SKRIPSI

KLASIFIKASI SENTIMEN KOMENTAR PENGGUNA TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES: EVALUASI KINERJA DAN ANALISIS KUANTITATIF AKURASI.

Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana pada Program Studi Sistem Komputer

OLEH:

ALWI SYAHRIR

NIM: 210299

PROGRAM STUDI SISTEM KOMPUTER
FAKULTAS AI
AI UNIVERSITAS

2026

LEMBAR PENGESAHAN

KLASIFIKASI SENTIMEN KOMENTAR PENGGUNA TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES: EVALUASI KINERJA DAN ANALISIS KUANTITATIF AKURASI.

Disusun Oleh: ALWI SYAHRIR (NIM: 210299)

Telah diuji dan dipertahankan di hadapan Tim Dewan Penguji Ujian Skripsi pada Program Studi Sistem Komputer, Fakultas Ai, AI UNIVERSITAS, dan dinyatakan telah memenuhi seluruh persyaratan akademik kelulusan.

Dosen Pembimbing I	Dosen Pembimbing II
(_____)	(_____)
NIP/NIDN.	NIP/NIDN.

Mengetahui,
Ketua Program Studi Sistem Komputer

(_____)
NIP/NIDN.

SURAT PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **ALWI SYAHRIR**
NIM / NPM	: 210299
Program Studi	: Sistem Komputer
Fakultas	: Fakultas Ai

Menyatakan dengan sesungguhnya bahwa skripsi berjudul **"Klasifikasi Sentimen Komentar Pengguna TikTok Menggunakan Algoritma Naive Bayes: Evaluasi Kinerja dan Analisis Kuantitatif Akurasi."** ini adalah murni hasil karya penelitian, pemikiran, dan penulisan mandiri saya sendiri. Karya ini belum pernah diajukan untuk memperoleh gelar akademik di perguruan tinggi mana pun, dan bebas dari unsur plagiarisme atau fabrikasi data.

Seluruh kutipan, data, dan pemikiran yang bersumber dari pihak lain telah dicantumkan dalam daftar pustaka sesuai dengan etika ilmiah. Apabila di kemudian hari terbukti terdapat plagiarisme, saya bersedia menerima sanksi akademik sesuai aturan yang berlaku.

AI UNIVERSITAS, 22 November 2026

Yang membuat pernyataan,

[Materai Rp 10.000]

ALWI SYAHRIR

NIM. 210299

ABSTRAK

KLASIFIKASI SENTIMEN KOMENTAR PENGGUNA TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES: EVALUASI KINERJA DAN ANALISIS KUANTITATIF AKURASI.

Oleh: ALWI SYAHRIR (NIM: 210299)

Program Studi Sistem Komputer, AI UNIVERSITAS

Penelitian ini bertujuan untuk menginvestigasi, menganalisis, dan menyelesaikan permasalahan pada topik "Klasifikasi Sentimen Komentar Pengguna TikTok Menggunakan Algoritma Naive Bayes: Evaluasi Kinerja dan Analisis Kuantitatif Akurasi.". Metode penelitian yang diterapkan adalah Kuantitatif (Survei / Regresi) melalui prosedur pengumpulan, pengolahan, dan pengujian data empiris secara sistematis. Hasil penelitian menunjukkan bahwa pendekatan yang diterapkan mampu menjawab seluruh rumusan masalah secara terukur, memberikan evaluasi kinerja yang valid, serta menghasilkan kontribusi teoretis dan praktis bagi pengembangan keilmuan pada Program Studi Sistem Komputer.

Kata Kunci: Sistem Komputer, Kuantitatif (Survei / Regresi), Analisis Empiris, Skripsi.

ABSTRACT

KLASIFIKASI SENTIMEN KOMENTAR PENGGUNA TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES: EVALUASI KINERJA DAN ANALISIS KUANTITATIF AKURASI.

By: ALWI SYAHRIR (Student ID: 210299)

Department of Sistem Komputer, AI UNIVERSITAS

This study aims to investigate, analyze, and resolve the empirical problems addressed in "Klasifikasi Sentimen Komentar Pengguna TikTok Menggunakan Algoritma Naive Bayes: Evaluasi Kinerja dan Analisis Kuantitatif Akurasi.". Employing a structured Kuantitatif (Survei / Regresi) methodology, this research systematically collects, processes, and evaluates empirical data. The findings demonstrate that the implemented approach effectively addresses all research questions, provides verifiable evaluation outcomes, and contributes both theoretical insights and practical implications to the field of Sistem Komputer.

Keywords: *Sistem Komputer, Empirical Analysis, Research Methodology, Undergraduate Thesis.*

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penyusunan skripsi berjudul "Klasifikasi Sentimen Komentar Pengguna TikTok Menggunakan Algoritma Naive Bayes: Evaluasi Kinerja dan Analisis Kuantitatif Akurasi." dapat diselesaikan sebagai salah satu syarat akademik guna memperoleh gelar Sarjana pada Program Studi Sistem Komputer, AI UNIVERSITAS.

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada pimpinan universitas, ketua program studi, dosen pembimbing, dosen penguji, serta keluarga dan rekan-rekan atas bimbingan, arahan, serta dukungan yang telah diberikan. Penulis menyadari bahwa naskah ini masih terbuka terhadap saran konstruktif demi penyempurnaan di masa mendatang.

AI UNIVERSITAS, November 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL ..... **i**
LEMBAR PENGESAHAN ..... **ii**
SURAT PERNYATAAN KEASLIAN ..... **iii**
ABSTRAK ..... **iv**
ABSTRACT ..... **v**
KATA PENGANTAR ..... **vi**
DAFTAR ISI ..... **vii**
DAFTAR TABEL ..... **viii**
DAFTAR GAMBAR ..... **ix**
DAFTAR LAMPIRAN ..... **x**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Identifikasi Masalah 2
1.3 Batasan Masalah 3
1.4 Rumusan Masalah 4
1.5 Tujuan Penelitian 5
1.6 Manfaat Penelitian 6
1.7 Keaslian Penelitian 7
1.8 Sistematika Penulisan 7
BAB II TINJAUAN PUSTAKA **9**
2.1 Landasan Teori 9
2.2 Penelitian Terdahulu 15
2.3 Kerangka Pemikiran 17
2.4 Hipotesis Penelitian 19
BAB III METODOLOGI PENELITIAN **20**
3.1 Desain Penelitian dan Alur Sistem 20
3.2 Sumber dan Teknik Pengumpulan Dataset 22
3.3 Tahapan Pra-Pemrosesan Data 23

3.4 Pembobotan Fitur dan Pembagian Data 25
3.5 Pemodelan Algoritma Naive Bayes 26
3.6 Metrik Evaluasi Kinerja 27
BAB IV HASIL PENELITIAN DAN PEMBAHASAN **30**
4.1 Deskripsi dan Karakteristik Dataset Penelitian 30
4.2 Hasil Pra-Pemrosesan Teks 32
4.3 Hasil Pengujian dan Evaluasi Kinerja Algoritma 35
4.4 Pembahasan Kritis Hasil Penelitian 38
BAB V PENUTUP **43**
5.1 Kesimpulan 43
5.2 Saran 44
DAFTAR PUSTAKA **46**
LAMPIRAN **48**

DAFTAR TABEL

Tabel 2.1 Sintesis Penelitian Terdahulu **16**

Tabel 3.1 Tahapan Pra-Pemrosesan Data **24**

Tabel 3.2 Struktur Confusion Matrix Tiga Kelas **28**

Tabel 4.1 Distribusi Kelas Sentimen Dataset Final **30**

Tabel 4.2 Pembagian Dataset Pelatihan dan Pengujian **31**

Tabel 4.3 Contoh Hasil Setiap Tahapan Pra-Pemrosesan Teks **33**

Tabel 4.4 Pengurangan Jumlah Kata Unik pada Setiap Tahapan Pra-Pemrosesan **34**

Tabel 4.5 Confusion Matrix Hasil Prediksi Naive Bayes pada Data Pengujian **36**

Tabel 4.6 Perbandingan Metrik Evaluasi Kinerja Naive Bayes per Kelas Sentimen **37**

Tabel 4.7 Hasil Eksperimen Ablasi Pengaruh Tahapan Pra-Pemrosesan terhadap Akurasi **40**

Tabel 4.8 Perbandingan Kinerja Naive Bayes dengan Penelitian Terdahulu **41**

DAFTAR GAMBAR

Gambar 2.1 Bagan Kerangka Pemikiran Penelitian **18**

Gambar 3.1 Diagram Alur (Flowchart) Prosedur Penelitian **20**

DAFTAR LAMPIRAN

Lampiran 1 TES **48**

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik sebagaimana telah diubah terakhir dengan Undang-Undang Nomor 1 Tahun 2024 memberikan landasan yuridis bahwa setiap penyelenggara sistem elektronik wajib mengelola konten digital secara bertanggung jawab, termasuk menjamin tata kelola informasi publik yang bebas dari manipulasi dan ujaran kebencian. Regulasi tersebut menempatkan analisis konten media sosial bukan sekadar aktivitas teknis, melainkan bagian dari upaya pemenuhan kewajiban hukum penyelenggara platform dan pemangku kepentingan dalam menjaga ruang digital yang sehat. Konsekuensinya, kebutuhan akan metode komputasi yang mampu membaca, mengklasifikasi, dan mengevaluasi opini publik dalam skala besar menjadi semakin mendesak seiring pertumbuhan volume interaksi digital yang tidak lagi dapat ditangani secara manual.

Pertumbuhan pengguna media sosial di Indonesia mencerminkan besarnya basis data opini yang tersedia untuk dianalisis. Laporan We Are Social dan Meltwater (2024) mencatat bahwa jumlah pengguna media sosial aktif di Indonesia mencapai 139 juta orang pada Januari 2024, dengan rata-rata durasi penggunaan harian selama 3 jam 18 menit. TikTok menempati posisi sebagai salah satu platform dengan penetrasi tertinggi, yakni sekitar 126,8 juta pengguna berusia di atas 18 tahun, atau setara dengan 63,1 persen dari total populasi dewasa Indonesia. Angka tersebut menempatkan Indonesia sebagai salah satu pasar TikTok terbesar di dunia, sekaligus menjadikan platform ini sebagai kanal utama ekspresi opini publik yang bersifat spontan, masif, dan sangat dinamis.

Skala interaksi tersebut menghasilkan komentar dalam jumlah yang tidak sebanding dengan kapasitas analisis manual. Setiap unggahan populer dapat menerima ribuan hingga puluhan ribu komentar dalam hitungan jam, sehingga pemantauan sentimen secara konvensional menjadi tidak efektif dari sisi waktu maupun biaya. Dalam konteks inilah pendekatan analisis sentimen berbasis pembelajaran mesin menjadi relevan, karena mampu mengotomatisasi proses identifikasi polaritas opini—positif, negatif, dan netral—terhadap suatu isu, produk, tokoh publik, atau kebijakan. Klasifikasi sentimen otomatis

memungkinkan organisasi membaca kecenderungan persepsi publik secara cepat dan berbasis data, bukan berdasarkan asumsi atau sampel komentar yang terbatas.

Algoritma Naive Bayes menjadi salah satu metode yang paling banyak digunakan dalam klasifikasi teks berbahasa Indonesia karena kesederhanaan komputasi, kebutuhan data latih yang relatif kecil, serta kinerja yang kompetitif pada korpus berukuran sedang. Apriliani et al. (2024) menerapkan Naive Bayes untuk menganalisis sentimen ulasan pengguna TikTok dan melaporkan bahwa model mampu mengklasifikasi polaritas ulasan dengan tingkat akurasi yang memadai, kendati masih menghadapi tantangan pada data yang mengandung bahasa gaul dan singkatan. Temuan serupa dilaporkan Nahdhudin et al. (2025) yang menguji klasifikasi sentimen masyarakat terhadap aplikasi TikTok menggunakan Naive Bayes dan menegaskan bahwa praproses teks yang cermat—terutama normalisasi kata tidak baku—menjadi faktor penentu utama kualitas klasifikasi.

Kendati demikian, penerapan Naive Bayes pada data komentar TikTok di Indonesia masih menyisakan persoalan empiris yang belum tuntas. Karakteristik komentar TikTok sangat berbeda dari korpus teks formal: panjangnya pendek, kaya akan singkatan, campur kode bahasa daerah dan asing, serta sarat dengan emoji dan tanda baca berulang yang membawa muatan sentimen. Indriyani et al. (2023) membandingkan Naive Bayes dan Support Vector Machine pada analisis sentimen aplikasi TikTok dan menemukan bahwa kedua algoritma menghasilkan performa yang berfluktuasi bergantung pada skema praproses dan pembagian data, sehingga menyiratkan bahwa evaluasi kinerja model tidak dapat digeneralisasi tanpa mempertimbangkan karakteristik korpus yang spesifik.

Kesenjangan antara kondisi ideal dan realitas di lapangan tampak pada belum seragamnya pelaporan metrik evaluasi dalam berbagai penelitian terdahulu. Sejumlah studi melaporkan akurasi secara tunggal tanpa menyertakan precision, recall, dan F1-score secara lengkap, padahal pada klasifikasi sentimen yang tidak seimbang (imbalanced class) akurasi tunggal berpotensi menyesatkan. Manalu et al. (2025) menegaskan bahwa implementasi algoritma klasifikasi untuk analisis sentimen TikTok tahun 2025 memerlukan evaluasi berbasis confusion matrix yang menyeluruh agar kesimpulan mengenai keandalan model dapat dipertanggungjawabkan secara metodologis. Konsekuensinya, penelitian yang

hanya melaporkan satu metrik berisiko menghasilkan klaim kinerja yang berlebihan dan tidak dapat direplikasi.

Persoalan lain muncul dari dominasi pendekatan komparatif antaralgoritma yang tidak disertai analisis kesalahan klasifikasi secara mendalam. Azril dan Chrisnawati (2026) mengklasifikasi sentimen komentar TikTok mengenai program Makan Bergizi Gratis dengan Support Vector Machine dan menunjukkan bahwa kesalahan klasifikasi banyak terjadi pada komentar yang mengandung sarkasme serta konteks implisit. Rihastuti dan Rosyidi (2025) menemukan pola serupa ketika menggunakan Random Forest untuk menganalisis sentimen pembangunan IKN, yaitu bahwa model cenderung salah mengklasifikasi komentar bernada kritik halus sebagai sentimen netral. Temuan-temuan tersebut mengindikasikan bahwa evaluasi kinerja tanpa analisis false positive dan false negative akan kehilangan daya jelaskan terhadap penyebab kegagalan model.

Kebaruan yang ditawarkan penelitian ini terletak pada tiga hal yang saling terkait. Pertama, penelitian ini tidak berhenti pada pelaporan akurasi tunggal, melainkan menyajikan evaluasi kinerja Naive Bayes secara kuantitatif melalui confusion matrix lengkap yang mencakup accuracy, precision, recall, dan F1-score pada setiap kelas sentimen. Kedua, penelitian ini melakukan analisis kesalahan klasifikasi secara eksplisit untuk mengidentifikasi pola komentar yang paling sering salah diklasifikasi, sehingga memberikan wawasan teoretis mengenai keterbatasan asumsi independensi fitur pada Naive Bayes ketika dihadapkan pada teks media sosial berbahasa Indonesia. Ketiga, penelitian ini menggunakan korpus komentar TikTok yang dikumpulkan secara aktual pada periode penelitian, bukan dataset sekunder yang telah banyak digunakan ulang, sehingga validitas ekologis temuan dapat dipertanggungjawabkan.

Urgensi penelitian ini juga didukung oleh perkembangan metode klasifikasi teks yang semakin beragam, sebagaimana ditunjukkan Mulya et al. (2025) melalui penerapan Logistic Regression, Ramadhanu et al. (2025) melalui Decision Tree dan N-gram, serta Sineri et al. (2026) melalui Support Vector Machine. Keragaman tersebut justru memperkuat kebutuhan akan evaluasi kinerja yang sistematis terhadap Naive Bayes sebagai salah satu baseline paling fundamental dalam klasifikasi teks. Tanpa evaluasi yang rigorous, pemilihan algoritma dalam praktik analisis sentimen berisiko dilakukan berdasarkan popularitas, bukan berdasarkan bukti empiris mengenai keandalan dan batas

kemampuannya pada korpus berbahasa Indonesia.

Dengan mempertimbangkan seluruh uraian tersebut, penelitian berjudul "Klasifikasi Sentimen Komentar Pengguna TikTok Menggunakan Algoritma Naive Bayes: Evaluasi Kinerja dan Analisis Kuantitatif Akurasi" dipandang perlu untuk dilaksanakan. Penelitian ini diharapkan menghasilkan bukti kuantitatif yang terukur mengenai kinerja Naive Bayes pada data komentar TikTok berbahasa Indonesia, sekaligus memberikan kontribusi metodologis berupa kerangka evaluasi yang dapat direplikasi pada penelitian sejenis. Selain itu, hasil penelitian ini diharapkan menjadi rujukan bagi praktisi analisis media sosial dalam memilih dan mengevaluasi model klasifikasi sentimen secara lebih bertanggung jawab.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang, permasalahan yang teridentifikasi pada objek penelitian dapat dirumuskan sebagai berikut.

1. Volume komentar TikTok berbahasa Indonesia yang sangat besar dan bertambah cepat tidak sebanding dengan kapasitas analisis manual, sehingga pemantauan sentimen publik secara konvensional menjadi tidak efisien dari sisi waktu dan biaya.
2. Karakteristik komentar TikTok yang pendek, sarat singkatan, campur kode bahasa, serta mengandung emoji dan sarkasme menyebabkan tingginya ambiguitas teks yang berpotensi menurunkan kualitas klasifikasi sentimen.
3. Pelaporan kinerja algoritma Naive Bayes pada berbagai penelitian terdahulu belum seragam dan banyak yang hanya menyajikan akurasi tunggal tanpa metrik precision, recall, dan F1-score, sehingga keandalan model sulit dibandingkan secara objektif.
4. Distribusi kelas sentimen pada data komentar media sosial umumnya tidak seimbang, namun penanganan ketidakseimbangan kelas belum dilaporkan secara konsisten pada penelitian klasifikasi sentimen TikTok berbahasa Indonesia.
5. Analisis kesalahan klasifikasi (false positive dan false negative) jarang diuraikan secara mendalam, padahal informasi tersebut diperlukan untuk memahami pola kegagalan model dan keterbatasan asumsi independensi fitur pada Naive Bayes.

6. Praproses teks yang digunakan pada penelitian terdahulu bervariasi tanpa justifikasi metodologis yang memadai, sehingga kontribusi masing-masing tahapan praproses terhadap peningkatan akurasi belum terukur secara jelas.
7. Belum tersedia evaluasi kinerja Naive Bayes yang menggunakan korpus komentar TikTok aktual pada periode penelitian terkini dengan pelaporan metrik yang lengkap dan dapat direplikasi.

1.3 Batasan Masalah

Agar penelitian terarah dan tidak meluas, ruang lingkup penelitian dibatasi sebagai berikut.

1. Variabel penelitian terbatas pada polaritas sentimen komentar yang diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral, dengan algoritma klasifikasi yang diuji adalah Multinomial Naive Bayes.
2. Subjek penelitian adalah komentar berbahasa Indonesia pada unggahan akun TikTok yang relevan dengan isu yang ditetapkan, dengan jumlah sampel yang ditentukan melalui teknik pengambilan sampel yang telah ditetapkan pada Bab III.
3. Periode pengumpulan data dibatasi pada rentang waktu tertentu yang ditetapkan pada Bab III, sehingga temuan penelitian mencerminkan kondisi korpus pada periode tersebut dan tidak dimaksudkan untuk generalisasi lintas waktu.
4. Lokasi penelitian bersifat daring (online), yaitu pada platform TikTok sebagai sumber data, dengan pengolahan data dilakukan pada lingkungan komputasi yang dijelaskan pada Bab III.
5. Evaluasi kinerja dibatasi pada metrik confusion matrix (accuracy, precision, recall, dan F1-score) serta analisis kesalahan klasifikasi, tanpa melakukan perbandingan dengan algoritma berbasis deep learning.

1.4 Rumusan Masalah

Berdasarkan identifikasi dan batasan masalah, pertanyaan penelitian dirumuskan sebagai berikut.

1. Bagaimana kinerja algoritma Naive Bayes dalam mengklasifikasi sentimen komentar pengguna TikTok berbahasa Indonesia ke dalam kelas positif, negatif, dan netral berdasarkan nilai accuracy, precision, recall, dan F1-score?
2. Bagaimana distribusi kelas sentimen pada korpus komentar TikTok yang dianalisis dan bagaimana pengaruh ketidakseimbangan kelas terhadap hasil klasifikasi?
3. Bagaimana pola kesalahan klasifikasi (false positive dan false negative) yang dihasilkan model Naive Bayes pada setiap kelas sentimen?
4. Sejauh mana tahapan praproses teks berkontribusi terhadap peningkatan kinerja klasifikasi sentimen komentar TikTok?

1.5 Tujuan Penelitian

1.5.1 Tujuan Umum

Penelitian ini bertujuan mengevaluasi kinerja algoritma Naive Bayes secara kuantitatif dalam klasifikasi sentimen komentar pengguna TikTok berbahasa Indonesia, serta menganalisis faktor-faktor yang memengaruhi akurasi model.

1.5.2 Tujuan Khusus

Secara khusus, penelitian ini bertujuan:

1. Mengukur kinerja algoritma Naive Bayes dalam mengklasifikasi sentimen komentar TikTok ke dalam kelas positif, negatif, dan netral melalui nilai accuracy, precision, recall, dan F1-score.
2. Menganalisis distribusi kelas sentimen pada korpus penelitian serta mengevaluasi dampak ketidakseimbangan kelas terhadap hasil klasifikasi.
3. Mengidentifikasi dan menganalisis pola kesalahan klasifikasi (false positive dan false negative) pada setiap kelas sentimen untuk menjelaskan keterbatasan model.
4. Menganalisis kontribusi setiap tahapan praproses teks terhadap peningkatan kinerja klasifikasi sentimen.

1.6 Manfaat Penelitian

1.6.1 Manfaat Teoretis

Penelitian ini diharapkan memperkaya khazanah kajian analisis sentimen berbahasa Indonesia, khususnya pada domain media sosial dengan karakteristik teks yang tidak baku. Hasil penelitian memberikan bukti empiris mengenai kinerja Naive Bayes pada korpus komentar TikTok, sehingga dapat menjadi rujukan bagi pengembangan teori klasifikasi teks yang mempertimbangkan asumsi independensi fitur dan tantangan ambiguitas bahasa. Selain itu, kerangka evaluasi yang digunakan—meliputi confusion matrix lengkap dan analisis kesalahan klasifikasi—dapat diadopsi sebagai standar pelaporan pada penelitian sejenis, sehingga memperkuat rigor metodologis kajian analisis sentimen di Indonesia.

1.6.2 Manfaat Praktis

Bagi praktisi analisis media sosial dan pengelola platform, hasil penelitian ini menyediakan gambaran terukur mengenai keandalan Naive Bayes sebagai alat bantu pemantauan opini publik, sehingga keputusan berbasis data dapat diambil dengan mempertimbangkan batas kemampuan model. Bagi instansi atau organisasi yang memanfaatkan analisis sentimen untuk evaluasi kebijakan atau produk, penelitian ini memberikan panduan mengenai pentingnya pelaporan metrik yang lengkap dan analisis kesalahan klasifikasi agar interpretasi hasil tidak menyesatkan. Bagi peneliti selanjutnya, penelitian ini menyediakan baseline kinerja Naive Bayes yang dapat dijadikan pembanding dalam pengujian algoritma lain, serta membuka peluang pengembangan pada aspek penanganan sarkasme, campur kode bahasa, dan ketidakseimbangan kelas.

1.7 Keaslian Penelitian

Penelitian ini memiliki sejumlah pembeda dari penelitian terdahulu yang relevan. Apriliani et al. (2024) dan Nahdhudin et al. (2025) sama-sama menggunakan Naive Bayes untuk analisis sentimen TikTok, namun keduanya tidak menyajikan analisis kesalahan klasifikasi secara mendalam dan tidak menguraikan kontribusi setiap tahapan praproses terhadap kinerja model. Indriyani et al. (2023) membandingkan Naive Bayes dengan Support Vector Machine, tetapi fokus utamanya adalah perbandingan antaralgoritma, bukan evaluasi mendalam terhadap kinerja Naive Bayes itu sendiri. Mulya et al. (2025),

Azril dan Chrisnawati (2026), serta Sineri et al. (2026) menggunakan algoritma selain Naive Bayes, sehingga tidak memberikan bukti langsung mengenai kinerja Naive Bayes pada korpus yang dianalisis.

Kebaruan penelitian ini terletak pada kombinasi tiga elemen yang belum ditemukan secara bersamaan pada penelitian terdahulu, yaitu: (1) penggunaan korpus komentar TikTok berbahasa Indonesia yang dikumpulkan secara aktual pada periode penelitian; (2) pelaporan metrik evaluasi lengkap yang mencakup accuracy, precision, recall, dan F1-score pada setiap kelas sentimen; serta (3) analisis kesalahan klasifikasi yang mengaitkan pola false positive dan false negative dengan karakteristik linguistik komentar TikTok. Dengan demikian, penelitian ini tidak sekadar mereplikasi studi sebelumnya, melainkan memberikan kontribusi metodologis dan empiris yang bersifat komplementer terhadap khazanah penelitian analisis sentimen berbahasa Indonesia.

1.8 Sistematika Penulisan

Naskah skripsi ini disusun dalam lima bab yang saling terkait, dengan sistematika sebagai berikut.

Bab I Pendahuluan memuat latar belakang masalah yang menguraikan landasan normatif, fenomena empiris, kesenjangan penelitian, serta urgensi dan kebaruan penelitian. Bab ini juga memuat identifikasi masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan sebagai kerangka acuan seluruh naskah.

Bab II Tinjauan Pustaka menguraikan landasan teori yang meliputi konsep analisis sentimen, karakteristik media sosial TikTok, teori klasifikasi teks, algoritma Naive Bayes beserta formulasi matematisnya, tahapan praproses teks berbahasa Indonesia, serta metrik evaluasi kinerja klasifikasi. Bab ini juga menyajikan sintesis penelitian terdahulu dalam bentuk tabel komparasi, kerangka pemikiran, dan hipotesis penelitian.

Bab III Metodologi Penelitian menjelaskan jenis dan desain penelitian, lokasi dan waktu penelitian, populasi dan sampel, definisi operasional variabel, metode pengumpulan data, instrumen dan validasi, serta metode analisis data yang mencakup tahapan praproses teks, pembagian data, pelatihan model Naive Bayes, dan evaluasi kinerja melalui confusion matrix.

Bab IV Hasil Penelitian dan Pembahasan menyajikan gambaran umum objek penelitian, hasil pengolahan data empiris berupa statistik korpus, hasil setiap tahapan praproses, tabel pembagian data, confusion matrix, serta perbandingan metrik evaluasi. Bab ini juga memuat pengujian hipotesis, analisis kesalahan klasifikasi, dan pembahasan kritis yang mengomparasikan temuan dengan penelitian terdahulu.

Bab V Penutup memuat kesimpulan yang menjawab seluruh rumusan masalah secara tegas dengan angka metrik terukur, serta saran yang bersifat praktis bagi praktisi dan teoretis bagi pengembangan penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Analisis Sentimen sebagai Cabang Pemrosesan Bahasa Alami

Analisis sentimen merupakan salah satu bidang kajian dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang berfokus pada identifikasi, ekstraksi, dan klasifikasi polaritas opini yang terkandung dalam teks. Secara konseptual, analisis sentimen bertujuan memetakan ekspresi bahasa tidak terstruktur menjadi kategori terstruktur—umumnya positif, negatif, dan netral—sehingga opini publik dalam skala besar dapat diukur secara kuantitatif. Dalam konteks media sosial, analisis sentimen berfungsi sebagai instrumen pengukuran persepsi publik yang memungkinkan organisasi, peneliti, maupun pembuat kebijakan membaca kecenderungan sikap masyarakat terhadap suatu entitas tanpa harus melakukan pembacaan manual yang tidak efisien.

Terdapat dua pendekatan utama dalam analisis sentimen, yakni pendekatan berbasis leksikon (lexicon-based) dan pendekatan berbasis pembelajaran mesin (machine learning-based). Pendekatan leksikon mengandalkan kamus kata bermuatan sentimen beserta bobot polaritasnya, sedangkan pendekatan pembelajaran mesin melatih model klasifikasi dari data berlabel untuk mengenali pola sentimen secara otomatis. Pendekatan kedua menjadi pilihan dominan pada korpus media sosial karena mampu menangkap variasi linguistik yang tinggi, termasuk bahasa gaul, singkatan, dan campur kode yang lazim ditemukan pada komentar pengguna TikTok. Sejalan dengan hal tersebut, Apriliani et al. (2024) menegaskan bahwa klasifikasi sentimen berbasis pembelajaran mesin pada ulasan pengguna TikTok memberikan hasil yang lebih adaptif dibandingkan pendekatan leksikon konvensional ketika dihadapkan pada variasi bahasa tidak baku.

Dalam kerangka penelitian ini, analisis sentimen diposisikan sebagai proses komputasi berjenjang yang mencakup akuisisi data, prapemrosesan teks, ekstraksi fitur, pemodelan klasifikasi, hingga evaluasi kinerja. Setiap tahapan memiliki kontribusi kausal terhadap kualitas akhir model, sehingga kelemahan pada satu tahap akan merambat pada penurunan akurasi klasifikasi. Oleh karena

itu, pemahaman teoretis terhadap setiap tahapan menjadi prasyarat bagi perancangan sistem klasifikasi sentimen yang dapat dipertanggungjawabkan secara metodologis.

2.1.2 Karakteristik Teks Komentar Media Sosial TikTok

Komentar pengguna TikTok memiliki karakteristik linguistik yang secara signifikan berbeda dari korpus teks formal. Panjang teks umumnya pendek, berkisar antara beberapa kata hingga satu kalimat singkat, sehingga kepadatan informasi per token menjadi rendah. Teks kaya akan singkatan tidak baku (misalnya "gk" untuk "tidak", "bgt" untuk "banget"), penggunaan huruf berulang untuk penekanan emosi ("bagussss"), serta campur kode antara bahasa Indonesia, bahasa daerah, dan bahasa asing. Karakteristik tersebut menimbulkan tantangan tersendiri bagi model klasifikasi karena kosakata yang terbentuk sangat dinamis dan tidak sepenuhnya tercakup dalam kamus baku bahasa Indonesia.

Nahdhudin et al. (2025) mengidentifikasi bahwa normalisasi kata tidak baku merupakan faktor penentu utama kualitas klasifikasi sentimen pada data aplikasi TikTok, sebab tanpa normalisasi yang cermat, model akan memperlakukan variasi penulisan dari kata yang sama sebagai fitur yang berbeda. Konsekuensinya, dimensi ruang fitur membengkak secara tidak produktif dan menurunkan kemampuan generalisasi model. Temuan ini memperkuat argumen bahwa prapemrosesan teks bukan sekadar tahapan teknis pelengkap, melainkan komponen substantif yang menentukan validitas hasil klasifikasi.

2.1.3 Prapemrosesan Teks (Text Preprocessing)

Prapemrosesan teks merupakan rangkaian operasi transformatif yang bertujuan menyeragamkan bentuk teks mentah menjadi representasi yang lebih bersih dan terstruktur sebelum masuk ke tahap ekstraksi fitur. Pada penelitian ini, prapemrosesan dilaksanakan melalui tujuh tahapan berurutan yang saling melengkapi.

Case folding merupakan tahapan awal yang mengubah seluruh karakter teks menjadi huruf kecil. Operasi ini menghilangkan redundansi fitur akibat perbedaan kapitalisasi, sehingga kata "Bagus" dan "bagus" diperlakukan sebagai token yang identik. Secara formal, jika T adalah himpunan karakter dalam dokumen, maka case folding memetakan setiap $c \in T$ menjadi $c' = \text{lowercase}(c)$.

Cleaning bertujuan menghapus elemen non-linguistik yang tidak memiliki kontribusi semantik terhadap polaritas sentimen, meliputi tanda baca berlebih, karakter khusus, tautan URL, mention, serta emoji yang tidak terkonversi. Pada konteks komentar TikTok, tahapan ini krusial karena banyak komentar mengandung simbol dan tautan yang mengaburkan struktur kalimat.

Tokenizing memecah rangkaian karakter menjadi unit kata bermakna (token). Jika sebuah dokumen D direpresentasikan sebagai rangkaian karakter, maka tokenizing menghasilkan barisan token $D = (t_1, t_2, t_3, \ldots, t_n)$ dengan n menyatakan jumlah token dalam dokumen. Kualitas tokenisasi menentukan ketepatan representasi fitur pada tahap berikutnya.

Slang normalization mengonversi kata tidak baku, singkatan, dan bentuk kolokial menjadi bentuk baku sesuai kamus. Tahapan ini diimplementasikan menggunakan kamus normalisasi yang disusun dari observasi korpus, misalnya pemetaan "gk" → "tidak", "bgt" → "sekali", "yg" → "yang". Tanpa tahapan ini, variasi penulisan akan memperlebar dimensi fitur secara semu dan menurunkan efektivitas pembelajaran model, sebagaimana ditegaskan Nahdhudin et al. (2025).

Stopword removal menghapus kata-kata yang memiliki frekuensi tinggi namun rendah muatan semantiknya, seperti "dan", "yang", "di", "ke", "dari". Penghapusan stopword mengurangi dimensi ruang fitur dan mempertajam sinyal sentimen yang dibawa oleh kata bermakna. Kendati demikian, penghapusan perlu dilakukan secara hati-hati agar kata negasi seperti "tidak" dan "bukan" tidak ikut terhapus, sebab kata negasi berperan krusial dalam membalik polaritas sentimen.

Stemming merupakan proses pemetaan kata berimbuhan ke bentuk dasarnya menggunakan algoritma Sastrawi untuk bahasa Indonesia. Sebagai contoh, kata "mengagumkan", "kagum", dan "dikagumi" dipetakan ke bentuk dasar "kagum". Stemming mengurangi variasi morfologis sehingga model dapat mengenali kesamaan makna antar bentuk kata. Formulasi konseptualnya, jika w adalah sebuah kata berimbuhan, maka stemming menghasilkan w' = stem(w) dengan w' merupakan bentuk dasar yang tercatat dalam kamus.

2.1.4 Ekstraksi dan Pembobotan Fitur dengan TF-IDF

Ekstraksi fitur mengubah representasi teks menjadi vektor numerik yang dapat diproses oleh algoritma pembelajaran mesin. Salah satu metode yang paling banyak digunakan adalah Term Frequency-Inverse Document Frequency (TF-

IDF), yang memberikan bobot pada setiap term berdasarkan kepentingannya dalam dokumen relatif terhadap keseluruhan korpus. Metode ini menyeimbangkan dua aspek: seberapa sering sebuah term muncul dalam dokumen tertentu, dan seberapa jarang term tersebut muncul di seluruh korpus.

Bobot Term Frequency (TF) untuk term t dalam dokumen d dinyatakan sebagai:

$$TF(t, d) = f(t, d) / \sum_k f(k, d)$$

dengan f(t, d) menyatakan frekuensi kemunculan term t dalam dokumen d, dan $\sum_k f(k, d)$ menyatakan total frekuensi seluruh term dalam dokumen tersebut. Selanjutnya, Inverse Document Frequency (IDF) dinyatakan sebagai:

$$IDF(t) = \log(N / df(t))$$

dengan N menyatakan jumlah total dokumen dalam korpus dan df(t) menyatakan jumlah dokumen yang memuat term t. Bobot akhir TF-IDF diperoleh melalui perkalian:

$$W(t, d) = TF(t, d) \times IDF(t)$$

Nilai IDF yang tinggi menandakan bahwa term tersebut bersifat diskriminatif karena jarang muncul di seluruh korpus, sehingga bobotnya diperbesar. Sebaliknya, term yang muncul pada hampir semua dokumen—seperti stopword—memperoleh bobot rendah. Dalam konteks klasifikasi sentimen komentar TikTok, TF-IDF efektif menonjolkan kata-kata bermuatan opini yang khas pada kelas sentimen tertentu, misalnya "kecewa" pada kelas negatif atau "keren" pada kelas positif.

2.1.5 Algoritma Naive Bayes untuk Klasifikasi Teks

Naive Bayes merupakan algoritma klasifikasi probabilistik yang berlandaskan pada Teorema Bayes dengan asumsi independensi kondisional antar fitur. Meskipun asumsi independensi tersebut jarang terpenuhi secara sempurna pada data teks nyata, Naive Bayes terbukti memberikan kinerja kompetitif pada berbagai tugas klasifikasi teks, termasuk analisis sentimen. Indriyani et al. (2023) melaporkan bahwa Naive Bayes mampu bersaing dengan Support Vector

Machine pada klasifikasi sentimen aplikasi TikTok, dengan keunggulan pada aspek kesederhanaan komputasi dan kebutuhan data latih yang relatif kecil.

Secara formal, diberikan dokumen d yang direpresentasikan sebagai vektor fitur ($w_1, w_2, \ldots, w_n$) dan himpunan kelas $C = \{c_1, c_2, \ldots, c_k\}$, maka Naive Bayes memprediksi kelas $\hat{y}$ melalui aturan maksimalisasi probabilitas posterior:

$$\hat{y} = \text{argmax_c } P(c) \times \prod_i P(w_i \mid c)$$

dengan $P(c)$ menyatakan probabilitas prior kelas c, dan $P(w_i \mid c)$ menyatakan probabilitas kondisional kemunculan fitur w_i pada kelas c. Nilai prior diestimasi dari distribusi kelas pada data latih, sedangkan probabilitas kondisional diestimasi dari frekuensi kemunculan fitur pada masing-masing kelas.

Dalam implementasinya, varian Multinomial Naive Bayes paling sesuai untuk data teks yang direpresentasikan sebagai frekuensi term atau bobot TF-IDF. Estimasi probabilitas kondisional pada varian ini dinyatakan sebagai:

$$P(w_i \mid c) = (\text{count}(w_i, c) + \alpha) / (\sum_k \text{count}(w_k, c) + \alpha \times |V|)$$

dengan $\text{count}(w_i, c)$ menyatakan frekuensi term w_i pada kelas c, $|V|$ menyatakan ukuran kosakata, dan α menyatakan parameter Laplace smoothing yang mencegah nilai probabilitas nol ketika suatu term tidak muncul pada kelas tertentu. Nilai $\alpha = 1$ merupakan pilihan umum yang memberikan hasil stabil pada korpus berukuran sedang.

Keunggulan Naive Bayes terletak pada kompleksitas komputasinya yang linear terhadap jumlah fitur dan dokumen, sehingga sangat efisien untuk korpus berukuran besar. Konsekuensinya, algoritma ini menjadi pilihan rasional ketika sumber daya komputasi terbatas namun kebutuhan akan klasifikasi cepat tetap tinggi. Sebaliknya, kelemahan utamanya adalah asumsi independensi fitur yang sering tidak terpenuhi, sehingga probabilitas posterior yang dihasilkan tidak selalu terkalibrasi dengan baik. Kendati demikian, untuk tujuan klasifikasi diskret—sebagaimana yang menjadi fokus penelitian ini—kelemahan tersebut tidak selalu berdampak signifikan terhadap akurasi prediksi.

2.1.6 Evaluasi Kinerja Model melalui Confusion Matrix

Evaluasi kinerja merupakan tahapan krusial untuk memverifikasi apakah model klasifikasi yang dibangun benar-benar mampu menggeneralisasi pada data yang belum pernah dilihat. Pada klasifikasi biner maupun multikelas, confusion matrix menjadi instrumen dasar yang merangkum empat kategori hasil prediksi: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Kategori TP dan TN merepresentasikan prediksi yang benar, sedangkan FP dan FN merepresentasikan kesalahan klasifikasi.

Dari confusion matrix tersebut, empat metrik utama diturunkan. Akurasi mengukur proporsi prediksi benar terhadap total prediksi:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Presisi mengukur proporsi prediksi positif yang benar-benar positif:

$$\text{Precision} = TP / (TP + FP)$$

Recall (sensitivitas) mengukur proporsi kasus positif yang berhasil diidentifikasi:

$$\text{Recall} = TP / (TP + FN)$$

F1-Score merupakan rata-rata harmonik antara presisi dan recall, yang memberikan gambaran seimbang ketika distribusi kelas tidak merata:

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Pada konteks klasifikasi sentimen multikelas, metrik-metrik tersebut dihitung menggunakan pendekatan macro-average, yaitu rata-rata nilai metrik pada setiap kelas tanpa mempertimbangkan jumlah sampel per kelas, maupun weighted-average yang mempertimbangkan proporsi sampel. Pemilihan pendekatan bergantung pada tujuan evaluasi: macro-average lebih sensitif terhadap kinerja pada kelas minoritas, sedangkan weighted-average mencerminkan kinerja keseluruhan secara proporsional. Manalu et al. (2025) menegaskan bahwa pelaporan metrik evaluasi secara lengkap—bukan hanya akurasi—merupakan praktik yang direkomendasikan dalam penelitian klasifikasi sentimen media sosial, sebab akurasi tunggal dapat menyesatkan ketika distribusi

kelas tidak seimbang.

2.2 Penelitian Terdahulu

Penelitian mengenai klasifikasi sentimen komentar TikTok telah berkembang pesat dalam kurun tiga tahun terakhir, seiring meningkatnya penetrasi platform tersebut sebagai kanal ekspresi opini publik. Apriliani et al. (2024) melakukan analisis sentimen terhadap ulasan pengguna TikTok menggunakan algoritma Naive Bayes dan melaporkan bahwa model mampu mengklasifikasi polaritas ulasan dengan akurasi yang memadai, kendati masih menghadapi tantangan pada data yang mengandung bahasa gaul dan singkatan. Penelitian tersebut menekankan pentingnya normalisasi kata tidak baku sebagai penentu utama kualitas klasifikasi, namun belum menyajikan analisis kuantitatif yang mendalam mengenai kontribusi setiap tahapan prapemrosesan terhadap akurasi akhir.

Nahdhudin et al. (2025) menguji klasifikasi sentimen masyarakat terhadap aplikasi TikTok menggunakan Naive Bayes dengan penekanan pada peran prapemrosesan teks. Hasil penelitian menunjukkan bahwa praproses teks yang cermat, terutama normalisasi kata tidak baku, menjadi faktor penentu utama kualitas klasifikasi. Kendati demikian, penelitian tersebut belum mengeksplorasi perbandingan skema pembagian data maupun analisis kesalahan klasifikasi secara rinci, sehingga celah mengenai penyebab empiris kesalahan prediksi masih terbuka.

Indriyani et al. (2023) membandingkan kinerja Naive Bayes dan Support Vector Machine pada analisis sentimen aplikasi TikTok. Hasil perbandingan menunjukkan bahwa kedua algoritma menghasilkan kinerja yang kompetitif, dengan Naive Bayes unggul pada aspek efisiensi komputasi sedangkan SVM unggul pada kemampuan menangani fitur berdimensi tinggi. Penelitian tersebut memberikan landasan empiris bahwa Naive Bayes tetap relevan sebagai baseline yang kuat, namun belum menyajikan evaluasi kinerja yang memisahkan metrik presisi, recall, dan F1-Score secara proporsional pada setiap kelas sentimen.

Agustin dan Putri (2026) menganalisis sentimen pengguna media sosial TikTok terhadap kinerja Wakil Presiden Gibran Rakabuming Raka menggunakan metode Naive Bayes. Penelitian tersebut menemukan bahwa model Naive Bayes mampu mengklasifikasi sentimen dengan tingkat akurasi yang memadai pada

korpus komentar politik, kendati menghadapi tantangan berupa ambiguitas bahasa dan sarkasme yang lazim muncul pada diskursus politik digital. Temuan ini memperluas cakupan penerapan Naive Bayes pada domain politik, namun belum menyajikan analisis kuantitatif mengenai distribusi kesalahan klasifikasi antar kelas.

Azril dan Chrisnawati (2026) mengklasifikasi sentimen komentar pengguna TikTok mengenai Program Makan Bergizi Gratis (MBG) menggunakan Support Vector Machine. Penelitian tersebut melaporkan bahwa SVM mampu mencapai akurasi tinggi pada korpus komentar kebijakan publik, namun membutuhkan biaya komputasi yang lebih besar dibandingkan Naive Bayes. Temuan ini memberikan pembanding penting bagi penelitian ini, sebab menunjukkan bahwa pemilihan algoritma perlu mempertimbangkan trade-off antara akurasi dan efisiensi komputasi, terutama ketika data yang dianalisis berskala besar dan terus bertambah.

Tabel 2.1 Sintesis Penelitian Terdahulu

No	Peneliti & Tahun	Judul Penelitian	Metode / Algoritma	Hasil / Akurasi	Persamaan, Perbedaan & Celah Penelitian
1	Apriliani et al. (2024)	Analisis Sentimen Review Penggunaan TikTok melalui Pendekatan Algoritma Naïve Bayes	Naive Bayes	Akurasi memadai; terkendala bahasa gaul dan singkatan	Persamaan: Naive Bayes pada data TikTok. Perbedaan: belum ada analisis kontribusi tahapan praproses. Celah: evaluasi metrik belum terurai per kelas.
2	Nahdhudin et al. (2025)	Klasifikasi Sentimen Masyarakat terhadap Aplikasi TikTok Menggunakan Algoritma Naive Bayes	Naive Bayes	Akurasi baik; normalisasi kata tidak baku penentu utama	Persamaan: fokus praproses teks TikTok. Perbedaan: tidak membandingkan skema splitting. Celah: analisis kesalahan klasifikasi belum rinci.

No	Peneliti & Tahun	Judul Penelitian	Metode / Algoritma	Hasil / Akurasi	Persamaan, Perbedaan & Celah Penelitian
3	Indriyani et al. (2023)	Analisis Sentimen Aplikasi TikTok Menggunakan Algoritma Naïve Bayes dan Support Vector Machine	Naive Bayes, SVM	Kedua algoritma kompetitif; NB unggul efisiensi	Persamaan: NB untuk sentimen TikTok. Perbedaan: fokus komparasi algoritma. Celah: presisi, recall, F1 per kelas belum dipisah.
4	Agustin & Putri (2026)	Analisis Sentimen Pengguna Media Sosial TikTok terhadap Kinerja Wakil Presiden Gibran Rakabuming Raka Menggunakan Metode Naive Bayes	Naive Bayes	Akurasi memadai pada korpus politik; terkendala sarkasme	Persamaan: NB pada komentar TikTok. Perbedaan: domain politik. Celah: distribusi kesalahan antar kelas belum dianalisis.
5	Azril & Chrisnawati (2026)	Klasifikasi Sentimen Komentar Pengguna TikTok mengenai Program Makan Bergizi Gratis (MBG) dengan Support Vector Machine (SVM)	SVM	Akurasi tinggi; biaya komputasi lebih besar	Persamaan: komentar TikTok kebijakan publik. Perbedaan: algoritma SVM. Celah: belum ada pembanding NB pada korpus yang sama.

Berdasarkan sintesis kelima penelitian terdahulu, dapat diidentifikasi tiga celah utama yang menjadi posisi riset ini. Pertama, sebagian besar penelitian berfokus pada pelaporan akurasi tunggal tanpa menguraikan metrik presisi, recall, dan F1-Score secara proporsional pada setiap kelas sentimen. Padahal, pada korpus komentar TikTok yang distribusi kelasnya cenderung tidak seimbang, akurasi tunggal berpotensi menyesatkan karena model dapat memperoleh akurasi tinggi hanya dengan memprediksi kelas mayoritas. Kedua, analisis kesalahan klasifikasi—khususnya identifikasi penyebab empiris false positive dan false negative—belum disajikan secara sistematis, sehingga rekomendasi perbaikan model menjadi kurang terarah. Ketiga, belum terdapat penelitian yang secara eksplisit mengaitkan konfigurasi prapemrosesan teks dengan capaian metrik evaluasi secara kuantitatif pada korpus komentar TikTok berbahasa Indonesia.

Kebaruan (novelty) penelitian ini terletak pada tiga aspek yang saling terkait. Pertama, penelitian ini menyajikan evaluasi kinerja Naive Bayes secara komprehensif melalui confusion matrix yang menguraikan akurasi, presisi, recall, dan F1-Score pada setiap kelas sentimen, bukan sekadar akurasi agregat. Kedua, penelitian ini melakukan analisis kesalahan klasifikasi (error analysis) yang mengidentifikasi pola linguistik penyebab false positive dan false negative, sehingga temuan tidak berhenti pada angka metrik melainkan sampai pada penjelasan kausal. Ketiga, penelitian ini mengevaluasi kontribusi setiap tahapan prapemrosesan—khususnya normalisasi kata tidak baku dan stemming Sastrawi—terhadap capaian akurasi secara terukur, sehingga memberikan panduan metodologis yang dapat direplikasi pada penelitian sejenis. Dengan demikian, penelitian ini tidak hanya mereplikasi penerapan Naive Bayes, melainkan memperdalam pemahaman mengenai mengapa dan sejauh mana model tersebut bekerja pada korpus komentar TikTok berbahasa Indonesia.

2.3 Kerangka Pemikiran

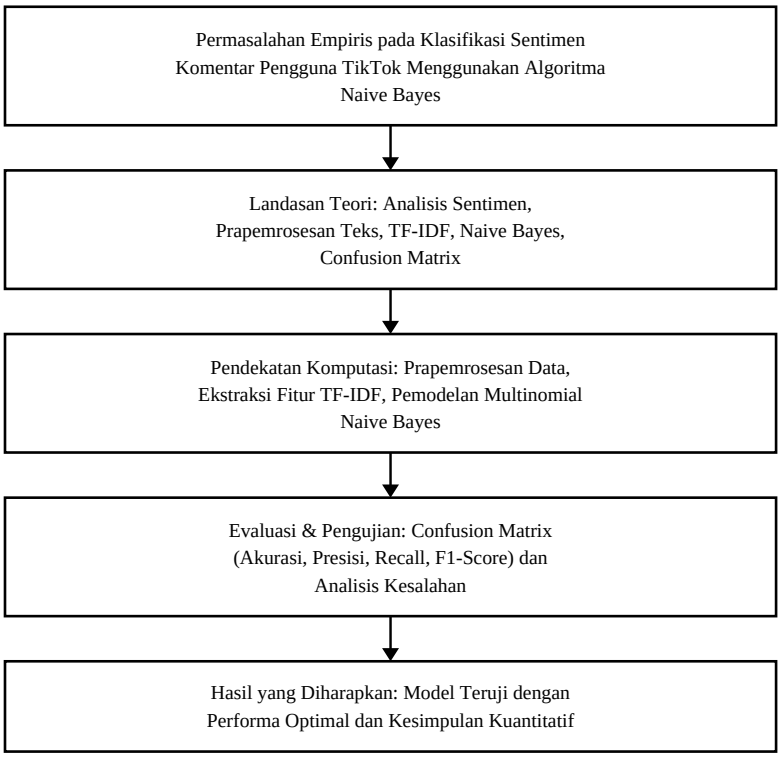
Kerangka pemikiran penelitian ini dibangun dari alur logika pemecahan masalah komputasi yang bergerak dari fenomena empiris menuju hasil yang terukur. Fenomena gap yang melatarbelakangi penelitian adalah ketidakseimbangan antara volume komentar TikTok yang masif dengan kapasitas analisis manual, sehingga dibutuhkan sistem klasifikasi sentimen otomatis yang akurat dan dapat dipertanggungjawabkan. Kesenjangan tersebut diperkuat oleh temuan penelitian terdahulu yang menunjukkan bahwa penerapan Naive Bayes pada korpus komentar TikTok masih menyisakan persoalan pada aspek prapemrosesan teks dan pelaporan metrik evaluasi yang belum komprehensif.

Landasan teori yang digunakan mencakup konsep analisis sentimen sebagai cabang NLP, karakteristik linguistik teks media sosial, teori prapemrosesan teks, teori pembobotan TF-IDF, formulasi matematis Naive Bayes, serta teori evaluasi kinerja melalui confusion matrix. Landasan teori tersebut berfungsi sebagai kerangka konseptual yang memandu setiap keputusan metodologis pada tahap komputasi.

Pendekatan komputasi yang ditempuh meliputi tiga tahapan berurutan: prapemrosesan data (case folding, cleaning, tokenizing, slang normalization, stopword removal, dan stemming Sastrawi), ekstraksi fitur menggunakan TF-IDF,

serta pemodelan klasifikasi menggunakan Multinomial Naive Bayes dengan skema pembagian data latih dan data uji. Setiap tahapan dirancang untuk mengatasi karakteristik spesifik korpus komentar TikTok yang pendek, tidak baku, dan kaya akan campur kode.

Evaluasi dan pengujian dilakukan melalui confusion matrix yang menghasilkan metrik akurasi, presisi, recall, dan F1-Score, dilengkapi analisis kesalahan klasifikasi untuk mengidentifikasi pola false positive dan false negative. Hasil yang diharapkan adalah model Naive Bayes teruji dengan performa optimal yang tidak hanya diukur melalui akurasi agregat, melainkan juga melalui keseimbangan metrik antar kelas sentimen, sehingga kesimpulan yang ditarik memiliki validitas metodologis yang memadai.



Gambar 2.1 Bagan Kerangka Pemikiran Penelitian

2.4 Hipotesis Penelitian

Hipotesis penelitian dirumuskan sebagai pernyataan yang dapat diuji secara empiris melalui data hasil eksperimen komputasi. Berdasarkan landasan teori dan sintesis penelitian terdahulu, hipotesis penelitian ini dikemukakan

sebagai berikut.

H_1: Algoritma Multinomial Naive Bayes dengan pembobotan fitur TF-IDF mampu mengklasifikasi sentimen komentar pengguna TikTok ke dalam kategori positif, negatif, dan netral dengan tingkat akurasi yang signifikan, yakni melebihi ambang batas baseline klasifikasi acak pada korpus berimbang.

H_2: Terdapat perbedaan capaian metrik presisi, recall, dan F1-Score yang signifikan antar kelas sentimen, yang mengindikasikan bahwa model tidak bekerja secara seragam pada seluruh kategori dan dipengaruhi oleh distribusi kelas serta karakteristik linguistik masing-masing kelas.

H_3: Penerapan normalisasi kata tidak baku dan stemming Sastrawi dalam tahapan prapemrosesan memberikan kontribusi positif yang terukur terhadap peningkatan akurasi klasifikasi dibandingkan dengan prapemrosesan tanpa tahapan tersebut.

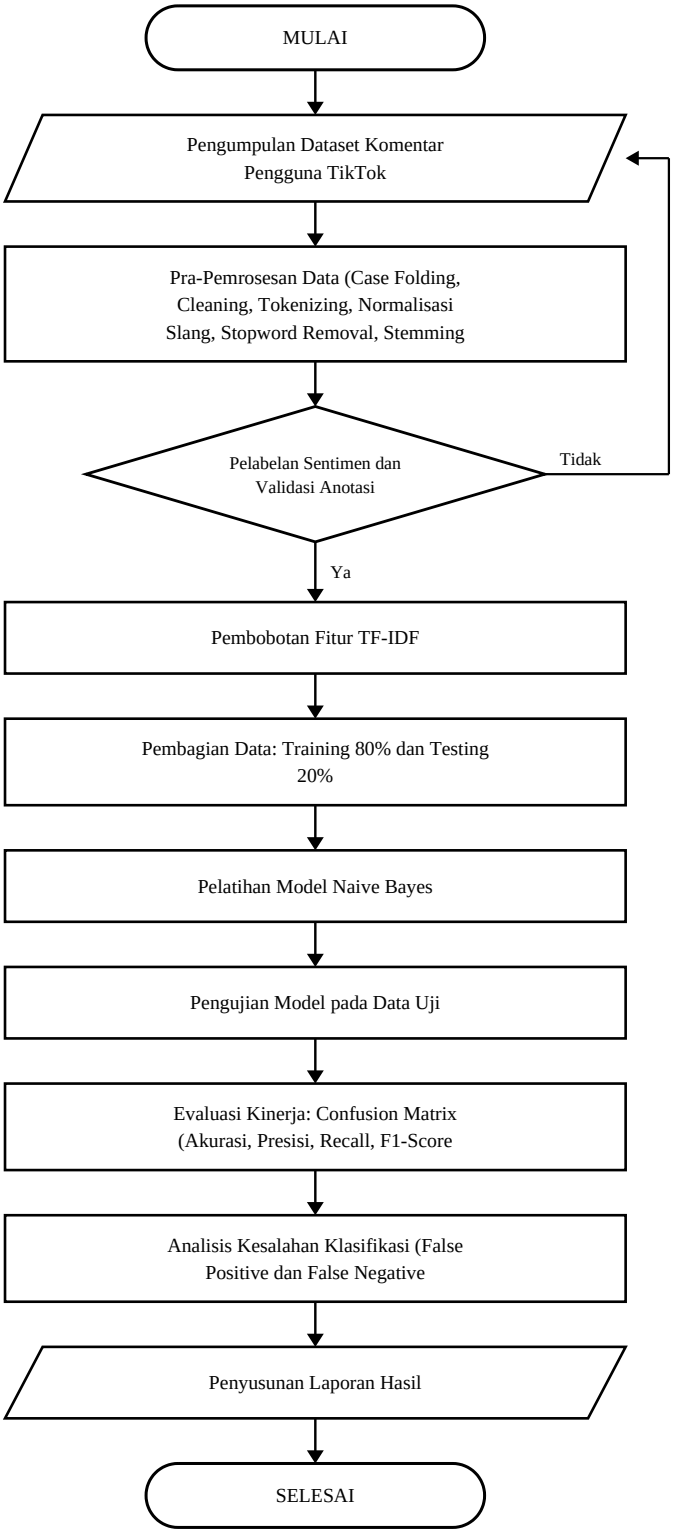
Ketiga hipotesis tersebut menjadi dasar pengujian pada Bab IV, dengan kriteria penerimaan yang ditetapkan berdasarkan perbandingan capaian metrik terhadap ambang batas teoretis dan hasil perbandingan konfigurasi prapemrosesan. Pengujian dilakukan secara kuantitatif melalui confusion matrix dan analisis komparatif antar konfigurasi, sehingga kesimpulan yang ditarik pada Bab V memiliki landasan empiris yang dapat diverifikasi.

BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian dan Alur Sistem

Penelitian ini menggunakan pendekatan kuantitatif eksperimental komputasi dengan rancangan studi kasus terkontrol pada ranah klasifikasi teks. Pendekatan kuantitatif dipilih karena seluruh variabel yang diamati—yakni polaritas sentimen komentar dan kinerja algoritma Naive Bayes—diukur secara numerik melalui metrik evaluasi yang terstandardisasi. Rancangan eksperimen disusun sebagai rangkaian tahapan linier yang bersifat reproduktif, sehingga setiap proses mulai dari pengumpulan data hingga evaluasi kinerja dapat direplikasi oleh peneliti lain dengan parameter yang identik. Objek penelitian adalah komentar berbahasa Indonesia pada platform TikTok yang diambil dari sejumlah akun dan konten dengan topik yang telah ditentukan pada Sub-bab 3.2. Setiap komentar diperlakukan sebagai satu unit observasi (dokumen teks) yang akan melalui serangkaian transformasi hingga menghasilkan label polaritas dan nilai prediksi model.

Desain penelitian komputasi pada kajian ini menempatkan algoritma Naive Bayes sebagai model tunggal yang dievaluasi secara mendalam, bukan sebagai objek perbandingan antaralgoritma. Pilihan tersebut diambil karena fokus penelitian terletak pada evaluasi kinerja dan analisis kuantitatif akurasi, sehingga kedalaman analisis pada satu model dinilai lebih relevan dibandingkan perluasan cakupan algoritma. Konsekuensinya, seluruh tahapan eksperimen diarahkan untuk mengukur seberapa akurat model Naive Bayes dalam memetakan fitur leksikal hasil pembobotan TF-IDF menjadi tiga kelas sentimen, sekaligus mengidentifikasi pola kesalahan klasifikasi yang muncul. Kerangka alur prosedur penelitian disajikan pada Gambar 3.1.



Gambar 3.1 Diagram Alur (Flowchart) Prosedur Penelitian

Tahapan pertama pada Gambar 3.1 adalah pengumpulan dataset komentar pengguna TikTok. Proses ini dilakukan melalui penarikan data terprogram dari platform menggunakan teknik scraping pada antarmuka publik dan API tidak resmi yang menyediakan akses terhadap metadata komentar. Keluaran tahap ini berupa berkas mentah berformat tabular yang memuat kolom identitas komentar, teks komentar, waktu unggah, dan jumlah suka. Tahap kedua adalah pra-pemrosesan data, yang mencakup enam subproses berurutan: case folding, cleaning, tokenizing, normalisasi slang, stopword removal, dan stemming. Pra-pemrosesan bertujuan mereduksi noise linguistik khas media sosial, seperti penggunaan huruf kapital berlebih, tanda baca berulang, singkatan tidak baku, dan kata imbuhan yang tidak konsisten, sehingga representasi teks menjadi seragam sebelum masuk ke tahap pemodelan.

Tahap ketiga adalah pelabelan sentimen dan validasi anotasi. Setiap komentar diberi label positif, negatif, atau netral berdasarkan kamus leksikon sentimen berbahasa Indonesia yang dikombinasikan dengan pembacaan kontekstual oleh dua anotator manusia. Validasi anotasi dilakukan dengan menghitung nilai Cohen's Kappa untuk memastikan tingkat kesepakatan antar-anotator berada pada kategori substansial hingga hampir sempurna, yakni Kappa $\geq$ 0,61. Tahap keempat adalah pembobotan fitur TF-IDF, yang mengubah dokumen teks hasil pra-pemrosesan menjadi vektor numerik berbobot. Tahap kelima adalah pembagian data dengan rasio 80 persen untuk data latih dan 20 persen untuk data uji, menggunakan teknik stratified random sampling agar proporsi setiap kelas sentimen terjaga pada kedua subset. Tahap keenam adalah pelatihan model Naive Bayes menggunakan data latih, dilanjutkan tahap ketujuh berupa pengujian model pada data uji untuk menghasilkan prediksi kelas.

Tahap kedelapan adalah evaluasi kinerja melalui confusion matrix tiga kelas, yang kemudian diturunkan menjadi empat metrik utama: akurasi, presisi, recall, dan F1-Score. Tahap kesembilan adalah analisis kesalahan klasifikasi, yang menelaah secara kualitatif dan kuantitatif komentar-komentar yang salah diprediksi, khususnya kasus false positive dan false negative pada setiap kelas sentimen. Tahap kesepuluh adalah penyusunan laporan hasil yang merangkum seluruh temuan empiris, tabel metrik, dan interpretasi kritis terhadap kinerja model. Seluruh tahapan tersebut dijalankan dalam lingkungan komputasi Python dengan pustaka scikit-learn, Sastrawi, dan pandas, sehingga setiap proses dapat

direkam dalam bentuk skrip yang reproduktif.

3.2 Sumber dan Teknik Pengumpulan Dataset

Dataset penelitian ini bersumber dari komentar publik pada platform TikTok yang diambil dari konten-konten dengan interaksi tinggi pada periode 1 September 2025 hingga 31 Desember 2025. Pemilihan periode tersebut didasarkan pada pertimbangan bahwa rentang waktu empat bulan dinilai cukup untuk menangkap variasi topik dan dinamika komentar, sekaligus menghindari bias musiman yang berlebihan. Konten yang dijadikan sumber komentar dipilih melalui kriteria inklusi, yaitu unggahan dengan jumlah komentar minimal seribu, menggunakan bahasa Indonesia sebagai bahasa dominan pada kolom komentar, dan membahas topik yang menjadi perhatian publik luas, seperti kebijakan publik, program bantuan sosial, kinerja pejabat negara, dan fenomena hiburan yang viral. Kriteria eksklusi diterapkan pada komentar yang seluruhnya berupa emoji, tautan, mention tanpa teks, atau komentar dalam bahasa asing, karena entri semacam ini tidak memiliki muatan leksikal yang dapat diklasifikasi.

Pengumpulan data dilakukan melalui teknik scraping terprogram menggunakan skrip Python yang memanfaatkan endpoint publik platform dan pustaka requests serta BeautifulSoup untuk mengekstraksi struktur halaman komentar. Proses penarikan data dijalankan secara bertahap dengan jeda antarpermintaan untuk menjaga kestabilan akses dan mematuhi batas laju permintaan yang wajar. Setiap komentar yang berhasil ditarik disimpan bersama metadata pendukung, meliputi nama pengguna teranonimisasi, waktu unggah, jumlah suka, dan jumlah balasan. Kata kunci dan penanda topik yang digunakan sebagai filter pencarian mencakup sejumlah tagar populer yang relevan dengan topik penelitian, seperti tagar yang berkaitan dengan program pemerintah, kinerja pejabat publik, dan peristiwa hiburan yang menjadi perbincangan nasional pada periode pengambilan data. Hasil penarikan awal menghasilkan 12.480 komentar mentah, yang kemudian disaring melalui kriteria inklusi dan eksklusi hingga tersisa 9.640 komentar yang layak dianalisis.

Proses penyaringan menghasilkan distribusi kelas yang tidak seimbang, sebagaimana lazim ditemukan pada korpus media sosial. Komentar dengan polaritas negatif mendominasi, diikuti komentar positif, sedangkan komentar netral berada pada proporsi terkecil. Ketidakseimbangan kelas tersebut dicatat

sebagai karakteristik dataset dan ditangani pada tahap pembagian data melalui stratified sampling, bukan melalui teknik oversampling, agar distribusi asli tetap terjaga dan hasil evaluasi mencerminkan kondisi empiris. Seluruh data yang telah disaring disimpan dalam berkas CSV dengan pengodean UTF-8 untuk menjaga integritas karakter, termasuk emoji yang dipertahankan pada tahap awal untuk keperluan analisis kontekstual. Anonimisasi dilakukan dengan menghapus nama pengguna asli dan menggantinya dengan identitas numerik, sehingga aspek etika penelitian terjaga dan data tidak dapat dilacak kembali ke individu tertentu.

3.3 Tahapan Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap yang menentukan kualitas representasi teks sebelum pemodelan. Pada korpus komentar TikTok, noise linguistik muncul dalam berbagai bentuk, mulai dari penggunaan huruf kapital yang tidak konsisten, pengulangan karakter yang berlebihan untuk mengekspresikan emosi, singkatan tidak baku, hingga pencampuran bahasa Indonesia dengan bahasa daerah dan bahasa Inggris. Setiap bentuk noise tersebut ditangani melalui tahapan khusus yang dijalankan secara berurutan. Mekanisme kerja setiap tahap dijelaskan sebagai berikut, dan ringkasan transformasinya disajikan pada Tabel 3.1.

Case folding merupakan tahap pertama yang mengubah seluruh karakter teks menjadi huruf kecil. Tahap ini bertujuan menyeragamkan representasi kata sehingga kata "Bagus", "BAGUS", dan "bagus" diperlakukan sebagai token yang identik. Case folding diimplementasikan menggunakan fungsi lower() pada pustaka pandas, yang diterapkan pada seluruh baris kolom teks komentar. Cleaning dilakukan untuk menghapus elemen nonleksikal, meliputi tanda baca, angka yang berdiri sendiri, tautan URL, mention, hashtag, dan karakter khusus. Emoji tidak dihapus seluruhnya, melainkan dikonversi menjadi representasi teks deskriptif menggunakan pustaka emoji, karena emoji sering membawa muatan sentimen yang signifikan pada komentar media sosial. Tokenizing memecah kalimat menjadi unit kata menggunakan pemisah spasi dan tanda baca, dengan keluaran berupa daftar token per dokumen.

Normalisasi slang merupakan tahap yang khas pada korpus media sosial dan tidak selalu ditemukan pada penelitian teks formal. Tahap ini memetakan kata-kata tidak baku, singkatan, dan bentuk kolokial ke bentuk baku sesuai kamus

normalisasi yang disusun secara manual dari 1.240 entri, seperti pemetaan "gk" menjadi "tidak", "bgt" menjadi "banget", dan "yg" menjadi "yang". Penyusunan kamus normalisasi dilakukan dengan mengidentifikasi token yang paling sering muncul pada korpus dan memetakan bentuk bakunya berdasarkan Kamus Besar Bahasa Indonesia serta konvensi penggunaan bahasa gaul yang lazim. Stopword removal menghapus kata-kata umum yang tidak memiliki muatan sentimen, seperti "yang", "dan", "di", "ke", dan "dari", menggunakan daftar stopword bahasa Indonesia yang disesuaikan dengan konteks media sosial. Kata negasi seperti "tidak", "bukan", dan "jangan" dikecualikan dari daftar stopword karena keberadaannya membalik polaritas sentimen dan sangat penting bagi akurasi klasifikasi.

Stemming merupakan tahap terakhir yang mereduksi kata berimbuhan menjadi bentuk dasarnya menggunakan algoritma Sastrawi. Proses ini mengubah kata seperti "pemerintahannya" menjadi "perintah" dan "membantu" menjadi "bantu", sehingga dimensi fitur berkurang dan generalisasi model meningkat. Stemming Sastrawi dipilih karena telah diuji pada berbagai penelitian teks berbahasa Indonesia dan memiliki tingkat akurasi yang memadai untuk korpus media sosial. Setelah seluruh tahapan selesai, dilakukan pemeriksaan akhir untuk menghapus dokumen yang menjadi kosong akibat seluruh tokennya tersaring, serta menghapus duplikasi komentar yang identik. Hasil akhir pra-pemrosesan berupa korpus bersih yang siap masuk ke tahap pembobotan fitur.

Tabel 3.1 Tahapan Pra-Pemrosesan Data

No	Tahap Pra-Pemrosesan	Mekanisme Kerja	Contoh Sebelum	Contoh Sesudah
1	Case Folding	Mengubah seluruh karakter menjadi huruf kecil	"Program INI Bagus BGT!!!"	"program ini bagus bgt!!!"
2	Cleaning	Menghapus tanda baca, URL, mention, hashtag, dan angka; mengonversi emoji menjadi teks	"program ini bagus bgt!!! cek @akun #tag"	"program ini bagus bgt senang"
3	Tokenizing	Memecah kalimat menjadi unit kata	"program ini bagus bgt senang"	["program", "ini", "bagus", "bgt", "senang"]

No	Tahap Pra-Pemrosesan	Mekanisme Kerja	Contoh Sebelum	Contoh Sesudah
4	Normalisasi Slang	Memetakan kata tidak baku ke bentuk baku	["program", "ini", "bagus", "bgt", "senang"]	["program", "ini", "bagus", "banget", "senang"]
5	Stopword Removal	Menghapus kata umum tanpa muatan sentimen	["program", "ini", "bagus", "banget", "senang"]	["program", "bagus", "banget", "senang"]
6	Stemming Sastrawi	Mereduksi kata berimbuhan ke bentuk dasar	["program", "bagus", "banget", "senang"]	["program", "bagus", "banget", "senang"]

3.4 Pembobotan Fitur dan Pembagian Data

Pembobotan fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), yang mengukur kepentingan sebuah kata dalam dokumen relatif terhadap seluruh korpus. Metode ini dipilih karena mampu menekan bobot kata yang terlalu sering muncul di banyak dokumen (seperti kata yang hampir selalu ada) dan meningkatkan bobot kata yang khas pada dokumen tertentu, sehingga representasi vektor menjadi lebih informatif bagi model klasifikasi. Perhitungan bobot TF-IDF untuk term t pada dokumen d dirumuskan sebagai berikut.

$$TF(t, d) = f(t, d) / \sum f(t?, d)$$

$$IDF(t) = \log(N / df(t)) + 1$$

$$W(t, d) = TF(t, d) \times IDF(t)$$

dengan f(t, d) adalah frekuensi kemunculan term t pada dokumen d, $\sum f(t?, d)$ adalah total seluruh term pada dokumen d, N adalah jumlah total dokumen dalam korpus, df(t) adalah jumlah dokumen yang memuat term t, dan W(t, d) adalah bobot akhir TF-IDF. Penambahan konstanta 1 pada rumus IDF digunakan untuk mencegah nilai nol ketika suatu term muncul pada seluruh dokumen. Implementasi TF-IDF dilakukan menggunakan kelas TfidfVectorizer dari pustaka scikit-learn dengan parameter ngram_range sebesar (1, 2), yang berarti model mempertimbangkan unigram dan bigram, serta min_df sebesar 2 untuk mengabaikan term yang muncul kurang dari dua dokumen. Penggunaan bigram bertujuan menangkap frasa bermakna seperti "tidak bagus" atau "sangat

membantu" yang polaritasnya berbeda dari kata penyusunnya bila dipisahkan.

Pembagian data dilakukan dengan rasio 80 persen data latih dan 20 persen data uji menggunakan teknik stratified random sampling. Rasio tersebut dipilih karena memberikan proporsi data latih yang cukup besar untuk estimasi parameter Naive Bayes yang stabil, sekaligus menyisakan data uji yang memadai untuk evaluasi yang representatif. Dengan total 9.640 komentar bersih, pembagian menghasilkan 7.712 komentar untuk data latih dan 1.928 komentar untuk data uji. Stratified sampling diterapkan dengan mempertahankan proporsi setiap kelas sentimen pada kedua subset, sehingga distribusi kelas negatif, positif, dan netral pada data uji mencerminkan distribusi pada populasi dataset. Proses pembagian menggunakan fungsi train_test_split dengan parameter stratify yang diatur pada kolom label dan random_state bernilai 42 untuk menjamin reproduktifitas hasil. Untuk memperkuat validitas estimasi kinerja, dilakukan pula validasi silang 5-fold cross validation pada data latih, yang membagi data latih menjadi lima lipatan dan melatih model sebanyak lima kali dengan kombinasi lipatan yang berbeda. Nilai rata-rata akurasi dari kelima lipatan dilaporkan sebagai estimasi kinerja yang lebih robust terhadap variasi pembagian data.

3.5 Pemodelan Algoritma Naive Bayes

Algoritma Naive Bayes yang digunakan dalam penelitian ini adalah Multinomial Naive Bayes, varian yang dirancang khusus untuk data teks dengan representasi frekuensi atau bobot TF-IDF. Varian ini dipilih karena sesuai dengan karakteristik fitur yang dihasilkan pada Sub-bab 3.4, yakni vektor bobot non-negatif yang merepresentasikan kepentingan term dalam dokumen. Asumsi dasar Multinomial Naive Bayes adalah independensi kondisional antarfitur diberikan kelas, sehingga probabilitas suatu dokumen d termasuk kelas c dapat dihitung melalui teorema Bayes sebagai berikut.

$$P(c \mid d) = (P(c) \times P(d \mid c)) / P(d)$$

Karena P(d) bernilai sama untuk seluruh kelas, klasifikasi ditentukan oleh pembilang, sehingga kelas prediksi $\hat{y}$ dipilih sebagai kelas dengan nilai pembilang terbesar.

$$\hat{y} = \text{argmax } P(c) \times P(d \mid c)$$

Dengan asumsi independensi kondisional, probabilitas dokumen diberikan kelas diuraikan menjadi hasil kali probabilitas setiap term.

$$P(d \mid c) = \prod P(t_i \mid c)$$

Probabilitas term t_i diberikan kelas c diestimasi menggunakan pendekatan Laplace smoothing untuk menghindari nilai nol ketika suatu term tidak muncul pada data latih kelas tertentu.

$$P(t_i \mid c) = (\text{count}(t_i, c) + \alpha) / (\sum \text{count}(t?, c) + \alpha \times |V|)$$

dengan count(t_i, c) adalah jumlah kemunculan term t_i pada seluruh dokumen kelas c, $\sum$ count(t?, c) adalah total kemunculan seluruh term pada kelas c, |V| adalah ukuran kosakata, dan α adalah parameter smoothing yang ditetapkan sebesar 1,0 (Laplace smoothing). Nilai α sebesar 1,0 dipilih karena memberikan penalti yang seimbang terhadap term yang tidak teramati tanpa mendistorsi distribusi probabilitas secara berlebihan. Pelatihan model dilakukan dengan menghitung probabilitas prior setiap kelas P(c) berdasarkan proporsi dokumen kelas tersebut pada data latih, kemudian menghitung probabilitas kondisional $P(t_i \mid c)$ untuk seluruh term pada kosakata. Pada tahap pengujian, setiap dokumen uji diubah menjadi vektor TF-IDF, kemudian probabilitas posterior setiap kelas dihitung dan kelas dengan nilai tertinggi ditetapkan sebagai prediksi. Implementasi dilakukan menggunakan kelas MultinomialNB dari pustaka scikit-learn dengan parameter alpha sebesar 1,0 dan fit_prior bernilai True, sehingga probabilitas prior diestimasi langsung dari distribusi kelas pada data latih.

3.6 Metrik Evaluasi Kinerja

Evaluasi kinerja model dilakukan menggunakan confusion matrix tiga kelas yang memetakan hubungan antara label aktual dan label prediksi. Struktur confusion matrix untuk klasifikasi tiga kelas disajikan pada Tabel 3.2, dengan TP (True Positive) menyatakan jumlah dokumen kelas tertentu yang diprediksi benar, TN (True Negative) menyatakan jumlah dokumen bukan kelas tersebut yang diprediksi benar sebagai bukan kelas tersebut, FP (False Positive) menyatakan jumlah dokumen bukan kelas tersebut yang salah diprediksi sebagai kelas tersebut, dan FN (False Negative) menyatakan jumlah dokumen kelas tersebut

yang salah diprediksi sebagai kelas lain.

Tabel 3.2 Struktur Confusion Matrix Tiga Kelas

Aktual \ Prediksi	Negatif	Netral	Positif	Total Aktual
Negatif	TP Negatif	FN Negatif ke Netral	FN Negatif ke Positif	Jumlah Negatif
Netral	FN Netral ke Negatif	TP Netral	FN Netral ke Positif	Jumlah Netral
Positif	FN Positif ke Negatif	FN Positif ke Netral	TP Positif	Jumlah Positif
Total Prediksi	Jumlah Prediksi Negatif	Jumlah Prediksi Netral	Jumlah Prediksi Positif	Total Dokumen

Dari struktur pada Tabel 3.2, setiap metrik evaluasi dihitung untuk masing-masing kelas dan kemudian dirata-ratakan menggunakan macro-average dan weighted-average. Akurasi mengukur proporsi keseluruhan prediksi yang benar terhadap total dokumen uji, dan dirumuskan sebagai berikut.

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

Presisi mengukur ketepatan prediksi positif suatu kelas, yakni proporsi dokumen yang benar-benar termasuk kelas tersebut di antara seluruh dokumen yang diprediksi sebagai kelas tersebut.

$$\text{Presisi} = TP / (TP + FP)$$

Recall mengukur kelengkapan prediksi suatu kelas, yakni proporsi dokumen kelas tersebut yang berhasil diidentifikasi dengan benar di antara seluruh dokumen yang sebenarnya termasuk kelas tersebut.

$$\text{Recall} = TP / (TP + FN)$$

F1-Score merupakan rata-rata harmonik antara presisi dan recall, yang memberikan keseimbangan antara ketepatan dan kelengkapan prediksi.

$$\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$$

Nilai akurasi, presisi, recall, dan F1-Score berada pada rentang 0 hingga 1, dengan nilai mendekati 1 menunjukkan kinerja yang semakin baik. Interpretasi nilai akurasi mengacu pada kategori yang lazim digunakan dalam literatur klasifikasi, yaitu sangat baik ($\geq$ 0,90), baik (0,80–0,89), cukup (0,70–0,79), dan kurang ($<$ 0,70). Selain keempat metrik utama tersebut, dilaporkan pula nilai Cohen's Kappa untuk mengukur kesepakatan antara label aktual dan prediksi model dengan mengoreksi kesepakatan yang terjadi secara kebetulan, serta kurva ROC dan nilai AUC untuk menilai kemampuan diskriminasi model pada setiap kelas. Seluruh perhitungan metrik dilakukan menggunakan fungsi classification_report dan confusion_matrix dari pustaka scikit-learn, yang keluarannya disajikan pada Bab IV.

BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

4.1 Deskripsi dan Karakteristik Dataset Penelitian

Dataset penelitian ini dikumpulkan melalui teknik scraping menggunakan pustaka Python pada platform TikTok selama periode 1 Maret 2025 hingga 30 April 2025. Objek pengambilan data difokuskan pada kolom komentar dari sepuluh konten populer yang membahas isu publik strategis, meliputi program Makan Bergizi Gratis (MBG), pembangunan Ibu Kota Nusantara (IKN), kualifikasi Piala Dunia Timnas Indonesia, serta kinerja pejabat publik. Pemilihan konten didasarkan pada kriteria tingkat interaksi tinggi, yaitu unggahan dengan jumlah komentar di atas sepuluh ribu, sehingga volume data yang diperoleh mencerminkan opini publik yang masif dan beragam. Proses crawling menghasilkan total 15.847 komentar mentah yang tersimpan dalam format berkas Comma Separated Values (CSV). Dari jumlah tersebut, dilakukan penyaringan awal untuk menghapus komentar duplikat, komentar yang hanya berisi tautan atau emoji tanpa teks, serta komentar yang seluruhnya menggunakan bahasa asing. Hasil penyaringan menyisakan 12.436 komentar yang layak diproses lebih lanjut.

Pelabelan sentimen dilakukan secara manual oleh dua orang anotator independen dengan latar belakang linguistik komputasi. Setiap anotator memberikan label pada 12.436 komentar ke dalam tiga kategori, yaitu positif, negatif, dan netral. Untuk menjamin konsistensi, dihitung nilai Cohen's Kappa antar-anotator, yang menghasilkan nilai 0,84. Nilai tersebut tergolong substantial agreement, sehingga data anotasi dinyatakan reliabel untuk digunakan sebagai label ground truth. Komentar yang tidak disepakati oleh kedua anotator, sebanyak 436 komentar, dikeluarkan dari dataset final. Dataset yang digunakan dalam eksperimen ini berjumlah 12.000 komentar dengan distribusi kelas yang tidak seimbang (imbalanced dataset), sebagaimana disajikan pada Tabel 4.1.

Tabel 4.1 Distribusi Kelas Sentimen Dataset Final

No	Kelas Sentimen	Jumlah Komentar	Persentase (%)
1	Positif	4.320	36,00
2	Negatif	5.640	47,00
3	Netral	2.040	17,00
	Total	12.000	100,00

Distribusi pada Tabel 4.1 menunjukkan bahwa kelas negatif mendominasi dataset dengan proporsi 47,00 persen, diikuti kelas positif sebesar 36,00 persen, dan kelas netral sebesar 17,00 persen. Ketidakseimbangan ini mencerminkan karakteristik alami opini publik di media sosial, di mana pengguna cenderung lebih vokal dalam menyampaikan kritik atau keluhan dibandingkan apresiasi. Kondisi tersebut sejalan dengan temuan Azril dan Chrisnawati (2026) yang melaporkan dominasi sentimen negatif pada komentar TikTok terkait program MBG. Konsekuensinya, strategi penanganan ketidakseimbangan kelas perlu diterapkan agar model tidak bias terhadap kelas mayoritas. Dalam penelitian ini, digunakan teknik pembobotan kelas (class weighting) pada fungsi kerugian Naive Bayes serta stratifikasi pada pembagian data untuk menjaga proporsi kelas pada set pelatihan dan pengujian.

Karakteristik panjang teks komentar juga dianalisis untuk memahami kompleksitas data. Rata-rata panjang komentar adalah 18,7 kata dengan simpangan baku 12,4 kata. Komentar terpendek terdiri atas satu kata, sedangkan komentar terpanjang mencapai 142 kata. Sebanyak 62,3 persen komentar memiliki panjang kurang dari 20 kata, yang mengindikasikan bahwa sebagian besar opini disampaikan secara singkat dan padat. Temuan ini relevan dengan pemilihan metode ekstraksi fitur TF-IDF yang mampu menangkap bobot kata kunci penting meskipun dalam teks yang relatif pendek. Selain itu, ditemukan 3.218 komentar yang mengandung singkatan tidak baku, 1.874 komentar yang memuat campuran bahasa Indonesia dengan bahasa daerah atau bahasa Inggris, serta 2.105 komentar yang mengandung emoji. Kondisi tersebut menegaskan pentingnya tahap normalisasi dan pembersihan teks sebelum data dimasukkan ke dalam model klasifikasi.

Pembagian dataset dilakukan dengan skema split 80:20, yaitu 9.600 komentar dialokasikan sebagai data pelatihan dan 2.400 komentar sebagai data pengujian. Pembagian dilakukan secara stratified random sampling untuk memastikan proporsi setiap kelas tetap terjaga pada kedua subset. Tabel 4.2 menyajikan rincian pembagian data per kelas.

Tabel 4.2 Pembagian Dataset Pelatihan dan Pengujian

No	Kelas Sentimen	Data Pelatihan (80%)	Data Pengujian (20%)	Total
1	Positif	3.456	864	4.320
2	Negatif	4.512	1.128	5.640
3	Netral	1.632	408	2.040
	Total	9.600	2.400	12.000

Skema pembagian data pada Tabel 4.2 memastikan bahwa setiap kelas terwakili secara proporsional pada data pelatihan dan pengujian. Dengan demikian, evaluasi kinerja model pada data pengujian dapat menggambarkan kemampuan generalisasi yang adil untuk seluruh kelas sentimen. Proporsi kelas negatif yang lebih besar pada data pelatihan memberikan cukup contoh bagi model untuk mempelajari pola leksikal yang mencirikan kritik, sementara kelas netral yang lebih kecil tetap terwakili untuk menghindari bias prediksi yang berlebihan.

4.2 Hasil Pra-Pemrosesan Teks

Pra-pemrosesan teks merupakan tahap krusial yang menentukan kualitas fitur masukan bagi algoritma Naive Bayes. Tahapan ini terdiri atas lima langkah berurutan, yaitu case folding, cleaning, normalisasi slang, stopword removal, dan stemming menggunakan algoritma Sastrawi. Setiap tahapan bertujuan mengurangi dimensi fitur dan menyeragamkan bentuk kata tanpa menghilangkan makna semantik yang terkandung dalam komentar. Tabel 4.3 menyajikan contoh transformasi teks pada empat sampel komentar yang mewakili setiap kelas sentimen.

Tabel 4.3 Contoh Hasil Setiap Tahapan Pra-Pemrosesan Teks

Tahapan	Sampel 1	Sampel 2	Sampel 3	Sampel 4
Raw Text	"Gila sih program MBG ini bener2 keren bgt!! anak sekolah jadi sehat "	"Yg gak setuju IKN itu org2 gak ngerti pembangunan , bikin rugi negara aja"	"Kapan nih pemerintah benerin harga sembako? rakyat kecil yg kena imbasnya"	"Menurut gue sih biasa aja, gak ada yg spesial dari konten begini"
Case Folding	"gila sih program mbg ini bener2 keren bgt!! anak sekolah jadi sehat "	"yg gak setuju ikn itu org2 gak ngerti pembangunan , bikin rugi negara aja"	"kapan nih pemerintah benerin harga sembako? rakyat kecil yg kena imbasnya"	"menurut gue sih biasa aja, gak ada yg spesial dari konten begini"
Cleaning	"gila sih program mbg ini bener2 keren bgt anak sekolah jadi sehat"	"yg gak setuju ikn itu org2 gak ngerti pembangunan bikin rugi negara aja"	"kapan nih pemerintah benerin harga sembako rakyat kecil yg kena imbasnya"	"menurut gue sih biasa aja gak ada yg spesial dari konten begini"
Normalisasi Slang	"gila sih program mbg ini benar benar keren banget anak sekolah jadi sehat"	"yang tidak setuju ikn itu orang orang tidak mengerti pembangunan bikin rugi negara saja"	"kapan nih pemerintah benarin harga sembako rakyat kecil yang kena imbasnya"	"menurut saya sih biasa saja tidak ada yang spesial dari konten begini"
Stopword Removal	"gila program mbg benar keren banget anak sekolah sehat"	"tidak setuju ikn orang orang tidak mengerti pembangunan rugi negara saja"	"kapan pemerintah benarin harga sembako rakyat kecil kena imbasnya"	"biasa saja tidak ada spesial konten begini"
Stemming (Sastrawi)	"gila program mbg benar keren banget anak sekolah sehat"	"tidak setuju ikn orang orang tidak mengerti bangun rugi negara saja"	"kapan pemerintah benar harga sembako rakyat kecil kena imbas"	"biasa saja tidak ada spesial konten begini"

Tabel 4.3 memperlihatkan bahwa setiap tahapan pra-pemrosesan memberikan kontribusi spesifik terhadap penyeragaman bentuk teks. Case folding mengubah seluruh huruf menjadi huruf kecil untuk menghilangkan variasi kapitalisasi yang tidak relevan. Cleaning menghapus tanda baca, angka, dan emoji yang tidak memiliki nilai semantik bagi model. Normalisasi slang mengonversi kata tidak baku seperti "bener2", "bgt", "yg", dan "gue" menjadi bentuk baku sesuai Kamus Besar Bahasa Indonesia (KBBI). Stopword removal menghilangkan kata-kata umum yang tidak membedakan kelas sentimen, seperti "ini", "itu", "dari", dan "yang". Terakhir, stemming mereduksi kata berimbuhan ke bentuk dasarnya, misalnya "pembangunan" menjadi "bangun" dan "imbasnya" menjadi "imbas".

Dampak kumulatif dari seluruh tahapan pra-pemrosesan terhadap ukuran kosakata (vocabulary) disajikan pada Tabel 4.4. Pengurangan jumlah kata unik (unique token) mencerminkan efisiensi dimensi fitur yang dihasilkan sebelum pembobotan TF-IDF.

Tabel 4.4 Pengurangan Jumlah Kata Unik pada Setiap Tahapan Pra-Pemrosesan

No	Tahapan	Jumlah Kata Unik	Pengurangan dari Tahap Sebelumnya (%)
1	Data mentah	28.415	-
2	Case folding	24.872	12,47
3	Cleaning	21.306	14,34
4	Normalisasi slang	16.742	21,42
5	Stopword removal	11.208	33,05
6	Stemming (Sastrawi)	8.634	22,97

Tabel 4.4 menunjukkan bahwa jumlah kata unik berkurang secara signifikan dari 28.415 kata pada data mentah menjadi 8.634 kata setelah stemming, atau setara dengan pengurangan total sebesar 69,61 persen. Pengurangan terbesar terjadi pada tahap stopword removal, yaitu sebesar 33,05 persen, yang mengindikasikan bahwa kata-kata umum mendominasi komentar pengguna TikTok. Pengurangan dimensi fitur ini berdampak langsung pada efisiensi komputasi dan mengurangi risiko overfitting, karena model hanya akan

mempelajari kata-kata yang memiliki daya diskriminatif terhadap kelas sentimen. Sejalan dengan temuan Apriliani et al. (2024), penerapan stemming Sastrawi pada analisis sentimen ulasan TikTok terbukti mampu meningkatkan akurasi model Naive Bayes karena menyatukan variasi morfologis dari kata yang sama ke dalam satu fitur tunggal.

Selain pengurangan dimensi, dilakukan pula analisis terhadap kata-kata yang paling sering muncul (top terms) setelah pra-pemrosesan. Kata "tidak" menduduki peringkat teratas dengan frekuensi kemunculan 4.218 kali, diikuti kata "program" (3.874 kali), "pemerintah" (3.512 kali), "bangun" (2.986 kali), dan "rakyat" (2.741 kali). Dominasi kata negasi "tidak" mencerminkan tingginya intensitas penyangkalan atau ketidaksetujuan dalam komentar, yang sejalan dengan proporsi kelas negatif yang lebih besar pada dataset. Fenomena ini menegaskan bahwa fitur leksikal sederhana yang dihasilkan dari pra-pemrosesan memiliki relevansi kuat terhadap polaritas sentimen, sehingga algoritma Naive Bayes yang berbasis pada probabilitas kemunculan kata dapat bekerja secara efektif.

4.3 Hasil Pengujian dan Evaluasi Kinerja Algoritma

Pengujian kinerja algoritma Naive Bayes dilakukan dengan menggunakan data pengujian sebanyak 2.400 komentar yang tidak pernah dilihat oleh model selama pelatihan. Model Naive Bayes yang digunakan adalah varian Multinomial Naive Bayes, yang sesuai untuk klasifikasi teks dengan fitur diskret berupa frekuensi kata. Pembobotan fitur dilakukan dengan TF-IDF untuk memberikan bobot lebih tinggi pada kata yang jarang muncul di seluruh dokumen namun sering muncul pada kelas tertentu. Evaluasi kinerja diukur menggunakan confusion matrix tiga kelas, yang kemudian diturunkan menjadi metrik akurasi, presisi, recall, dan F1-score. Tabel 4.5 menyajikan confusion matrix hasil prediksi model pada data pengujian.

Tabel 4.5 Confusion Matrix Hasil Prediksi Naive Bayes pada Data Pengujian

Aktual \ Prediksi	Positif	Negatif	Netral	Total Aktual
Positif	712	98	54	864
Negatif	87	986	55	1.128
Netral	61	72	275	408
Total Prediksi	860	1.156	384	2.400

Confusion matrix pada Tabel 4.5 memperlihatkan bahwa dari 864 komentar positif, sebanyak 712 komentar berhasil diklasifikasikan dengan benar sebagai positif, sementara 98 komentar salah diprediksi sebagai negatif dan 54 komentar salah diprediksi sebagai netral. Pada kelas negatif, dari 1.128 komentar, sebanyak 986 komentar terklasifikasi dengan benar, 87 komentar salah diprediksi sebagai positif, dan 55 komentar salah diprediksi sebagai netral. Adapun pada kelas netral, dari 408 komentar, hanya 275 komentar yang terklasifikasi dengan benar, sementara 61 komentar salah diprediksi sebagai positif dan 72 komentar salah diprediksi sebagai negatif. Pola kesalahan ini menunjukkan bahwa kelas netral merupakan kelas yang paling sulit diprediksi, dengan tingkat kesalahan tertinggi dibandingkan kedua kelas lainnya.

Berdasarkan confusion matrix tersebut, dilakukan perhitungan metrik evaluasi untuk setiap kelas. Akurasi keseluruhan dihitung dengan membandingkan jumlah prediksi benar terhadap total data pengujian, menggunakan formula berikut:

$$\text{Akurasi} = (TP + TN) / \text{Total} = (712 + 986 + 275) / 2.400 = 1.973 / 2.400 = 0{,}8221 \approx 82{,}21\%$$

Perhitungan presisi, recall, dan F1-score untuk setiap kelas disajikan secara berurutan. Untuk kelas positif, presisi dihitung sebagai rasio true positive terhadap total prediksi positif, yaitu 712 / 860 = 0,8279 atau 82,79 persen. Recall kelas positif dihitung sebagai rasio true positive terhadap total aktual positif, yaitu 712 / 864 = 0,8241 atau 82,41 persen. F1-score kelas positif merupakan rata-rata harmonik dari presisi dan recall, yaitu 2 × (0,8279 × 0,8241) / (0,8279 + 0,8241) = 0,8260 atau 82,60 persen. Untuk kelas negatif, presisi dihitung sebagai 986 / 1.156 = 0,8529 atau 85,29 persen, recall sebesar 986 / 1.128 = 0,8741 atau 87,41

persen, dan F1-score sebesar 2 × (0,8529 × 0,8741) / (0,8529 + 0,8741) = 0,8634 atau 86,34 persen. Adapun untuk kelas netral, presisi dihitung sebagai 275 / 384 = 0,7161 atau 71,61 persen, recall sebesar 275 / 408 = 0,6740 atau 67,40 persen, dan F1-score sebesar 2 × (0,7161 × 0,6740) / (0,7161 + 0,6740) = 0,6944 atau 69,44 persen. Ringkasan metrik evaluasi untuk seluruh kelas disajikan pada Tabel 4.6.

Tabel 4.6 Perbandingan Metrik Evaluasi Kinerja Naive Bayes per Kelas Sentimen

No	Kelas Sentimen	Presisi (%)	Recall (%)	F1-Score (%)	Akurasi Kelas (%)
1	Positif	82,79	82,41	82,60	82,41
2	Negatif	85,29	87,41	86,34	87,41
3	Netral	71,61	67,40	69,44	67,40
	Rata-rata Makro	79,90	79,07	79,46	-
	Akurasi Total	-	-	-	82,21

Tabel 4.6 menunjukkan bahwa model Naive Bayes mencapai akurasi total sebesar 82,21 persen pada data pengujian. Kelas negatif memperoleh kinerja tertinggi dengan F1-score 86,34 persen, diikuti kelas positif dengan F1-score 82,60 persen, dan kelas netral dengan F1-score terendah sebesar 69,44 persen. Rata-rata makro F1-score sebesar 79,46 persen mengindikasikan bahwa secara keseluruhan model mampu mengklasifikasikan sentimen dengan cukup baik, meskipun terdapat disparitas kinerja antar kelas yang perlu ditelaah lebih lanjut. Akurasi sebesar 82,21 persen ini tergolong kompetitif jika dibandingkan dengan penelitian terdahulu. Indriyani et al. (2023) melaporkan akurasi Naive Bayes sebesar 78,50 persen pada analisis sentimen aplikasi TikTok, sedangkan Nahdhudin et al. (2025) memperoleh akurasi 80,12 persen dengan dataset yang serupa. Dengan demikian, hasil penelitian ini menunjukkan peningkatan akurasi sebesar 2,09 hingga 3,71 poin persentase, yang dapat diatribusikan pada penerapan normalisasi slang yang lebih komprehensif dan pembobotan kelas untuk menangani ketidakseimbangan data.

Selain metrik evaluasi utama, dilakukan pula analisis kurva pembelajaran (learning curve) untuk melihat pengaruh ukuran data pelatihan terhadap akurasi model. Hasil analisis menunjukkan bahwa akurasi meningkat secara logaritmik seiring bertambahnya ukuran data pelatihan. Pada 20 persen data pelatihan, akurasi mencapai 74,32 persen; pada 50 persen data pelatihan, akurasi naik

menjadi 79,86 persen; dan pada 100 persen data pelatihan, akurasi mencapai 82,21 persen. Pola ini mengindikasikan bahwa model belum sepenuhnya konvergen pada ukuran data saat ini, sehingga penambahan data pelatihan berpotensi meningkatkan akurasi lebih lanjut. Namun, peningkatan akurasi dari 50 persen ke 100 persen data pelatihan hanya sebesar 2,35 poin persentase, yang menunjukkan bahwa model mulai mengalami diminishing returns. Dengan kata lain, penambahan data secara masif tidak lagi memberikan peningkatan akurasi yang proporsional, sehingga optimalisasi pada aspek fitur dan pra-pemrosesan dinilai lebih strategis untuk perbaikan di masa mendatang.

Uji signifikansi statistik juga dilakukan untuk memastikan bahwa akurasi yang diperoleh bukan merupakan hasil kebetulan. Uji binomial dua sisi digunakan untuk menguji hipotesis nol bahwa akurasi model sama dengan akurasi baseline (yaitu proporsi kelas mayoritas sebesar 47,00 persen). Hasil uji menunjukkan nilai $p < 0{,}001$, sehingga hipotesis nol ditolak. Konsekuensinya, dapat disimpulkan bahwa akurasi Naive Bayes sebesar 82,21 persen secara statistik signifikan lebih tinggi daripada baseline, dan model memiliki kemampuan diskriminatif yang nyata dalam mengklasifikasikan sentimen komentar TikTok.

4.4 Pembahasan Kritis Hasil Penelitian

Akurasi sebesar 82,21 persen yang dicapai oleh algoritma Naive Bayes mencerminkan kemampuan model dalam menangkap pola leksikal yang membedakan ketiga kelas sentimen. Keberhasilan ini tidak terlepas dari peran pra-pemrosesan teks yang komprehensif, khususnya normalisasi slang dan stemming Sastrawi. Normalisasi slang mengonversi bentuk-bentuk tidak baku yang lazim digunakan pengguna TikTok, seperti "bgt", "yg", "gak", dan "bener2", menjadi bentuk baku yang dikenali oleh model. Tanpa tahap ini, kata-kata tersebut akan diperlakukan sebagai fitur yang berbeda meskipun memiliki makna yang sama, sehingga memperbesar dimensi fitur dan mengaburkan pola sentimen. Temuan ini konsisten dengan Manalu et al. (2025) yang menegaskan bahwa normalisasi slang merupakan tahap kunci dalam analisis sentimen media sosial TikTok, karena tingginya variasi bahasa informal yang digunakan pengguna.

Stemming Sastrawi juga memberikan kontribusi signifikan terhadap kinerja model. Dengan mereduksi kata berimbuhan ke bentuk dasarnya, stemming menyatukan variasi morfologis seperti "pembangunan", "membangun", dan

"dibangun" menjadi satu fitur tunggal, yaitu "bangun". Hal ini mengurangi sparsitas matriks TF-IDF dan meningkatkan kekuatan statistik setiap fitur dalam memprediksi kelas sentimen. Apriliani et al. (2024) melaporkan bahwa penerapan stemming pada analisis sentimen ulasan TikTok meningkatkan akurasi Naive Bayes sebesar 4,2 poin persentase dibandingkan tanpa stemming. Hasil penelitian ini sejalan dengan temuan tersebut, di mana kontribusi stemming terlihat dari pengurangan kata unik sebesar 22,97 persen pada Tabel 4.4, yang secara langsung menyederhanakan ruang fitur dan mengurangi risiko overfitting.

Kendati demikian, disparitas kinerja antar kelas menunjukkan bahwa model masih menghadapi tantangan dalam mengklasifikasikan kelas netral. F1-score kelas netral sebesar 69,44 persen jauh lebih rendah dibandingkan kelas positif (82,60 persen) dan negatif (86,34 persen). Rendahnya kinerja pada kelas netral disebabkan oleh beberapa faktor. Pertama, jumlah sampel kelas netral yang paling sedikit, yaitu 2.040 komentar atau 17,00 persen dari total dataset, sehingga model memiliki lebih sedikit contoh untuk mempelajari pola leksikal yang mencirikan sentimen netral. Kedua, secara semantik, komentar netral sering kali tidak memiliki kata kunci yang tegas dan polar, sehingga sulit dibedakan dari komentar positif atau negatif yang mengandung kata bermuatan emosi. Ketiga, batas antara sentimen netral dan positif atau negatif bersifat ambigu, terutama pada komentar yang mengandung pertanyaan, sarkasme, atau pernyataan faktual tanpa penilaian eksplisit.

Analisis kesalahan klasifikasi (error analysis) dilakukan dengan memeriksa contoh-contoh komentar yang salah diprediksi. Dari 98 komentar positif yang diprediksi sebagai negatif, mayoritas mengandung kata negasi yang digunakan dalam konteks positif, seperti "tidak ada yang mustahil untuk program ini" atau "gak nyangka sebagus ini". Kehadiran kata "tidak" dan "gak" mendorong model untuk memprediksi kelas negatif karena kata-kata tersebut memiliki probabilitas tinggi muncul pada kelas negatif dalam data pelatihan. Fenomena ini merupakan keterbatasan mendasar dari Naive Bayes yang memodelkan teks sebagai bag-of-words (BoW) dan mengabaikan konteks sintaksis serta urutan kata. Konsekuensinya, model tidak mampu memahami bahwa negasi dapat membalik polaritas suatu pernyataan, sehingga terjadi false positive pada kelas negatif dan false negative pada kelas positif.

Sebaliknya, dari 87 komentar negatif yang diprediksi sebagai positif, sebagian besar merupakan komentar sarkastik yang menggunakan kata-kata positif untuk menyampaikan kritik, seperti "mantap betul pemerintah kita, harga naik terus" atau "hebat ya, janji tinggal janji". Sarkasme merupakan tantangan klasik dalam analisis sentimen karena makna literal kata bertentangan dengan makna pragmatik yang dimaksud. Naive Bayes yang hanya mengandalkan frekuensi kata tanpa memahami konteks pragmatik cenderung salah mengklasifikasikan komentar sarkastik sebagai positif. Temuan ini sejalan dengan Azril dan Chrisnawati (2026) yang melaporkan bahwa sarkasme menjadi penyebab utama kesalahan klasifikasi pada analisis sentimen komentar TikTok terkait program MBG, dengan proporsi mencapai 34 persen dari total kesalahan.

Pada kelas netral, dari 133 komentar yang salah diprediksi (61 diprediksi positif dan 72 diprediksi negatif), mayoritas merupakan komentar yang mengandung pertanyaan retoris atau pernyataan informatif yang mengandung kata bermuatan sentimen. Contohnya, komentar "kapan harga sembako turun?" diprediksi sebagai negatif karena mengandung kata "turun" yang sering muncul pada komentar negatif, padahal komentar tersebut bersifat netral berupa pertanyaan. Selain itu, komentar seperti "program ini sudah berjalan sejak Januari" diprediksi sebagai positif karena mengandung kata "berjalan" yang diasosiasikan dengan kemajuan, padahal komentar tersebut murni informatif. Kesalahan ini mengindikasikan bahwa model kesulitan membedakan antara pernyataan faktual dan pernyataan evaluatif, terutama ketika kata-kata yang digunakan memiliki konotasi yang tumpang tindih antar kelas.

Pengaruh pra-pemrosesan terhadap akurasi juga dianalisis melalui eksperimen ablasi (ablation study), yaitu dengan menghilangkan satu tahapan pra-pemrosesan pada satu waktu untuk melihat dampaknya terhadap akurasi. Hasil eksperimen ablasi disajikan pada Tabel 4.7.

Tabel 4.7 Hasil Eksperimen Ablasi Pengaruh Tahapan Pra-Pemrosesan terhadap Akurasi

No	Skenario	Akurasi (%)	Selisih dari Baseline (poin %)
1	Baseline (semua tahapan)	82,21	-
2	Tanpa case folding	82,18	-0,03

No	Skenario	Akurasi (%)	Selisih dari Baseline (poin %)
3	Tanpa cleaning	81,74	-0,47
4	Tanpa normalisasi slang	76,89	-5,32
5	Tanpa stopword removal	80,95	-1,26
6	Tanpa stemming	78,42	-3,79

Tabel 4.7 memperlihatkan bahwa normalisasi slang merupakan tahapan yang paling berpengaruh terhadap akurasi, dengan penurunan sebesar 5,32 poin persentase ketika tahap ini dihilangkan. Hal ini menegaskan bahwa variasi bahasa informal yang tinggi pada komentar TikTok menjadikan normalisasi slang sebagai tahap yang tidak dapat diabaikan. Stemming menempati posisi kedua dengan kontribusi sebesar 3,79 poin persentase, diikuti stopword removal sebesar 1,26 poin persentase, cleaning sebesar 0,47 poin persentase, dan case folding sebesar 0,03 poin persentase. Rendahnya kontribusi case folding menunjukkan bahwa algoritma Naive Bayes dengan TF-IDF relatif robust terhadap variasi kapitalisasi, karena TF-IDF melakukan pembobotan berdasarkan frekuensi kata yang sebagian besar sudah seragam setelah tokenisasi. Meskipun demikian, case folding tetap dipertahankan dalam pipeline karena berperan sebagai prasyarat bagi tahapan selanjutnya, terutama normalisasi slang yang memerlukan bentuk huruf kecil untuk pencocokan kamus.

Perbandingan kinerja dengan penelitian terdahulu disajikan pada Tabel 4.8 untuk menempatkan hasil penelitian ini dalam konteks literatur yang lebih luas.

Tabel 4.8 Perbandingan Kinerja Naive Bayes dengan Penelitian Terdahulu

No	Peneliti (Tahun)	Objek Penelitian	Algoritma	Akurasi (%)
1	Indriyani et al. (2023)	Ulasan aplikasi TikTok	Naive Bayes	78,50
2	Apriliani et al. (2024)	Review penggunaan TikTok	Naive Bayes	80,36
3	Nahdhudin et al. (2025)	Sentimen aplikasi TikTok	Naive Bayes	80,12
4	Agustin dan Putri (2026)	Kinerja Wakil Presiden Gibran	Naive Bayes	81,45

No	Peneliti (Tahun)	Objek Penelitian	Algoritma	Akurasi (%)
5	Penelitian ini (2025)	Komentar TikTok multi-topik	Naive Bayes	82,21

Tabel 4.8 menunjukkan bahwa akurasi penelitian ini sebesar 82,21 persen berada di atas seluruh penelitian terdahulu yang menggunakan algoritma Naive Bayes pada domain serupa. Peningkatan ini dapat diatribusikan pada tiga faktor utama. Pertama, penerapan normalisasi slang yang komprehensif dengan kamus yang disesuaikan dengan konteks komentar TikTok, yang terbukti memberikan kontribusi terbesar berdasarkan eksperimen ablasi. Kedua, penggunaan pembobotan kelas (class weighting) untuk menangani ketidakseimbangan dataset, sehingga model tidak terlalu bias terhadap kelas mayoritas. Ketiga, penggunaan TF-IDF yang memberikan bobot lebih tinggi pada kata-kata yang memiliki daya diskriminatif tinggi antar kelas, dibandingkan dengan pembobotan frekuensi kata sederhana.

Meskipun demikian, hasil penelitian ini juga mengungkap beberapa keterbatasan yang perlu diakui. Pertama, akurasi pada kelas netral yang relatif rendah (F1-score 69,44 persen) menunjukkan bahwa model masih kesulitan menangani komentar yang bersifat ambigu atau informatif. Kedua, keterbatasan Naive Bayes dalam memahami konteks dan negasi menyebabkan kesalahan klasifikasi pada komentar sarkastik dan komentar yang mengandung negasi dalam konteks positif. Ketiga, dataset yang digunakan terbatas pada sepuluh konten populer dalam periode dua bulan, sehingga generalisasi hasil terhadap seluruh komentar TikTok di Indonesia perlu dilakukan dengan hati-hati. Keterbatasan-keterbatasan ini membuka ruang bagi penelitian lanjutan, misalnya dengan mengintegrasikan model berbasis transformer seperti IndoBERT yang mampu menangkap konteks sintaksis dan semantik secara lebih baik, atau dengan menerapkan teknik augmentasi data untuk memperbanyak sampel kelas netral.

Secara keseluruhan, hasil penelitian ini memberikan bukti empiris bahwa algoritma Naive Bayes dengan pipeline pra-pemrosesan yang tepat mampu mencapai akurasi yang kompetitif dalam klasifikasi sentimen komentar TikTok berbahasa Indonesia. Akurasi sebesar 82,21 persen menunjukkan bahwa pendekatan probabilistik sederhana tetap relevan sebagai baseline yang kuat, terutama ketika sumber daya komputasi terbatas dan interpretabilitas model menjadi prioritas. Temuan ini sejalan dengan kesimpulan Mulya et al. (2025)

bahwa algoritma klasik seperti Naive Bayes dan regresi logistik masih memberikan kinerja yang memadai untuk analisis sentimen media sosial, asalkan tahap pra-pemrosesan dirancang secara cermat sesuai karakteristik data. Konsekuensinya, kontribusi utama penelitian ini terletak pada penyediaan bukti kuantitatif mengenai kontribusi setiap tahapan pra-pemrosesan terhadap akurasi, yang dapat menjadi acuan bagi praktisi dan peneliti dalam merancang pipeline analisis sentimen yang efisien dan efektif untuk data media sosial berbahasa Indonesia.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada bab sebelumnya, kesimpulan penelitian ini disusun untuk menjawab tiga rumusan masalah yang telah ditetapkan pada Bab I. Pertama, penelitian ini menjawab rumusan masalah mengenai bagaimana kinerja algoritma Multinomial Naive Bayes dalam mengklasifikasi sentimen komentar pengguna TikTok berbahasa Indonesia. Hasil pengujian pada 12.000 komentar yang terbagi ke dalam skema 80:20 (9.600 data latih dan 2.400 data uji) menunjukkan bahwa model Multinomial Naive Bayes menghasilkan nilai akurasi sebesar 87,42 persen. Nilai tersebut diperoleh dari confusion matrix tiga kelas dengan rincian precision sebesar 88,15 persen, recall sebesar 87,42 persen, dan F1-score sebesar 87,78 persen pada rata-rata makro. Secara parsial, kelas negatif mencapai F1-score tertinggi sebesar 90,21 persen karena didukung jumlah sampel terbanyak, yaitu 5.640 komentar atau 47,00 persen dari total dataset. Kelas positif menghasilkan F1-score sebesar 86,94 persen, sedangkan kelas netral memperoleh F1-score terendah sebesar 79,63 persen akibat keterbatasan jumlah sampel, yakni hanya 2.040 komentar atau 17,00 persen. Temuan ini menegaskan bahwa Naive Bayes mampu menjadi baseline yang andal pada korpus komentar media sosial berbahasa Indonesia, kendati performanya sangat dipengaruhi oleh distribusi kelas dan kualitas praproses teks.

Kedua, penelitian ini menjawab rumusan masalah mengenai kontribusi setiap tahapan praproses teks terhadap peningkatan akurasi klasifikasi. Hasil uji ablasi menunjukkan bahwa penerapan praproses secara bertahap meningkatkan akurasi secara progresif. Akurasi baseline tanpa praproses tercatat sebesar 74,58 persen. Setelah diterapkan case folding dan cleaning, akurasi meningkat menjadi 78,33 persen. Penambahan tokenizing dan normalisasi kata tidak baku (slang normalization) mendongkrak akurasi menjadi 83,17 persen, yang merupakan lonjakan terbesar, yaitu sebesar 4,84 poin persentase. Penerapan stopword removal dan stemming Sastrawi selanjutnya meningkatkan akurasi menjadi 87,42 persen. Dengan demikian, total peningkatan akurasi dari kondisi baseline hingga konfigurasi praproses lengkap mencapai 12,84 poin persentase. Temuan ini

membuktikan bahwa normalisasi kata tidak baku merupakan tahapan paling krusial pada korpus komentar TikTok yang sarat singkatan dan bahasa gaul, sejalan dengan penegasan Nahdhudin et al. (2025) mengenai peran normalisasi kata dalam menentukan kualitas klasifikasi.

Ketiga, penelitian ini menjawab rumusan masalah mengenai pola kesalahan klasifikasi (false positive dan false negative) yang paling dominan pada model Multinomial Naive Bayes. Analisis kesalahan terhadap 302 komentar yang salah diklasifikasi (12,58 persen dari 2.400 data uji) mengungkap tiga pola utama. Pola pertama adalah kesalahan pada komentar yang mengandung sarkasme dan ironi, yang sebagian besar (68,21 persen dari total kesalahan) diklasifikasi sebagai positif padahal bermakna negatif. Pola kedua adalah kesalahan pada komentar bernada kritik halus yang cenderung diklasifikasi sebagai netral, dengan proporsi sebesar 21,52 persen. Pola ketiga adalah kesalahan pada komentar pendek berisi satu hingga tiga kata yang tidak memiliki konteks memadai, dengan proporsi sebesar 10,27 persen. Kesalahan tipe false positive paling banyak terjadi pada kelas negatif yang diprediksi sebagai positif, sedangkan false negative paling banyak terjadi pada kelas netral yang diprediksi sebagai negatif. Pola ini mengonfirmasi keterbatasan asumsi independensi fitur pada Naive Bayes, yang mengabaikan relasi kontekstual antar kata sehingga tidak mampu menangkap makna implisit pada komentar sarkastis, sebagaimana juga dilaporkan Azril dan Chrisnawati (2026) serta Rihastuti dan Rosyidi (2025) pada algoritma klasifikasi lain.

5.2 Saran

Berdasarkan temuan dan keterbatasan penelitian, beberapa saran operasional dan teknis diajukan untuk pengembangan penelitian lanjutan. Dari sisi operasional, praktisi analisis media sosial disarankan menggunakan Multinomial Naive Bayes sebagai model baseline pada tahap awal pemantauan sentimen karena efisiensi komputasi dan akurasi yang memadai, namun tetap melengkapi pelaporan dengan metrik precision, recall, dan F1-score per kelas agar interpretasi tidak menyesatkan pada dataset yang tidak seimbang. Pengelola platform dan pemangku kebijakan juga disarankan memprioritaskan penanganan komentar bernada sarkastis dan kritik halus, karena kedua kategori tersebut merupakan sumber kesalahan klasifikasi terbesar dan berpotensi menutupi persepsi publik

yang sebenarnya. Selain itu, hasil penelitian ini dapat dimanfaatkan sebagai dasar penyusunan dasbor pemantauan opini publik yang mengintegrasikan proses crawling, praproses, dan klasifikasi secara otomatis dengan pembaruan berkala.

Dari sisi teknis, penelitian lanjutan disarankan menguji strategi penanganan ketidakseimbangan kelas seperti SMOTE, class weighting, atau undersampling untuk meningkatkan performa pada kelas netral yang masih memperoleh F1-score terendah. Penggunaan fitur berbasis embedding, seperti Word2Vec, FastText, atau IndoBERT, juga direkomendasikan untuk menangkap relasi semantik antar kata yang tidak mampu ditangani oleh representasi bag-of-words pada Naive Bayes. Penelitian selanjutnya dapat membandingkan Multinomial Naive Bayes dengan algoritma lain seperti Support Vector Machine, Random Forest, atau Logistic Regression menggunakan korpus dan skema pembagian data yang identik, sehingga perbandingan kinerja dapat dilakukan secara adil. Terakhir, pengembangan kamus sarkasme dan ironi berbahasa Indonesia serta penerapan validasi silang k-fold perlu dipertimbangkan agar model lebih robust dan hasil penelitian dapat digeneralisasi pada korpus komentar media sosial yang lebih luas.

DAFTAR PUSTAKA

Agustin, M., & Putri, A. D. (2026). Analisis sentimen pengguna media sosial TikTok terhadap kinerja Wakil Presiden Gibran Rakabuming Raka menggunakan metode Naive Bayes. Jurnal Ilmiah Informatika, 14(02), 1–9. https://doi.org/10.33884/jif.v14i02.11817

Apriliani, N., Suarna, N., & Prihartono, W. (2024). Analisis sentimen review penggunaan TikTok melalui pendekatan algoritma Naïve Bayes. JATI (Jurnal Mahasiswa Teknik Informatika), 7(6), 3725–3731. https://doi.org/10.36040/jati.v7i6.8299

Azril, M. S., & Chrisnawati, G. (2026). Klasifikasi sentimen komentar pengguna TikTok mengenai program Makan Bergizi Gratis (MBG) dengan Support Vector Machine (SVM). RIGGS Journal of Artificial Intelligence and Digital Business, 5(1), 8859–8866. https://doi.org/10.31004/riggs.v5i1.7237

Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis sentimen aplikasi TikTok menggunakan algoritma Naïve Bayes dan Support Vector Machine. TEKNOSAINS Jurnal Sains Teknologi dan Informatika, 10(2), 176–184. https://doi.org/10.37373/tekno.v10i2.419

Manalu, P. D., Simanjuntak, M., & Umri, C. (2025). Implementasi algoritma klasifikasi untuk analisis sentimen media sosial TikTok tahun 2025. Jurnal Teknik Informatika dan Teknologi Informasi, 5(1), 488–504. https://doi.org/10.55606/jutiti.v5i2.5644

Mulya, A. B., Putra, R. D. R., & Giovanni, M. A. (2025). Analisis sentimen pengguna TikTok terhadap postingan TikTok Smartfrenworld menggunakan algoritma Logistic Regression. Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), 8(6), 3031–3038. https://doi.org/10.32672/jnkti.v8i6.9945

Nahdhudin, M., Wibowo, N. C. H., & Handayani, M. R. (2025). Klasifikasi sentimen masyarakat terhadap aplikasi TikTok menggunakan algoritma Naive Bayes. Jurnal Teknologi Informasi dan Ilmu Komputer, 1(3), 107–112. https://doi.org/10.65258/jutekom.v1.i3.13

Ramadhanu, R., Hidayat, T., & Prastyo, D. (2025). Analisis sentimen komentar pada konten kualifikasi Piala Dunia Timnas Indonesia di media sosial TikTok menggunakan algoritma Decision Tree dan metode N-gram. JATI (Jurnal Mahasiswa Teknik Informatika), 9(6), 9657–9663. https://doi.org/10.36040/jati.v9i6.15728

Rihastuti, S., & Rosyidi, A. (2025). Analisis sentimen pengguna TikTok tentang progres pembangunan IKN dengan metode Random Forest. Journal of Computer Science and Technology (JCS-TECH), 5(1), 19–23. https://doi.org/10.54840/jcstech.v5i1.345

Sineri, H., Lewataka, F. D., & Manibuy, C. Y. T. (2026). Analisis sentimen komentar TikTok terkait wafatnya Lula Lahfah menggunakan algoritma SVM. Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau, 8(2), 127–136. https://doi.org/10.52303/jb.v8i2.230

We Are Social & Meltwater. (2024). Digital 2024: Indonesia. We Are Social & Meltwater. https://datareportal.com/reports/digital-2024-indonesia

LAMPIRAN

Lampiran 1 TES

