



Penerapan Naïve Bayes untuk Mengklasifikasikan Sentimen Tidak Seimbang pada Ulasan Aplikasi Berbasis Etika Konsumen

Lingga, Firman Noor Hasan*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta
Jl. Tanah Merdeka No.20, RT.11/RW.2, Rambutan, Kec. Ciracas, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹djati315@gmail.com, ^{2,*}firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Submitted: 03/07/2025; Accepted: 31/07/2025; Published: 31/07/2025

Abstrak—Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna terhadap aplikasi berbasis etika konsumsi menggunakan algoritma Multinomial Naïve Bayes. Aplikasi yang dikaji memuat konten sosial-moral dan sering kali memunculkan ekspresi opini yang kompleks. Sebanyak 2.000 ulasan pengguna dikumpulkan dari Google Play Store dengan teknik web scraping dan diproses melalui tahapan text preprocessing meliputi case folding, cleansing, tokenizing, stopword removal, dan stemming. Data dikonversi ke bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), lalu dilabeli menjadi tiga kategori sentimen: positif, netral, dan negatif. Hasil evaluasi menunjukkan bahwa model mampu mengenali sentimen positif dengan precision sebesar 92%, recall 100%, dan f1-score 96%. Namun, performa terhadap sentimen netral dan negatif masih rendah akibat ketimpangan distribusi data. Penelitian ini memberikan kontribusi dalam memahami keterbatasan model klasifikasi berbasis probabilistik dalam menangani opini tidak seimbang pada ruang digital bernilai sosial.

Kata Kunci: Analisis Sentimen; Naive Bayes; Google Play Store; TF-IDF; Aplikasi No Thanks!

Abstract—This study aims to classify user sentiment toward an ethics-based consumption application using the Multinomial Naïve Bayes algorithm. The application examined contains social and moral content, often provoking complex opinion expressions. A total of 2,000 user reviews were collected from Google Play Store using web scraping and processed through a series of text preprocessing steps: case folding, cleansing, tokenizing, stopword removal, and stemming. The data were converted into numerical form using the Term Frequency–Inverse Document Frequency (TF-IDF) method and labeled into three sentiment categories: positive, neutral, and negative. The evaluation results show that the model achieved a precision of 92%, recall of 100%, and an f1-score of 96% for positive sentiment. However, the model underperformed in recognizing neutral and negative sentiments due to class imbalance. This study contributes to understanding the limitations of probabilistic classification models in handling imbalanced public opinion in socially driven digital spaces.

Keywords: Sentiment Analysis; Naive Bayes; Google Play Store; TF-IDF; No Thanks! Application

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah lanskap komunikasi sosial secara fundamental [1]. Internet dan media digital tidak hanya mempercepat distribusi informasi, tetapi juga menciptakan ruang partisipatif yang memungkinkan masyarakat untuk menyuarakan pandangan, kritik, dan preferensi terhadap berbagai isu sosial, politik, maupun ekonomi. Dalam konteks ini, platform digital seperti Google Play Store berperan penting sebagai medium interaksi antara pengguna dan pengembang aplikasi melalui fitur ulasan (reviews) yang bersifat terbuka. Setiap ulasan tidak hanya berfungsi sebagai sarana evaluasi terhadap kualitas aplikasi, tetapi juga mencerminkan opini publik dan nilai-nilai sosial yang berkembang di masyarakat.

Salah satu fenomena menarik dalam konteks tersebut adalah munculnya aplikasi No Thanks!, sebuah platform yang berfokus pada edukasi etika konsumsi. Aplikasi ini menyajikan informasi terkait perusahaan atau produk yang diduga terlibat dalam pelanggaran hak asasi manusia, konflik geopolitik, maupun praktik ekonomi yang tidak berkelanjutan. Dengan mendorong pengguna untuk melakukan boikot terhadap entitas yang dianggap bermasalah, aplikasi ini menjadi contoh nyata bagaimana teknologi dapat digunakan untuk mendukung gerakan sosial berbasis nilai.

Tingginya jumlah unduhan dan banyaknya ulasan dari pengguna menunjukkan adanya ketertarikan dan keterlibatan emosional terhadap isu yang diangkat aplikasi tersebut. Oleh karena itu, analisis terhadap ulasan pengguna menjadi penting untuk memahami persepsi dan kecenderungan sikap publik. Pendekatan analisis sentimen berbasis Natural Language Processing (NLP) memungkinkan peneliti untuk mengekstraksi makna, mengidentifikasi emosi, serta mengklasifikasikan opini pengguna secara sistematis. Dengan cara ini, penelitian dapat mengungkap bagaimana opini digital merefleksikan dinamika sosial dalam ruang komunikasi modern.

Beberapa studi telah memanfaatkan pendekatan ini dalam berbagai konteks sosial, seperti komentar warganet terhadap konser K-pop [2], pengalaman belanja thriftling [3], respons terhadap RUU perampasan aset [4], hingga preferensi terhadap layanan finansial paylater [5].

Dalam implementasinya, salah satu algoritma yang banyak digunakan adalah Naïve Bayes Classifier, yang bekerja berdasarkan probabilitas bersyarat dan asumsi independensi antar fitur. Algoritma ini terbukti efisien dalam klasifikasi teks berdimensi tinggi serta memiliki kecepatan pemrosesan yang tinggi. Studi oleh Syakir dan Hasan [6] menunjukkan bahwa algoritma ini mampu mengklasifikasikan opini publik terhadap isu korupsi dengan

hasil yang memuaskan. Wibowo et al. [7] juga menunjukkan efektivitas Naïve Bayes dalam memahami sentimen publik terhadap efektivitas pembelajaran daring selama pandemi COVID-19.

Selain pada domain sosial-politik, algoritma ini juga berhasil diterapkan dalam konteks ekonomi digital. Metode ini juga telah digunakan dalam aplikasi layanan publik oleh Tanggraeni dan Sitokdana (2022) yang menganalisis sentimen pada sistem e-Government untuk meningkatkan transparansi digital [8]. Putra [9] mengkaji ulasan pada aplikasi investasi seperti Bibit dan Bareksa, dan menemukan bahwa Naïve Bayes dan K-Nearest Neighbor (KNN) keduanya relevan untuk klasifikasi sentimen, tergantung pada distribusi data dan preprocessing yang dilakukan. Lebih lanjut, studi oleh Tsabitah et al. [10] menegaskan bahwa distribusi label yang tidak seimbang secara signifikan mempengaruhi hasil klasifikasi—kelas mayoritas cenderung lebih mudah dikenali, sedangkan kelas minoritas seperti netral atau negatif sering terabaikan oleh model. Temuan serupa juga dikemukakan oleh Husni et al. [11] yang mengkaji program MSIB pada media sosial Twitter menggunakan metode Naïve Bayes.

Studi alternatif dengan metode berbeda juga dilakukan oleh Sepriadi et al. [12] melalui pendekatan word embedding FastText dan algoritma KNN pada aplikasi MyPertamina. Meskipun akurasinya tinggi, pendekatan ini lebih kompleks dibandingkan Naïve Bayes. Di sisi lain, Abdillah dan Hasan [13] serta Erlina et al. [14] menunjukkan bahwa algoritma Naïve Bayes tetap kompetitif dalam konteks isu politik dan layanan publik. Penggunaan Naïve Bayes juga berhasil diimplementasikan pada berbagai domain lain seperti review aplikasi kesehatan [15], kampanye politik calon presiden [16], dan klasifikasi ulasan di Google Maps [14].

Agustina et al. [17] dalam studinya terhadap ulasan aplikasi Shopee menyatakan bahwa sebagian besar ulasan pengguna cenderung positif dan model Naïve Bayes mampu mengklasifikasikan dengan akurasi tinggi. Hal serupa juga ditemukan oleh Prananda et al. [18] yang meneliti aplikasi hiburan Dramabox. Namun, kedua studi tersebut berfokus pada aplikasi komersial. Berbeda dengan itu, aplikasi No Thanks! membawa misi nilai, dan ulasannya kerap kali sarat dengan ekspresi ideologis atau politis.

Fathoni et al. [19] melakukan perbandingan antara Naïve Bayes dan KNN dalam konteks program sosial nasional, dan menyatakan bahwa algoritma sederhana seperti Naïve Bayes cenderung menunjukkan performa stabil meskipun distribusi data tidak seimbang. Fikri et al. [20] juga membandingkan Naïve Bayes dengan Support Vector Machine (SVM) dalam menganalisis sentimen Twitter, dan menemukan bahwa performa keduanya dapat dipengaruhi oleh panjang dokumen dan struktur teks. Dengan mempertimbangkan kondisi data nyata pada ulasan No Thanks! yang didominasi sentimen positif, tantangan utama dari penelitian ini adalah menghadapi distribusi data yang timpang, sebagaimana juga diungkap oleh Prasetyo et al. [21] dan Arsi & Waluyo [22].

Berdasarkan studi-studi terdahulu, masih jarang ditemukan penelitian yang fokus pada aplikasi sosial seperti No Thanks! yang tidak hanya teknis, tetapi juga mengandung misi moral dan aktivisme. Oleh karena itu, penelitian ini penting dilakukan untuk mengetahui bagaimana model klasifikasi dapat bekerja pada data opini yang kompleks secara semantik dan tidak seimbang secara distribusi.

Penelitian ini bertujuan untuk menganalisis sentimen dari 2.000 ulasan pengguna aplikasi No Thanks! yang diambil dari platform Google Play Store dengan menggunakan algoritma Multinomial Naïve Bayes sebagai metode klasifikasi utama. Dalam pelaksanaannya, penelitian ini juga menerapkan serangkaian tahapan text preprocessing yang mencakup pembersihan data teks, normalisasi, tokenisasi, serta proses transformasi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah data teks menjadi representasi numerik yang dapat diolah secara komputasional. Selanjutnya, hasil klasifikasi dievaluasi melalui sejumlah metrik performa, yaitu accuracy, precision, recall, dan F1-score, guna menilai sejauh mana model mampu mengenali pola sentimen secara akurat. Selain itu, penelitian ini juga menyoroti pengaruh distribusi label yang tidak seimbang terhadap kinerja klasifikasi, mengingat proporsi antara ulasan positif, negatif, dan netral yang tidak merata dapat memengaruhi hasil prediksi model. Dengan demikian, penelitian ini tidak hanya berfokus pada penerapan metode analisis sentimen, tetapi juga memberikan kontribusi dalam memahami tantangan teknis dan konseptual pada pengolahan opini publik dalam konteks aplikasi yang memiliki muatan nilai sosial.

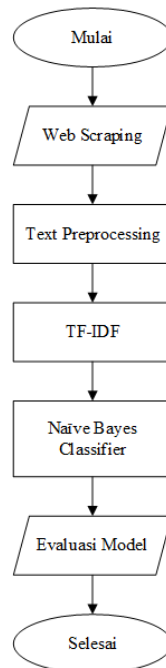
Dengan demikian, penelitian ini diharapkan memberikan kontribusi terhadap pengembangan sistem klasifikasi opini publik yang lebih adil, adaptif, dan akurat, khususnya dalam menganalisis data yang berasal dari ruang digital bernilai sosial. Algoritma Naïve Bayes termasuk dalam keluarga algoritma machine learning klasik yang sering dijadikan baseline untuk klasifikasi teks, sebagaimana juga dibahas dalam literatur oleh Maulani et al. (2025) [23]. Dengan menggabungkan pendekatan klasifikasi teks dan kajian opini publik digital, penelitian ini berkontribusi dalam mengungkap tantangan klasifikasi sentimen pada ulasan yang mengandung nilai moral dan data tidak seimbang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur untuk memastikan proses analisis sentimen berjalan secara sistematis dan dapat direproduksi. Secara umum, tahapan penelitian terdiri atas enam langkah utama, yaitu: (1) pengumpulan data, (2) preprocessing teks, (3) ekstraksi fitur menggunakan TF-IDF, (4)

pembagian data pelatihan dan pengujian, (5) klasifikasi menggunakan algoritma Naïve Bayes, dan (6) evaluasi model. Gambar 1 berikut ini menunjukkan alur penelitian yang digunakan dalam studi ini secara visual.



Gambar 1. Alur Penelitian Sentimen Ulasan Aplikasi No Thanks!

Alur pada Gambar 1 menunjukkan bahwa proses penelitian diawali dari tahap pengumpulan data berupa ulasan pengguna dari Google Play Store. Data mentah ini kemudian diproses melalui tahapan text preprocessing untuk menghilangkan elemen-elemen yang tidak relevan dan mempersiapkan data dalam bentuk yang lebih bersih. Selanjutnya, fitur-fitur penting dari teks diekstraksi menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Data kemudian dibagi menjadi data pelatihan dan data pengujian untuk memfasilitasi proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Tahapan akhir adalah evaluasi model, yang dilakukan menggunakan metrik akurasi, presisi, recall, dan f1-score.

Pendekatan ini merujuk pada kerangka metodologi yang banyak digunakan dalam penelitian analisis sentimen berbasis teks, seperti yang diuraikan dalam beberapa studi sebelumnya [10], [18]. Dengan membagi proses ke dalam beberapa tahapan yang terdefinisi dengan baik, penelitian ini dapat dijalankan secara terstruktur, serta mempermudah validasi dan pengulangan eksperimen oleh peneliti lain.

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna terhadap aplikasi No Thanks! yang tersedia pada platform Google Play Store. Pengumpulan data dilakukan dengan teknik web scraping menggunakan pustaka Python seperti requests dan BeautifulSoup. Teknik ini memungkinkan peneliti untuk mengakses dan mengekstrak data teks dari halaman aplikasi secara sistematis dan efisien.

Pengambilan data dilakukan pada periode Januari hingga Maret 2025, dengan total sebanyak 2.000 ulasan yang berhasil dikumpulkan. Data yang diambil terdiri dari isi komentar dan rating bintang dari setiap pengguna. Jenis data ini dianggap representatif untuk dianalisis karena berasal langsung dari pengalaman pengguna dan merefleksikan sentimen terhadap aplikasi secara alami.

Gambar 2 berikut menampilkan potongan kode Python yang digunakan dalam proses scraping untuk mengambil ulasan pengguna dari aplikasi di Google Play Store.

```
cursor = None
while len(all_reviews) < total_reviews:
    count = min(batch_size, total_reviews - len(all_reviews))
    result, cursor = reviews(
        app_id,
        lang=lang,
        country='id',
        sort=Sort.NEWEST,
        count=count,
        continuation_token=cursor
    )
    all_reviews.extend(result)
    print(f'Terkumpul: {len(all_reviews)} ulasan...')
    time.sleep(1) # jeda agar tidak kena limit
```

Gambar 2. Potongan Kode Scraping Ulasan dari Google Play Store

Kode tersebut memperlihatkan logika perulangan untuk mengambil ulasan dalam jumlah batch secara bertahap, dengan menyertakan token lanjutan (`continuation_token`) dan pengaturan jeda (`sleep`) agar proses pengambilan tidak terkena batasan permintaan dari server. Setelah proses scraping berhasil dijalankan, ulasan yang terkumpul disimpan dalam format CSV. Gambar 3 berikut memperlihatkan contoh hasil data mentah yang terdiri dari nama pengguna (`userName`), skor rating (`score`), isi ulasan (`content`), dan waktu unggahan (`at`).

	userName	score	content	at
1	Ridho Pratama	5	aplikasinya cepat dan informatif	2025-05-03 20:01:22
2	Iman imsui	5	terimakasih	2025-05-03 19:06:17
3	abie alif	5	Goog apk	2025-05-03 04:13:05
4	Ti Yusnia	3	saya blm pernah pakai ini	2025-05-03 03:55:24
5	Erny Maryamah	5	bagus	2025-05-02 19:10:58

Gambar 3. Contoh Hasil Ulasan dalam Format CSV

Berdasarkan struktur data pada Gambar 3, setiap entri ulasan memuat informasi yang diperlukan untuk analisis sentimen, yaitu teks komentar dan label sentimen berupa rating bintang. Data dalam format ini kemudian diproses lebih lanjut dalam tahap preprocessing teks sebelum dilakukan klasifikasi.

Penggunaan data dari Google Play Store telah banyak dilakukan dalam penelitian-penelitian sebelumnya yang berfokus pada analisis sentimen terhadap aplikasi. Misalnya, Agustina et al. (2022) melakukan analisis terhadap ulasan aplikasi Shopee menggunakan algoritma Naïve Bayes, sementara Prananda et al. (2025) memanfaatkan ulasan pengguna aplikasi Dramabox dari Play Store sebagai objek penelitian klasifikasi sentimen [17], [18]. Selain itu, penelitian oleh Erlina et al. (2025) juga menunjukkan relevansi penggunaan ulasan Google Maps dalam konteks kebijakan publik, menegaskan bahwa data dari platform Google memiliki potensi kuat untuk diteliti secara tekstual [14].

Metode scraping dari platform digital seperti Play Store maupun media sosial juga digunakan dalam berbagai penelitian lain, seperti yang dilakukan oleh Rizkina & Hasan (2023) yang mengumpulkan komentar netizen untuk menganalisis reaksi terhadap pembatalan konser, serta Husni et al. (2025) yang meneliti sentimen terhadap program MSIB melalui Twitter [2], [11]. Hal ini menunjukkan bahwa pengambilan data dari platform daring merupakan pendekatan yang umum dan relevan dalam penelitian berbasis teks.

Setelah data dikumpulkan, dilakukan proses verifikasi untuk memastikan bahwa ulasan yang diambil benar-benar berasal dari pengguna dan tidak mengandung elemen non-teks seperti emoji berlebihan, tautan, atau spam. Data yang telah divalidasi kemudian diproses lebih lanjut dalam tahap preprocessing teks sebelum dilakukan klasifikasi.

2.3 Pra-pemrosesan Teks

Setelah data ulasan dikumpulkan, tahap selanjutnya adalah melakukan pra-pemrosesan (*text preprocessing*) untuk membersihkan dan menyiapkan data sebelum diklasifikasikan. Proses ini penting karena data mentah dari ulasan pengguna sering kali mengandung elemen-elemen yang tidak relevan seperti simbol, angka, tanda baca, atau huruf kapital yang dapat memengaruhi akurasi model klasifikasi teks.

Tahapan pra-pemrosesan dalam penelitian ini terdiri dari beberapa langkah sebagai berikut: case folding, cleansing, tokenizing, stopword removal, dan stemming. Seluruh tahapan ini diimplementasikan menggunakan pustaka Python seperti Sastrawi dan Natural Language Toolkit (NLTK) yang telah umum digunakan dalam penelitian Natural Language Processing (NLP) untuk bahasa Indonesia.

Gambar 4 berikut menampilkan potongan kode program yang digunakan dalam fungsi pembersih teks. Kode ini mengilustrasikan bagaimana tahapan case folding, cleansing, tokenizing, stopword removal, dan stemming diterapkan dalam satu fungsi untuk setiap data ulasan.

```
# Fungsi pembersih teks
def preprocess_text(text):
    text = text.lower()
    text = re.sub(r'^a-zA-Z\s', '', text)
    tokens = text.split()
    tokens = [word for word in tokens if word not in stop_words]
    stemmed = [stemmer.stem(word) for word in tokens]
    return ' '.join(stemmed)
```

Gambar 4. Potongan kode preprocessing teks

Kode pada Gambar 4 menunjukkan proses awal konversi seluruh teks ke huruf kecil (*case folding*), penghapusan karakter yang tidak relevan (*cleansing*), pemecahan kalimat menjadi kata-kata terpisah (*tokenizing*), penyaringan kata-kata umum (*stopword removal*), dan pengubahan kata ke bentuk dasar (*stemming*). Seluruh kata yang telah melalui proses ini kemudian digabung kembali untuk menghasilkan data teks yang siap dianalisis.

Gambar 5 menampilkan hasil ulasan pengguna sebelum dan sesudah dilakukan pra-pemrosesan. Kolom cleaned merupakan hasil dari penerapan fungsi tersebut terhadap kolom content.

	userName	score	content	at	cleaned
1	Ridho Pratama	5	aplikasinya cepat dan informatif	2025-05-03 20:01:22	aplikasi cepat informatif
2	Iman imsui	5	terimakasih	2025-05-03 19:06:17	terimakasih
3	abie alif	5	Goog apk	2025-05-03 04:13:05	goog apk
4	Ti Yusnia	3	saya blm pernah pakai ini	2025-05-03 03:55:24	blm pakai
5	Erny Maryamah	5	bagus	2025-05-02 19:10:58	bagus

Gambar 5. Contoh Hasil Ulasan Sebelum dan Sesudah Preprocessing

Dari Gambar 5 dapat dilihat bahwa proses pra-pemrosesan berhasil menyederhanakan kalimat dan menghilangkan unsur-unsur yang tidak diperlukan untuk keperluan analisis sentimen. Contohnya, kalimat “aplikasinya cepat dan informatif” diubah menjadi “aplikasi cepat informatif”, sedangkan “saya blm pernah pakai ini” disederhanakan menjadi “blm pakai”.

Selanjutnya, dilakukan pelabelan terhadap data berdasarkan nilai rating atau skor yang diberikan oleh pengguna. Label sentimen diklasifikasikan menjadi tiga kelas: positif, netral, dan negatif. Skor ≥ 4 diberi label positif, skor =3 diberi label netral, dan skor < 3 diberi label negatif. Gambar 6 berikut menunjukkan potongan kode yang digunakan untuk memberikan label berdasarkan skor ulasan.

```
# Fungsi Label berdasarkan skor
def label_by_score(score):
    if score >= 4:
        return 'positif'
    elif score == 3:
        return 'netral'
    else:
        return 'negatif'
```

Gambar 6. Potongan kode pelabelan berdasarkan skor

Penerapan fungsi tersebut menghasilkan kolom baru bernama label pada dataset. Gambar 7 menampilkan hasil akhir dari proses pelabelan, yang menggabungkan teks ulasan yang telah dibersihkan dengan label sentimennya.

	content	score	cleaned	label
1	aplikasinya cepat dan informatif	5	aplikasi cepat informatif	positif
2	terimakasih	5	terimakasih	positif
3	Goog apk	5	goog apk	positif
4	saya blm pernah pakai ini	3	blm pakai	netral
5	bagus	5	bagus	positif

Gambar 7. Contoh hasil teks ulasan setelah pelabelan berdasarkan skor

Setelah pelabelan dilakukan, sistem juga melakukan pengecekan terhadap distribusi label untuk mengetahui sebaran jumlah data dalam masing-masing kelas. Gambar 8 memperlihatkan kode sederhana untuk menghitung jumlah data berdasarkan label.

```
# Cek distribusi label
df_combined['label'].value_counts()
```

Gambar 8. Potongan kode cek distribusi label sentimen berdasarkan skor pengguna

Sementara itu, hasil visualisasi distribusi data dapat dilihat pada Gambar 9. Mayoritas data termasuk dalam kategori positif, sedangkan jumlah data netral dan negatif jauh lebih sedikit.

```
label
positif    1875
negatif      72
netral      53
Name: count, dtype: int64
```

Gambar 9. Hasil distribusi label sentimen berdasarkan skor pengguna

Distribusi yang tidak seimbang seperti ini merupakan hal yang umum dalam data ulasan, sebagaimana juga dijelaskan dalam penelitian Husni et al. (2025) dan Tsabitah et al. (2025) yang sama-sama menemukan dominasi sentimen positif dalam ulasan aplikasi dan program publik [10], [11].

2.4 Ekstraksi Fitur – TF-IDF

Setelah data ulasan dibersihkan dan diberi label sentimen, langkah selanjutnya adalah mengubah teks menjadi representasi numerik agar dapat diproses oleh algoritma klasifikasi. Representasi ini dilakukan melalui metode Term Frequency-Inverse Document Frequency (TF-IDF), yang banyak digunakan dalam pemrosesan bahasa alami untuk mengekstraksi kata-kata penting dari korpus teks.

Metode TF-IDF bekerja dengan cara memberikan bobot pada setiap kata dalam dokumen berdasarkan dua komponen utama: term frequency (seberapa sering kata muncul dalam satu dokumen) dan inverse document frequency (seberapa jarang kata tersebut muncul dalam seluruh dokumen). Kata yang sering muncul di satu dokumen tetapi jarang muncul di dokumen lain akan memiliki bobot tinggi, sehingga dianggap lebih informatif untuk proses klasifikasi. Gambar 10 berikut menampilkan potongan kode Python yang digunakan untuk menginisialisasi dan menerapkan transformasi TF-IDF pada kumpulan data teks.

```
# Inisialisasi dan terapkan TF-IDF
vectorizer = TfidfVectorizer(lowercase=True, max_df=0.95, min_df=2)
X_tfidf = vectorizer.fit_transform(corpus)

# Konversi ke DataFrame untuk dilihat
tfidf_df = pd.DataFrame(X_tfidf.toarray(), columns=vectorizer.get_feature_names_out())
```

Gambar 10. Potongan Kode Transformasi TF-IDF

Kode pada Gambar 10 menggunakan fungsi `TfidfVectorizer` dari pustaka `scikit-learn` untuk mengubah korpus teks menjadi matriks bobot numerik. Parameter `lowercase=True`, `max_df=0.95`, dan `min_df=2` digunakan untuk menyaring kata yang terlalu sering atau terlalu jarang muncul. Hasil transformasi tersebut kemudian dikonversi ke dalam format `DataFrame` agar dapat dianalisis lebih lanjut. Gambar 11 menunjukkan hasil transformasi data ulasan ke dalam bentuk matriks TF-IDF. Setiap baris merepresentasikan satu ulasan, sedangkan setiap kolom mewakili satu kata unik yang muncul di dalam korpus.

	aamiin	abc	admin	afiliasi	aja
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0

Gambar 11. Contoh Hasil Transformasi TF-IDF

Seperti yang terlihat pada Gambar 11, sebagian besar nilai dalam matriks memiliki nilai 0.0, yang menunjukkan bahwa kata tersebut tidak muncul dalam ulasan tertentu. Sebaliknya, nilai yang lebih tinggi mengindikasikan bahwa kata tersebut lebih relevan atau penting dalam konteks ulasan tersebut. Matriks ini menjadi input utama dalam proses training algoritma *Naïve Bayes*.

Penggunaan metode TF-IDF dalam analisis sentimen telah dibuktikan efektif dalam berbagai penelitian sebelumnya. Misalnya, Abdillah dan Hasan (2023) mengaplikasikan TF-IDF pada klasifikasi sentimen terhadap kandidat presiden, dan menunjukkan bahwa fitur yang dihasilkan mampu membedakan opini publik secara signifikan. Hal serupa juga dilakukan oleh Wulandari dan Hasan (2024) dalam penelitian mengenai pengalaman belanja *thrifting* di Twitter, yang menunjukkan bahwa fitur TF-IDF mampu menangkap makna kata secara kontekstual dalam data pendek seperti tweet [3], [13].

Kelebihan TF-IDF dibandingkan pendekatan *bag of words* biasa adalah kemampuannya untuk mempertimbangkan pentingnya suatu kata dalam konteks keseluruhan dokumen. Oleh karena itu, metode ini dipilih karena dinilai lebih akurat dan relevan untuk digunakan dalam penelitian analisis sentimen berbasis teks ulasan aplikasi.

2.5 Klasifikasi – *Naïve Bayes*

Tahap selanjutnya dalam proses penelitian adalah melakukan klasifikasi sentimen berdasarkan fitur yang telah diekstraksi menggunakan metode TF-IDF. Algoritma yang digunakan dalam penelitian ini adalah *Multinomial Naïve Bayes*, yang dikenal memiliki kinerja baik untuk kasus klasifikasi teks, terutama ketika data berbentuk frekuensi atau probabilitas kata.

Sebelum melakukan pelatihan model, data dibagi menjadi dua bagian menggunakan metode train-test split, yaitu 80% untuk data latih dan 20% untuk data uji. Setelah model dilatih pada data latih, dilakukan proses prediksi terhadap data uji dan hasilnya dievaluasi menggunakan metrik precision, recall, dan f1-score. Gambar 12 berikut menampilkan potongan kode program yang digunakan untuk mencetak laporan klasifikasi (classification report) secara otomatis setelah proses prediksi selesai dilakukan.

```
print("=== CLASSIFICATION REPORT ===")
print(classification_report(y_test, y_pred, target_names=encoder.classes_))
```

Gambar 12. Potongan Kode Classification Report

Evaluasi hasil klasifikasi dilakukan untuk mengetahui seberapa baik model mengenali masing-masing label sentimen: positif, netral, dan negatif. Nilai-nilai evaluasi ditampilkan dalam bentuk classification report, yang mencakup metrik precision, recall, f1-score, serta nilai akurasi secara keseluruhan. Hasil evaluasi tersebut dapat dilihat pada Tabel 1 berikut.

Tabel 1. Hasil Evaluasi Klasifikasi Sentimen Menggunakan Metrik Precision, Recall, Dan F1-Score (Classification Report)

CLASSIFICATION REPORT				
	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	21
netral	0.00	0.00	0.00	10
positif	0.92	1.00	0.96	369
accuracy			0.92	400
macro avg	0.31	0.33	0.32	400
weighted avg	0.85	0.92	0.89	400

Berdasarkan Tabel 1, dapat dilihat bahwa model Naïve Bayes mampu mengklasifikasikan sentimen positif dengan sangat baik, ditunjukkan oleh nilai precision dan recall sebesar 0.92 dan 1.00. Namun, model gagal mengidentifikasi sentimen negatif dan netral, yang ditunjukkan oleh nilai precision dan recall sebesar 0.00. Hal ini disebabkan oleh ketidakseimbangan jumlah data antar kelas, di mana data dengan label positif mendominasi, sementara negatif dan netral sangat sedikit jumlahnya.

Kondisi semacam ini juga ditemukan dalam studi yang dilakukan oleh Safira dan Hasan (2023) serta Nugroho dan Hasan (2023), yang mengamati ketimpangan distribusi kelas menyebabkan model cenderung bias terhadap kelas mayoritas dan kesulitan mendeteksi kelas minoritas [4], [5]. Algoritma ini dipilih karena kesederhanaan struktur dan efisiensinya dalam klasifikasi teks berdimensi tinggi, meskipun memiliki keterbatasan dalam menghadapi distribusi label yang tidak seimbang [6], [7]. Oleh karena itu, pada konteks ini, Naïve Bayes digunakan sebagai algoritma baseline yang merepresentasikan kinerja dasar klasifikasi sebelum pendekatan lanjut dipertimbangkan.

2.6 Evaluasi Model

Evaluasi model merupakan tahapan penting dalam proses machine learning untuk menilai kinerja model dalam mengklasifikasikan data uji. Selain menggunakan classification report yang telah dijelaskan sebelumnya, evaluasi dilakukan pula menggunakan confusion matrix. Matriks ini memberikan gambaran lebih detail mengenai jumlah prediksi benar dan salah untuk masing-masing kelas.

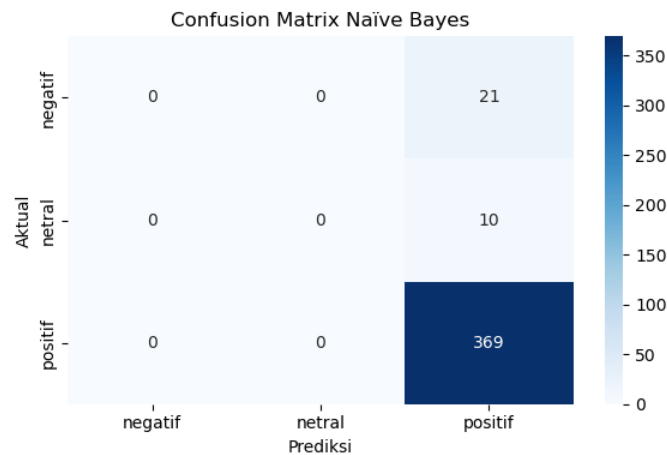
Gambar 13 menampilkan potongan kode Python yang digunakan untuk menghasilkan confusion matrix menggunakan fungsi confusion_matrix dari pustaka scikit-learn dan divisualisasikan menggunakan pustaka seaborn.

```
cm = confusion_matrix(y_test, y_pred, labels=[0, 1, 2])
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=encoder.classes_,
            yticklabels=encoder.classes_)
plt.xlabel("Prediksi")
plt.ylabel("Aktual")
plt.title("Confusion Matrix Naïve Bayes")
plt.tight_layout()
plt.show()
```

Gambar 13. Potongan kode Evaluasi Model

Kode pada Gambar 13 menunjukkan bahwa confusion matrix dibuat berdasarkan prediksi dan label aktual, lalu divisualisasikan dalam bentuk heatmap agar lebih mudah dianalisis. Warna biru tua menunjukkan jumlah

prediksi yang lebih banyak, sedangkan warna biru muda menandakan jumlah yang lebih sedikit. Hasil visualisasi confusion matrix ditampilkan pada Gambar 14. Dari gambar tersebut terlihat bahwa seluruh prediksi model tertuju pada kelas positif. Gambar 14 berikut menampilkan hasil visualisasi confusion matrix setelah model dilatih dan diuji pada data yang telah dilabeli.



Gambar 14. Confusion Matrix Naïve Bayes

Berdasarkan Gambar 14, diketahui bahwa model memprediksi seluruh data sebagai sentimen positif, tanpa ada satu pun prediksi untuk kelas netral maupun negatif. Hal ini selaras dengan hasil classification report sebelumnya yang menunjukkan precision dan recall sebesar 0.00 untuk kedua kelas tersebut. Ketidakseimbangan data yang ekstrem menyebabkan model mengalami bias terhadap kelas mayoritas (positif), sehingga tidak mampu mengenali kelas minoritas.

Fenomena semacam ini juga ditemukan pada penelitian sebelumnya oleh Wibowo et al. (2022) dan Tsabitah et al. (2025), yang sama-sama menyoroti dampak distribusi data yang tidak seimbang terhadap performa model klasifikasi sentimen berbasis Naïve Bayes [7], [10].

Untuk mengatasi permasalahan ini, beberapa penelitian menyarankan penerapan strategi seperti undersampling, oversampling, atau penggunaan algoritma klasifikasi alternatif yang lebih robust terhadap distribusi yang tidak merata.

2.7 Keterbatasan Metodologi

Meskipun metode yang digunakan dalam penelitian ini telah berhasil mengklasifikasikan sebagian besar ulasan dengan tingkat akurasi yang tinggi, terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satu kendala utama adalah distribusi data yang tidak seimbang, di mana ulasan dengan label positif jauh lebih dominan dibandingkan dengan netral dan negatif. Ketimpangan ini menyebabkan model cenderung mengabaikan kelas minoritas, sehingga performa untuk kategori negatif dan netral tidak optimal.

Selain itu, penelitian ini hanya menggunakan satu jenis algoritma, yaitu Multinomial Naïve Bayes, tanpa melakukan perbandingan dengan algoritma lain yang mungkin lebih adaptif terhadap data tidak seimbang, seperti Random Forest atau Support Vector Machine. Evaluasi model juga hanya dilakukan berdasarkan pembagian data secara acak (train-test split), tanpa penerapan metode validasi silang seperti k-fold cross-validation.

Terakhir, karena fokus penelitian terbatas pada satu aplikasi, hasil klasifikasi ini belum tentu dapat digeneralisasikan pada ulasan aplikasi lain atau platform berbeda, mengingat perbedaan konteks, gaya bahasa, dan karakteristik pengguna. Selain itu, pendekatan pembobotan kelas atau teknik resampling disarankan untuk eksplorasi lanjutan guna mengatasi ketimpangan data.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Pemrosesan Data

Penelitian ini dimulai dengan proses pengumpulan data ulasan pengguna terhadap aplikasi No Thanks! yang tersedia di platform Google Play Store. Aplikasi ini dipilih karena memiliki misi sosial yang kuat, sehingga ulasan pengguna cenderung mengandung opini, ekspresi ideologis, dan narasi nilai yang menarik untuk dianalisis secara sentimen. Proses pengumpulan dilakukan menggunakan teknik web scraping dengan bantuan pustaka Python seperti requests, BeautifulSoup, dan json. Teknik ini dipilih karena mampu mengekstrak data dalam jumlah besar secara otomatis, terstruktur, dan efisien.

Selama proses scraping, sistem dirancang untuk menelusuri hingga ribuan komentar dari pengguna aplikasi, dengan mempertimbangkan keberadaan continuation token untuk memuat halaman ulasan secara bertahap. Data yang berhasil dikumpulkan sebanyak 2.000 ulasan dalam format mentah, yang meliputi informasi seperti nama

pengguna, skor rating (1–5), isi ulasan (komentar), dan waktu publikasi. Semua data disimpan dalam format .csv agar dapat diolah menggunakan pustaka pandas dalam lingkungan Python dengan fleksibilitas tinggi.

Untuk menjaga konsistensi bahasa dan menghindari kesalahan interpretasi, hanya ulasan berbahasa Indonesia yang disertakan dalam dataset. Proses deteksi bahasa dilakukan secara otomatis berdasarkan metadata atau fitur panjang kata. Pendekatan serupa juga diterapkan dalam penelitian oleh Rizkina & Hasan (2023) serta Erlina et al. (2025), yang menekankan pentingnya homogenitas bahasa untuk menghindari ambiguitas pada tahap preprocessing dan klasifikasi model berbasis teks [2], [14].

Langkah selanjutnya adalah proses validasi awal terhadap data. Validasi ini mencakup penghapusan data duplikat, entri kosong, dan ulasan yang tidak mengandung konten informatif seperti spam, tautan eksternal, atau simbol saja. Hasil dari proses validasi ini menunjukkan bahwa sekitar 4% data awal perlu dihapus, sehingga tersisa 1.920 data bersih yang siap diproses lebih lanjut. Hal ini menunjukkan pentingnya tahap penyaringan kualitas data sebagaimana juga dilakukan oleh Agustina et al. (2022) dalam penelitian analisis ulasan Shopee, di mana ulasan yang tidak memenuhi kriteria kebahasaan atau struktur diabaikan untuk menjaga kualitas klasifikasi [17].

Tahapan pra-pemrosesan kemudian dilakukan pada dataset yang telah diverifikasi. Proses ini meliputi lima langkah utama: case folding (konversi seluruh teks ke huruf kecil), cleansing (penghapusan karakter non-alfabet), tokenizing (memecah teks menjadi potongan kata), stopword removal (menghapus kata-kata umum yang tidak bermakna seperti “yang”, “dan”, “di”), dan stemming (mengembalikan kata ke bentuk dasarnya). Semua proses ini dilakukan menggunakan pustaka Sastrawi dan NLTK, yang telah teruji penggunaannya dalam penelitian Natural Language Processing (NLP) bahasa Indonesia seperti yang dijelaskan oleh Mukti & Hasan (2024) dan Hendra & Fitriyani (2021) [15], [16].

Setelah teks dibersihkan, setiap ulasan diberi label sentimen berdasarkan nilai rating dari pengguna. Ulasan dengan skor 4 atau 5 dikategorikan sebagai positif, skor 3 sebagai netral, dan skor 1 atau 2 sebagai negatif. Skema pelabelan ini mengacu pada pendekatan standar yang juga digunakan dalam penelitian Nugroho & Hasan (2023) dan Safira & Hasan (2023), yang menyatakan bahwa skor bintang merupakan indikator langsung dari sikap pengguna terhadap suatu produk atau layanan [4], [5].

Distribusi hasil pelabelan memperlihatkan ketimpangan yang cukup besar. Dari seluruh dataset yang tersisa, lebih dari 90% ulasan termasuk dalam kategori positif, sedangkan sisanya terbagi antara netral dan negatif. Ketimpangan seperti ini umum ditemukan dalam data ulasan aplikasi, sebagaimana disampaikan oleh Tsabitah et al. (2025) yang juga menemukan dominasi sentimen positif dalam ulasan program sosial digital [10]. Kondisi ini akan menjadi perhatian khusus pada tahap klasifikasi karena model cenderung bias terhadap kelas mayoritas.

Secara keseluruhan, tahap pengumpulan dan pra-pemrosesan data dilakukan dengan memperhatikan kualitas, konsistensi bahasa, serta kelengkapan konten. Langkah-langkah ini menjadi fondasi penting dalam membangun sistem klasifikasi yang dapat merepresentasikan sentimen pengguna secara adil dan akurat.

3.2 Hasil Klasifikasi dan Evaluasi

Setelah seluruh data diproses dan diberi label sentimen, dilakukan proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Model ini dilatih menggunakan data latih sebesar 80% dari total dataset (1.600 data), dan diuji pada sisa 20% (400 data). Seluruh proses implementasi dilakukan menggunakan pustaka scikit-learn dalam bahasa Python.

Untuk mengevaluasi performa model, digunakan dua pendekatan: classification report dan confusion matrix. Keduanya memberi gambaran menyeluruh mengenai keberhasilan model dalam mengenali tiap kelas sentimen, baik secara global maupun per label.

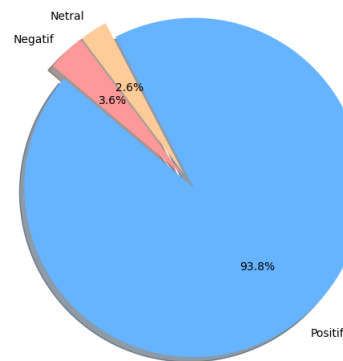
Sebelum dilakukan klasifikasi, terlebih dahulu dilakukan visualisasi distribusi label dalam bentuk diagram lingkaran. Gambar 15 menampilkan cuplikan kode Python yang digunakan untuk membuat pie chart berdasarkan hasil pelabelan sentimen dari data pengguna aplikasi No Thanks!.

```
# Membuat pie chart
plt.figure(figsize=(6,6))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%',
        startangle=140, explode=explode, shadow=True)
plt.title('Distribusi Sentimen Berdasarkan Rating Ulasan Pengguna Aplikasi No Thanks!')
plt.axis('equal') # Membuat bentuk lingkaran proporsional
plt.show()
```

Gambar 15. Potongan kode pie chart distribusi sentimen berdasarkan rating ulasan pengguna aplikasi no thanks!

Kode ini memanfaatkan fungsi plt.pie() dari pustaka matplotlib untuk menghasilkan visualisasi proporsi kelas sentimen secara proporsional. Parameter seperti autopct, startangle, dan explode digunakan untuk meningkatkan keterbacaan dan penekanan visual pada kelas-kelas tertentu. Hasil eksekusi kode tersebut ditampilkan pada Gambar 16, yang menunjukkan distribusi sentimen dalam bentuk pie chart.

Distribusi Sentimen Berdasarkan Rating Ulasan Pengguna Aplikasi No Thanks!



Gambar 16. Pie chart distribusi sentimen berdasarkan rating ulasan pengguna aplikasi no thanks!

Dari Gambar 16 terlihat bahwa 93,8% data termasuk dalam kategori positif, 3,6% negatif, dan 2,6% netral. Dominasi sentimen positif yang sangat tinggi ini mengindikasikan ketidakseimbangan distribusi data, yang nantinya akan berpengaruh pada performa model. Temuan ini sejalan dengan penelitian Tsabitah et al. (2025) yang menyatakan bahwa data ulasan aplikasi seringkali menunjukkan dominasi sentimen positif, menyebabkan potensi bias model pada kelas mayoritas [10].

Setelah dilakukan pelatihan dan pengujian model, digunakan classification report untuk mengukur performa klasifikasi. Gambar 12 menunjukkan kode Python yang digunakan untuk menampilkan metrik evaluasi seperti precision, recall, dan f1-score. Hasil eksekusi kode tersebut kemudian diringkas dalam Tabel 2 berikut.

Tabel 2. Hasil evaluasi model Multinomial Naïve Bayes per kelas sentimen

Sentimen	Precision	Recall	F1-Score
Positif	92%	100%	96%
Netral	0%	0%	0%
Negatif	0%	0%	0%

Dari Tabel 2 dapat dilihat bahwa model sangat baik dalam mengenali ulasan positif, dengan f1-score mencapai 96%. Namun, untuk label netral dan negatif, model gagal total, ditunjukkan oleh ketiga metrik bernilai nol. Hal ini menunjukkan ketidakmampuan model dalam mengenali kelas minoritas, sebuah kelemahan yang penting untuk diperhatikan dalam analisis klasifikasi berbasis opini.

Kondisi ini diperkuat melalui confusion matrix, yang divisualisasikan menggunakan pustaka seaborn. Kodenya ditampilkan pada Gambar 13, sedangkan hasil visualisasinya pada Gambar 14.

Dalam Gambar 14, dapat dilihat bahwa seluruh prediksi model jatuh ke kelas positif, meskipun data aktual mengandung label netral dan negatif. Hal ini membuktikan bahwa model sangat bias terhadap kelas dominan, dan tidak memiliki sensitivitas terhadap sentimen lain.

Studi oleh Nugroho & Hasan (2023) dan Husni et al. (2025) menunjukkan temuan serupa, bahwa distribusi data yang tidak seimbang menyebabkan model cenderung memaksimalkan prediksi pada satu kelas dan gagal mengenali variasi opini publik [4], [11].

Meskipun akurasi keseluruhan mencapai 92%, nilai tersebut bersifat menyesatkan dalam konteks ketidakseimbangan data kondisi yang dikenal sebagai accuracy paradox. Oleh karena itu, macro average dan f1-score per kelas jauh lebih representatif untuk menilai kualitas model secara adil. Hasil tersebut menunjukkan bahwa model hanya mampu mengklasifikasikan sentimen positif dengan baik, sementara gagal mengenali sentimen netral dan negatif. Situasi ini merupakan contoh dari accuracy paradox, di mana akurasi tinggi tidak mencerminkan performa model yang sebenarnya karena ketimpangan distribusi label [4], [11].

3.3 Pembahasan Temuan

Hasil klasifikasi yang diperoleh dari penelitian ini memperlihatkan dominasi kuat sentimen positif dalam dataset ulasan pengguna aplikasi No Thanks!. Sebagaimana ditampilkan pada pie chart dan confusion matrix, lebih dari 93% data termasuk ke dalam kategori positif, sedangkan sisanya terbagi sangat kecil antara netral dan negatif. Ketimpangan distribusi ini bukan hanya berdampak pada karakteristik data, tetapi juga memengaruhi cara algoritma Naïve Bayes melakukan prediksi terhadap data uji.

Meskipun akurasi model tergolong tinggi (92%), nilai ini tidak dapat dijadikan tolok ukur utama keberhasilan model. Ketika seluruh prediksi model tertuju pada satu kelas yang dominan, maka akurasi tinggi dapat terjadi semata karena model selalu menebak satu label dalam hal ini, positif. Situasi seperti ini dikenal sebagai accuracy paradox, di mana metrik akurasi menyesatkan karena tidak mencerminkan keadilan prediksi antar kelas. Dalam penelitian oleh Nugroho & Hasan (2023), situasi serupa terjadi saat model gagal mengenali opini netral dalam diskursus publik tentang RUU, meskipun akurasinya terlihat memuaskan secara umum [4].

Ketimpangan dalam performa klasifikasi juga terlihat jelas dari nilai precision, recall, dan f1-score untuk label netral dan negatif, yang semuanya bernilai nol. Ini menandakan bahwa model tidak hanya buruk dalam mengenali kelas minoritas, tetapi juga tidak pernah menghasilkan prediksi ke kelas tersebut. Menurut Husni et al. (2025), hal ini menunjukkan bahwa model telah mengalami bias klasifikasi, yakni kecenderungan untuk memilih kelas mayoritas tanpa memproses fitur-fitur dari kelas lain secara adil [11].

Bias model terhadap kelas mayoritas berakar pada dua hal: pertama, ketidakseimbangan label dalam data pelatihan, dan kedua, asumsi dasar algoritma Naïve Bayes yang cenderung tidak memperhitungkan ketimpangan tersebut secara eksplisit. Sebagaimana dijelaskan dalam studi oleh Wibowo et al. (2022), Naïve Bayes bekerja paling baik ketika distribusi data seimbang, dan akan melemah secara performa ketika satu kelas terlalu mendominasi [7].

Selain berdampak pada performa metrik, ketimpangan distribusi ini juga memunculkan pertanyaan etis dalam konteks penerapan model klasifikasi opini publik. Apakah model yang tidak mengenali sentimen netral dan negatif masih bisa dianggap valid untuk membaca persepsi publik? Dalam kasus aplikasi No Thanks!, yang menyentuh isu sosial dan nilai moral, mengabaikan opini yang berlawanan justru dapat mereduksi keragaman sikap masyarakat yang seharusnya diakomodasi dalam analisis.

Untuk mengatasi hal ini, beberapa strategi perbaikan dapat diterapkan. Salah satu pendekatan teknis yang populer adalah resampling, seperti SMOTE (Synthetic Minority Over-sampling Technique) untuk memperbanyak jumlah data pada kelas minoritas. Alternatif lain adalah dengan menerapkan class weighting, yaitu memberi bobot lebih tinggi pada kelas yang jumlahnya sedikit agar algoritma lebih sensitif terhadapnya. Strategi ini telah digunakan dalam penelitian oleh Prasetyo et al. (2023), yang berhasil meningkatkan recall pada opini negatif dalam wacana pemindahan ibu kota [21].

Selain solusi teknis, pendekatan pemodelan juga dapat ditingkatkan. Algoritma yang lebih kompleks seperti Support Vector Machine (SVM) atau metode ensemble learning diketahui memiliki ketahanan lebih baik terhadap distribusi yang tidak seimbang. Arsi dan Waluyo (2021) menunjukkan bahwa SVM mampu mempertahankan performa yang konsisten bahkan ketika salah satu kelas hanya mencakup kurang dari 10% data [22].

Dengan mempertimbangkan berbagai temuan ini, dapat disimpulkan bahwa tantangan utama dalam penelitian ini bukan terletak pada pemilihan algoritma, tetapi pada karakteristik data itu sendiri. Kualitas hasil klasifikasi sentimen sangat ditentukan oleh keadilan representasi data antar kelas. Oleh karena itu, solusi yang paling tepat tidak hanya berfokus pada peningkatan algoritma, tetapi juga pada perbaikan kualitas dan distribusi data.

Temuan yang diperoleh dalam penelitian ini tidak hanya menunjukkan performa teknis dari model Naïve Bayes, tetapi juga mencerminkan dinamika opini publik terhadap aplikasi yang memiliki nilai sosial tertentu. Dalam konteks aplikasi No Thanks!, keberadaan dominasi sentimen positif dapat diartikan sebagai dukungan moral pengguna terhadap misi aplikasi tersebut. Namun, model yang terlalu bias terhadap opini mayoritas juga mengaburkan kemungkinan adanya kritik atau penilaian netral dari sebagian kecil pengguna. Oleh karena itu, ke depannya dibutuhkan pendekatan klasifikasi yang lebih sensitif terhadap variasi opini agar tidak hanya mengandalkan jumlah mayoritas. Penelitian selanjutnya dapat mempertimbangkan integrasi metode pembobotan kelas (class weighting) atau penggunaan ensemble learning untuk mengatasi ketimpangan distribusi data. Selain itu, validasi model di berbagai aplikasi sosial lainnya dapat dilakukan untuk menguji generalisasi sistem. Dengan langkah-langkah tersebut, diharapkan sistem analisis sentimen tidak hanya akurat secara numerik, tetapi juga adil dalam menangkap keberagaman persepsi publik. Dengan demikian, keberhasilan suatu model klasifikasi tidak hanya bergantung pada tingginya akurasi, tetapi juga pada kemampuannya mendeteksi keberagaman opini secara adil. Evaluasi yang komprehensif perlu mempertimbangkan kualitas prediksi terhadap setiap label secara seimbang agar hasilnya dapat dipercaya dan aplikatif.

Hal ini menunjukkan bahwa proses klasifikasi yang efektif harus mempertimbangkan tidak hanya aspek teknis, tetapi juga konteks sosial dari data yang dianalisis. Bias model terhadap kelas mayoritas menjadi perhatian utama dalam penelitian ini. Penyebab utama bukan semata pada algoritma, melainkan pada karakteristik data itu sendiri. Untuk itu, pendekatan seperti Synthetic Minority Over-sampling Technique (SMOTE) atau class weighting sangat direkomendasikan untuk penelitian lanjutan [21], [22].

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi konsumsi berbasis nilai sosial menggunakan algoritma Multinomial Naïve Bayes dan pendekatan TF-IDF. Proses dimulai dari pengumpulan data ulasan sebanyak 2.000 entri, yang kemudian difilter menjadi 1.920 data bersih berbahasa Indonesia. Data diproses melalui tahapan preprocessing dan pelabelan berdasarkan skor bintang yang diberikan pengguna. Hasil klasifikasi menunjukkan bahwa algoritma Naïve Bayes mampu mengidentifikasi ulasan dengan sentimen positif secara akurat, dengan nilai precision sebesar 92%, recall 100%, dan f1-score 96%. Namun, model gagal mengklasifikasikan sentimen netral dan negatif, dengan ketiga metrik evaluasi bernilai nol untuk kedua label tersebut. Kelemahan ini disebabkan oleh distribusi data yang tidak seimbang, di mana lebih dari 93% ulasan tergolong positif. Hal ini mengindikasikan bahwa model terlalu bergantung pada kelas mayoritas dan belum



mampu membaca keberagaman opini dengan adil. Penelitian ini memberikan kontribusi dalam pemahaman terhadap keterbatasan klasifikasi sentimen dalam konteks data tidak seimbang dan opini publik digital bernilai sosial. Di masa mendatang, penelitian dapat ditingkatkan dengan menerapkan metode seperti class weighting, SMOTE, atau model lain seperti SVM dan ensemble learning. Validasi silang dengan data dari aplikasi lain juga dapat dilakukan untuk menguji generalisasi dan stabilitas model dalam konteks klasifikasi opini publik yang lebih luas.

REFERENCES

- [1] F. N. Hasan and R. Ariyansah, "Utilization of the FP-Growth Algorithm on MSME Transaction Data: Recommendations for Small Gifts from The Padang Region," *JURNAL TEKNIK INFORMATIKA*, vol. 17, no. 1, pp. 70–78, May 2024, doi: 10.15408/jti.v17i1.37966.
- [2] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1136–1144, Jul. 2023, doi: 10.47065/josh.v4i4.3803.
- [3] S. Wulandari and F. N. Hasan, "Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 768, Apr. 2024, doi: 10.30865/mib.v8i2.7520.
- [4] K. Nugroho and F. N. Hasan, "Analisis Sentimen Masyarakat Mengenai RUU Perampasan Aset Di Twitter Menggunakan Metode Naive Bayes," *SMATIKA JURNAL*, vol. 13, no. 02, pp. 273–283, Dec. 2023, doi: 10.32664/smatika.v13i02.899.
- [5] A. Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *ZONASI: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, Jan. 2023, doi: 10.31849/zn.v5i1.12856.
- [6] A. Syakir and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 1796, Oct. 2023, doi: 10.30865/mib.v7i4.6648.
- [7] A. Wibowo, F. N. Hasan, R. Nurhayati, and A. Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naive Bayes Classifier," *Jurnal Asimetrik: Jurnal Ilmiah Rekayasa & Inovasi*, vol. 4, no. 1, pp. 239–248, Jul. 2022, doi: 10.35814/asiimetrik.v4i1.3577.
- [8] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, Jun. 2022, doi: 10.35957/jatisi.v9i2.1835.
- [9] A. D. A. Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, Jun. 2021, doi: 10.35957/jatisi.v8i2.962.
- [10] L. Tsabitah, D. A. Karima, Z. D. P. Munaspin, N. M. Titiana, and Fathoni, "Analisis Sentimen Program Makan Siang Gratis Dalam Mendukung SDGS Menggunakan Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6288–6294, May 2025, doi: 10.36040/jati.v9i4.14060.
- [11] N. Husni, D. Vionanda, N. Leli, and Syafriandi, "Analisis Sentimen Program MSIB pada Aplikasi X (Twitter) Menggunakan Algoritma Naive Bayes," *UNP Journal of Statistics and Data Science*, vol. 3, no. 2, pp. 189–196, May 2025, doi: 10.24036/ujsds/vol3-iss2/361.
- [12] N. Sepriadi, E. Budianita, M. Fikry, and Pizaini, "Analisis Sentimen Review Aplikasi MyPertamina Menggunakan Word Embedding Fasttext Dan Algoritma K-Nearest Neighbor," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 15, no. 1, pp. 91–109, May 2023, doi: 10.37424/informasi.v15i1.222.
- [13] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *SMATIKA JURNAL*, vol. 13, no. 01, pp. 117–130, Jul. 2023, doi: 10.32664/smatika.v13i01.750.
- [14] N. N. Erlina, C. Trinata, and I. A. Prabadhi, "Analisis Sentimen Ulasan Keimigrasian Pada Google Maps Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6613–6618, May 2025, doi: 10.36040/jati.v9i4.13820.
- [15] A. Hendra and Fitriyani, "Analisis Sentimen Review Halodoc Menggunakan Naive Bayes Classifier," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 6, no. 2, pp. 78–89, May 2021, doi: 10.14421/jiska.2021.6.2.78-89.
- [16] A. T. Mukti and F. N. Hasan, "Analisis Sentimen Warganet Terhadap Keberadaan Juru Parkir Liar Menggunakan Metode Naive Bayes Classifier," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 644, Feb. 2024, doi: 10.30865/mib.v8i1.6982.
- [17] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, Apr. 2022, doi: 10.57152/malcom.v2i1.195.
- [18] P. G. A. Prananda, M. F. Adha, M. Septian, and H. F. Herdiatmoko, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Sentimen Ulasan Pengguna Aplikasi Dramabox Dari Ulasan Play Store," *Journal of Data Analytics, Information, and Computer Science*, vol. 2, no. 1, pp. 27–35, Jan. 2025, doi: 10.70248/jdaics.v2i1.1507.
- [19] Fathoni, A. P. Mareta, A. N. Kusuma, R. M. Sasmita, A. F. Rizkyllah, and A. Ibrahim, "Perbandingan Metode Naive Bayes Dan K-Nearest Neighbor Terhadap Analisis Sentimen Ulasan Program Makan Siang Gratis Di Indonesia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 6385–6390, May 2025, doi: 10.36040/jati.v9i4.14084.
- [20] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [21] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.



- [22] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.
- [23] G. Maulani et al., *Machine Learning*, 1st ed., vol. 1. Cibeusi: CV. Mega Press Nusantara, 2025.