemenus=[{
        "buttons": [
            {
                "label": "📊 Performa Metrik",
                "method": "update",
                "args": [{"visible": [True, True, False, False]}]
            },
            {
                "label": "🧩 Detail Metrik",
                "method": "update",
                "args": [{"visible": [False, False, True, True]}]
            }
        ],
        "direction": "down",
        "showactive": True,
        "x": 0.5, "y": 1.25
    }],
    template="plotly_white",
    title="Dashboard Komparasi Sentimen Ulasan Buku"
)
```

Kode Sumber 3. Logika Navigasi *Dashboard* Interaktif.

4. KESIMPULAN

Penelitian ini berhasil menunjukkan efektivitas komparatif algoritma *Multinomial Naive Bayes* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen spesifik dalam ulasan produk buku di platform *e-commerce* Shopee. Melalui pengujian *Stratified 10-Fold Cross Validation* pada *dataset* seimbang yang dikurasi secara manual, SVM mencatat akurasi rata-rata 86,47%, mengungguli *Naive Bayes* yang mencapai 86,29%. Meskipun uji signifikansi *McNemar* menunjukkan bahwa perbedaan akurasi global tidak signifikan secara statistik, analisis kesalahan membuktikan bahwa SVM jauh lebih unggul dan lebih akurat (88,37%) dalam menangani kompleksitas teks deskriptif panjang, bahasa informal, dan struktur kalimat negasi berkat kemampuannya untuk memodelkan bobot vektor pendukung secara holistik, di mana *Naive Bayes* seringkali tertipu oleh asumsi independensi kata.

Implikasi praktis dari temuan ini sangat relevan bagi bisnis literatur digital yang beroperasi di platform *e-commerce*. Presisi SVM yang tinggi menjadikannya model yang sangat direkomendasikan untuk diintegrasikan ke dalam sistem pemantauan toko online guna menghindari *false positive*, yang dapat merusak kredibilitas manajemen dalam menilai kepuasan pelanggan.

Namun, penelitian ini memiliki beberapa keterbatasan. Fokus dataset pada 1.101 ulasan buku spesifik domain membuat model berisiko mengalami penurunan kinerja jika diterapkan pada kategori produk lain seperti elektronik atau fesyen tanpa penyempurnaan lebih lanjut. Selain itu, prosedur pelabelan silang oleh tiga annotator manusia memakan waktu dan tidak efisien ketika diimplementasikan pada dataset skala besar. Oleh karena itu, penelitian lebih lanjut disarankan untuk memperluas cakupan objek *e-commerce* analisis sentimen lintas domain dan membandingkan kinerja SVM dengan arsitektur pembelajaran mendalam, seperti *Long Short-Term*

Memory (LSTM) atau BERT Indonesia, untuk mengevaluasi keterbatasan pengenalan semantik dalam teks tidak terstruktur.

PERNYATAAN PENGGUNAAN ARTIFICIAL INTELLIGENCE

Penulis menggunakan chatGPT untuk membantu perbaikan tata bahasa pada bagian Pendahuluan dan hasil. Seluruh output dari chatGPT telah diperiksa dan diverifikasi oleh penulis, dan penulis bertanggung jawab penuh atas isi artikel ini.

KONTRIBUSI PENULIS

Penulis Pertama: Konsep, Metodologi, Perangkat Lunak, Penulisan Draft.

Penulis Kedua: Konsep, Analisis formal, Validasi, Penulisan – review & editing, Supervisi.

Penulis Ketiga: Visualisasi, Penulisan – review & editing, supervisi.

PENDANAAN

Penelitian ini tidak menerima hibah atau pendanaan khusus dari lembaga pendanaan mana pun, baik sektor publik, komersial, maupun nirlaba.

KONFLIK KEPENTINGAN

Penulis menyatakan tidak terdapat konflik kepentingan terkait publikasi artikel ini..

DAFTAR PUSTAKA

[1] Steven and I. Fenriana, "Analisis Sentimen Membandingkan Pengguna Aplikasi E-Commerce Tokopedia Dan Shopee Menggunakan Algoritma," *J. POTERS (Proceedings Technol. Eng. Comput.*, vol. 1, no. 1, pp. 32–39, 2025, [Online]. Available: https://jurnal.buddhidharma.ac.id/index.php/poters/article/view/3588

[2] D. Setyowati, "Transaksi E-Commerce Pada 2024," katadata.co.id. [Online]. Available: https://katadata.co.id/digital/e-commerce/685ba4184c37a/transaksi-belanja-online-di-indonesia-melambat-thailand-dan-malaysia-melesat

[3] R. C. Rivaldi and T. D. Wismarini, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP)," *J. Elektron. dan Komput.*, vol. 17, no. 1, pp. 120–128, 2024, doi: https://doi.org/10.51903/elkom.v17i1.1680.

[4] H. Diandra Tasya, Dea Vahira Wani, Madhani Nur Syafitri, Naura Nitajria, "Menyelami Teks Ulasan Dalam Bahasa Indonesia Beserta Struktur," *J. Innov. Creat.*, vol. 5, no. 3, pp. 12995–12999, 2025, doi: https://doi.org/10.31004/joecy.v5i3.3972.

[5] M. K. Shabrina Rasyid Munthe, S.Pd.I., M.S, Samsir, S.T., M. Rina Asriani Levianti, S.E., M.Ak, Dr. Basyarul Ulya, SH., M. K. Nurhanna Harahap, M.Hum, Abdul Hakim Dalimunthe, M. . M. Rusdi, S.Pd.I., M.M, Nurhabibah Ritonga, S.E., and M. K. Ibnu Rasyid Munthe, S.T., *Buku Monograf Pemodelan Sentimen Dalam Ulasan Produk Dan Penerapannya Pada E-Commerce*. Batu Bara, Sumatera Utara: PT. JASA NIAGA DIGITAL INDONESIA, 2025. [Online]. Available: https://jurnal.jndi.my.id/index.php/buku/article/view/339

[6] B. I. Hasbi and I. S. Putro, "Perbandingan Kinerja Algoritma Naive Bayes Dan Support Vector Machine untuk Analisis Sentimen Ulasan Produk E- Commerce Menggunakan Pembobotan TF-IDF," *J. Ikraith-Informatika*, vol. 9, no. 2, pp. 123–133, 2025, doi: https://doi.org/10.37817/ikraith-informatika.v9i2.

[7] A. D. Wirahadi, S. Lestanti, and A. P. Kusuma, "Analisis Sentimen Terhadap Produk Kecantikan Emina Daily Matte Loose Powder di Emina Official Shop Menggunakan Metode Support Vector Machine," *J. Ekon. Manajemen, Akunt. dan Keuang.*, no. 1, pp. 1–10, 2026, doi: https://doi.org/10.53697/emak.v7i1.3624.

[8] T. P. R. Sanjaya, A. Fauzi, and A. F. N. Masruriyah, "Analisis Sentimen Ulasan Pada E-Commerce Shopee Menggunakan Algoritma Naive Bayes Dan Support Vector Machine Analysis," *J. INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.

[9] A. Hanafiah *et al.*, "Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier," *J. INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, pp. 230–236, 2023, doi: https://doi.org/10.31539/intecoms.v6i1.5845.

[10] M. Y. Alghifari, M. R. Sanjaya, D. R. Indah, and E. L. Ruskan, "Perbandingan Algoritma Svm Dan Naive Bayes Dalam Analisis Sentimen Terhadap Ulasan Pengguna Bukalapak," *J. INOVTEK Polbeg - Seri Inform.*, vol. 10, no. 3, pp. 1623–1633, 2025, doi: https://doi.org/10.35314/dqhpkb12.

[11] S. Setyabudi and E. Aryanny, "Analisis Sentimen Penilaian Pengguna Marketplace Lazada Dengan Metode Naïve Bayes Dan Support Vector Machine," *J. INOVTEK Polbeg - Seri Inform.*, vol. 10, no. 1, pp. 422–433, 2025, doi: https://doi.org/10.35314/sww8cg21.

[12] F. Dikananda, A. R. Dikananda, and S. Anwar, "Analisis Sentimen Terhadap Marketplace Menggunakan Algoritma Naïve Bayes (Studi Kasus : Lazada Indonesia)," *J. JATI*, vol. 8, no. 4, pp. 8219–8225, 2024, doi: https://doi.org/10.36040/jati.v8i4.10965.

[13] E. Indrayuni and A. Nurhadi, "Analisis Sentimen Aplikasi Tiktok Shop Seller Center Menggunakan Naive Bayes, Svm Dan Logistic Regression," *J. INTI Nusa Mandiri*, vol. 20, no. 1, pp. 26–34, 2025, doi: https://doi.org/10.33480/inti.v20i1.6851.

[14] M. R. Emiliana, M. R. I. Fata, and M. G. Adam, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi X," *J. Pustaka Data*, vol. 5, no. 2, pp. 331–341, 2025, doi: https://doi.org/10.55382/jurnalpustakadata.v5i2.1460.

[15] A. Albladi, M. Islam, and C. Seals, "Sentiment Analysis of Twitter Data Using NLP Models: A Comprehensive Review," *IEEE Access*, vol. 13, no. January, pp. 30444–30468, 2025, doi: 10.1109/ACCESS.2025.3541494.

[16] K. Alahmadi, S. Alharbi, J. Chen, and X. Wang, "Generalizing Sentiment Analysis: A Review Of Progress, Challenges, And Emerging Directions," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 45, Apr. 2025, doi: 10.1007/s13278-025-01461-8.

[17]J. Riebesell *et al.*, "A Framework To Evaluate Machine Learning Crystal Stability Predictions," *Nat. Mach. Intell.*, vol. 7, no. 6, pp. 836–847, 2025, doi: 10.1038/s42256-025-01055-1.

[18] P. Maxmiliano, Y. F. Riti, I. Y. Nugroho, and C. E. C. Juniarto, "Perbandingan Algoritma Naive Bayes Dan Bert Untuk Analisis Sentimen Ulasan Produk Shopee Berdasarkan Rating Dan Atribut Produk (Warna/Kategori)," *J. Media Inform.*, vol. 6, no. 5, pp. 2552–2565, 2025, [Online]. Available: https://ejournal.sisfokomtek.org/index.php/jumin/article/download/6761/4284

[19] L. Zhang, "Features Extraction Based On Naive Bayes Algorithm And TF-IDF For News Classification," *Journals Plos One*, vol. 20, no. 7 July, pp. 1–17, 2025, doi: 10.1371/journal.pone.0327347.

[20] S. U. Royan, N. Suarna, I. Ali, D. Solihudin, S. V. Machine, and S. V. Machine, "Analisis Sentimen Ulasan Produk Skincare Di Shopee Untuk Meningkatkan Kualitas Produk Menggunakan Metode Support Vector Machine," *J. Inf. Dan Komput.*, vol. 13, pp. 96–105, 2025, doi: https://doi.org/10.35959/jik.v13i01.656.

[21] C. L. R. Dina Siti Nurrochmah, Nining Rahaningsih, Raditya Danar Dana, "Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Ulasan Aplikasi Kitalulus Di Google Play Store," *J. Inform. Terpadu*, vol. 8, no. 1, pp. 1–7, 2022, doi: https://doi.org/10.54914/jit.v11i1.1544.

[22] A. T. Bisono and A. Zulherry, "Analisis Sentimen Game Genshin Impact Untuk Mengetahui Reaksi Dan Harapan Pemain Menggunakan Metode Naïve Bayes," *sudo J. Tek. Inform.*, vol. 4, no. 2, pp. 183–193, 2025, doi: 10.56211/sudo.v4i2.1131.

[23] S. D. Parameswari, M. Lubis, S. Suakanto, Y. Z. Ramadhan, R. N. Amanah, and R. A. Dila, "Studi Perbandingan Naïve Bayes Dan Support Vector Machine (SVM) Dalam Analisis Sentimen Pengguna Metaverse," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 4, no. 3, pp. 1059–1065, 2025, doi: 10.55826/jtmit.v4i3.1122.

[24] 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," *Int. J. Informatics Vis.*, vol. 9, no. 2, pp. 796–807, 2025, doi: 10.62527/joiv.9.2.2905.

[25] I. Kurniawan, A. L. Hananto, S. S. Hilabi, and A. Hananto, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *J. JATISI*, vol. 10, no. 1, pp. 731–740, 2023, doi: https://doi.org/10.35957/jatisi.v10i1.3582.

[26] R. Rahmat, A. Rahim, and A. Arbansyah, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Alibaba Di Google Play Store," *J. JATI*, vol. 9, no. 2, pp. 3050–3057, 2025, doi: https://doi.org/10.36040/jati.v9i2.13269.