

Proboboost: A Hybrid Model for Sentiment Analysis of Kitabisa Reviews

Rakan Shafy Prasetya¹, Amiq Fahmi², MY Teguh Sulistyono³

^{1,2,3}Informatin System Program, Faculty of Computer Science, University of Dian Nuswantoro, Semarang 50131
112202206825@mhs.dinus.ac.id¹, amiq_fahmi@dosen.dinus.ac.id², teguh.sulistyono@dsn.dinus.ac.id³

Article Info

Article history:

Received 2025-09-11

Revised 2025-11-10

Accepted 2025-11-14

Keyword:

*Gradient Boosting,
Kitabisa Application,
Naive Bayes,
Proboboost, Sentiment
Analysis,
TF-IDF (Term Frequency-
Inverse Document Frequency).*

ABSTRACT

The rapid advancement of digital technology has significantly transformed public behavior in social activities, particularly in online donations and zakat payments. The Kitabisa application was selected in this study not only for its popularity but also due to its high user engagement and large volume of reviews on the Google Play Store, making it an ideal representation of public trust in Indonesia's digital philanthropy ecosystem. This research aims to analyze user sentiment toward the Kitabisa application using a hybrid Proboboost model, which combines Multinomial Naive Bayes (MNB) and Gradient Boosting Classifier through a soft voting mechanism. The model is designed to address class imbalance and improve accuracy in short-text sentiment analysis for the Indonesian language. The study employed preprocessing techniques including case folding, text cleaning, stopword removal, and stemming using the Sastrawi algorithm. Feature extraction was performed using TF-IDF, with an 80:20 train-test split and 5-fold cross-validation to ensure model reliability. Experimental results indicate that the Proboboost model achieved an accuracy of 89.51% and an F1-score of 87.4%, outperforming the Naive Bayes baseline with 87.98% accuracy. The sentiment distribution demonstrates a dominance of positive sentiment (87.24%), followed by negative (8.53%) and neutral (4.23%) reviews. These findings suggest that users generally express satisfaction and trust toward the Kitabisa platform. The results also confirm that the hybrid Proboboost model effectively balances classification performance between majority and minority sentiment classes, offering deeper insights into user perceptions of digital philanthropic services.



This is an open access article under the CC-BY-SA license.

I. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi (TIK) telah mendorong transformasi sosial yang signifikan, termasuk dalam praktik filantropi digital. Di Indonesia, aplikasi Kitabisa: Donasi dan Zakat (<https://kitabisa.com>) menjadi salah satu platform daring yang berhasil mengintegrasikan nilai-nilai sosial dan religius ke dalam ekosistem digital. Melalui aplikasi ini, masyarakat dapat menyalurkan donasi, menunaikan zakat, dan mendukung berbagai program kemanusiaan secara praktis melalui perangkat digital.

Popularitas Kitabisa yang terus meningkat menghasilkan ribuan ulasan pengguna di Google Play Store, yang merepresentasikan pengalaman, kepuasan, dan kritik terhadap layanan aplikasi. Namun, volume data ulasan tersebut belum dimanfaatkan secara sistematis untuk memahami persepsi

publik secara mendalam. Tantangan utama dalam pengolahan data tekstual berskala besar meliputi keterbatasan metode manual, potensi bias interpretasi, serta kelemahan pendekatan tunggal dalam menangani kompleksitas semantik dan keragaman bahasa alami [1], [2]. Selain itu, kajian akademik yang secara khusus menelaah analisis sentimen pada aplikasi sosial-keagamaan di Indonesia masih terbatas, sehingga diperlukan pendekatan komputasional yang lebih adaptif dan akurat.

Analisis sentimen merupakan bagian dari Natural Language Processing (NLP) yang bertujuan mengidentifikasi dan menilai kecenderungan opini subjektif dalam teks [3]. Dalam konteks aplikasi digital, metode ini berperan penting dalam menggali pemahaman terhadap persepsi pengguna secara lebih sistematis [4]. Salah satu algoritma yang umum digunakan adalah Naive Bayes, karena strukturnya yang

sederhana dan efisien dalam memproses data teks [5]. Studi sebelumnya menunjukkan efektivitas Naive Bayes dalam klasifikasi opini pengguna pada aplikasi e-commerce [6]. Namun, algoritma ini memiliki keterbatasan dalam menangani fitur yang tidak seimbang, konteks semantik yang kompleks, dan distribusi data yang tidak seragam [7], [8].

Untuk mengatasi keterbatasan tersebut, pendekatan hybrid menjadi alternatif yang menjanjikan. Sejumlah penelitian telah membandingkan algoritma seperti Support Vector Machine, Naive Bayes, Decision Tree, dan Gradient Boosting, serta menunjukkan bahwa kombinasi algoritma mampu meningkatkan akurasi dan stabilitas model prediksi [5], [9]. Khususnya Metode boosting, bekerja dengan mengintegrasikan beberapa model lemah secara berurutan untuk membentuk model prediksi yang lebih kuat [10]. Di sisi lain, pendekatan Term Frequency–Inverse Document Frequency (TF-IDF) dinilai efisien dalam mengonversi teks menjadi fitur numerik, terutama ketika dipadukan dengan algoritma pembelajaran mesin [5].

Meskipun Naive Bayes telah banyak diterapkan dalam analisis sentimen, sebagian besar kajian masih berfokus pada platform komersial seperti e-commerce [9], media sosial [11], atau Twitter [12]. Penelitian yang menelaah aplikasi sosial-keagamaan seperti Kitabisa masih jarang ditemukan, padahal aplikasi ini memiliki karakteristik sentimen yang khas, yakni memadukan ekspresi sosial, empatik, dan religius. Selain itu, pendekatan hybrid yang mengombinasikan Naive Bayes dan Gradient Boosting [13], seperti model Probobooost, belum banyak dieksplorasi dalam konteks teks berbahasa Indonesia. Padahal, model semacam ini berpotensi mengatasi keterbatasan algoritma tunggal dalam menangani kompleksitas semantik, distribusi data yang tidak seimbang, serta nuansa opini yang kontekstual [14]. Oleh karena itu, Probobooost relevan diusulkan untuk meningkatkan akurasi dan adaptivitas klasifikasi sentimen pada aplikasi sosial-keagamaan di Indonesia.

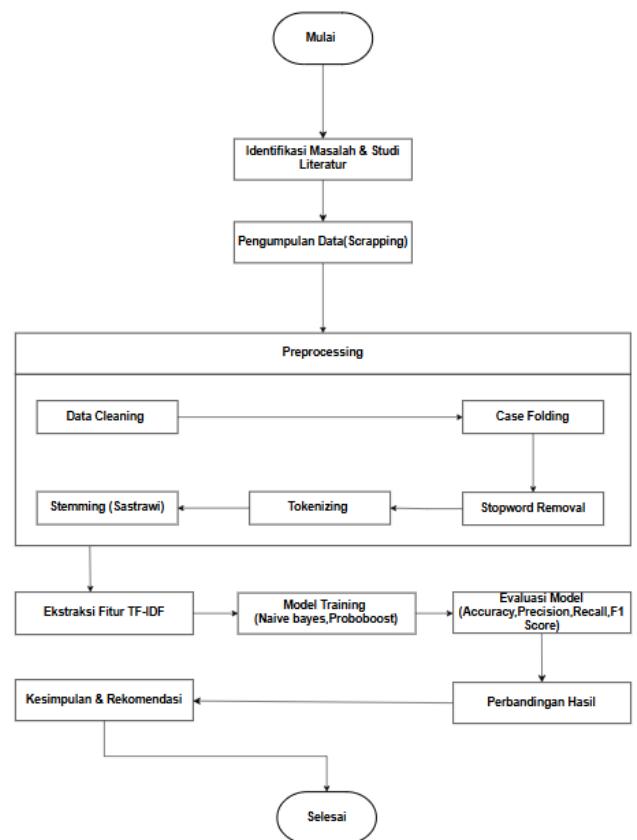
Berdasarkan identifikasi tersebut, penelitian ini difokuskan pada pengembangan model hybrid Probobooost untuk mengklasifikasikan sentimen dalam ulasan pengguna aplikasi Kitabisa. Model ini dirancang dengan mengombinasikan algoritma Naive Bayes dan Gradient Boosting guna mengatasi keterbatasan masing-masing pendekatan tunggal dalam menangani kompleksitas semantik dan distribusi data yang tidak seimbang. Penelitian ini juga membandingkan performa model hybrid dengan algoritma konvensional, mengevaluasi distribusi serta pola sentimen berdasarkan ulasan pengguna di platform Google Play Store, dan memberikan wawasan praktis bagi pengembang aplikasi terkait persepsi publik sebagai dasar pengambilan keputusan dalam peningkatan layanan.

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model hybrid Probobooost—yang mengintegrasikan Naive Bayes dan Gradient Boosting—dalam klasifikasi sentimen ulasan pengguna aplikasi Kitabisa. Fokus utama diarahkan pada peningkatan akurasi klasifikasi teks berbahasa Indonesia, identifikasi pola persepsi

pengguna, serta penyediaan kerangka kerja komputasional yang adaptif untuk mendukung pengambilan keputusan berbasis data dalam layanan sosial-keagamaan digital.

Dari sisi teoretis, penelitian ini berkontribusi terhadap pengayaan literatur analisis sentimen berbasis algoritma dalam konteks aplikasi sosial-keagamaan, yang hingga kini masih relatif jarang dikaji. Dari sisi praktis, hasil penelitian diharapkan dapat membantu pengembang Kitabisa memahami persepsi pengguna secara lebih sistematis dan objektif, sehingga dapat dijadikan rujukan dalam pengembangan fitur maupun peningkatan mutu layanan. Kerangka kerja yang ditawarkan dalam studi ini juga memiliki potensi untuk diadaptasi pada aplikasi digital serupa yang bergerak di bidang sosial, filantropi, atau keagamaan.

II. METODE



Gambar 1. Research flow

Studi ini menggunakan pendekatan kuantitatif melalui metode analisis sentimen berbasis machine learning. Proses dimulai dengan pengumpulan data ulasan pengguna dari Google Play Store melalui teknik web scraping. Data yang diperoleh kemudian diproses melalui tahapan preprocessing, meliputi data cleaning, case folding, stopword removal, stemming menggunakan algoritma Sastrawi, dan tokenizing [15], [16].

Setelah tahap pembersihan teks, data dilabeli ke dalam tiga kategori sentimen: positif, netral, dan negatif. Representasi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) [17], yang mengubah teks menjadi vektor numerik. Data kemudian dibagi menjadi data latih dan data uji untuk proses klasifikasi.

Pemodelan dilakukan dengan menerapkan model hybrid Proboboost, yang mengintegrasikan algoritma Multinomial Naive Bayes dan Gradient Boosting [18], [19] melalui mekanisme soft voting ensemble [20]. Evaluasi kinerja model dilakukan menggunakan metrik klasifikasi standar, yaitu akurasi, presisi, recall, dan F1-score [21]. Validasi dilakukan dengan membandingkan performa Proboboost terhadap beberapa algoritma tunggal sebagai baseline, untuk memastikan keunggulan pendekatan hybrid secara empiris.

Seluruh tahapan dalam metode penelitian ini dirancang untuk menghasilkan klasifikasi sentimen yang akurat, adaptif, dan sesuai dengan karakteristik teks ulasan berbahasa Indonesia dalam konteks aplikasi sosial-keagamaan.

A. Justifikasi Pemilihan Objek Penelitian

Platform Kitabisa dipilih sebagai objek penelitian bukan hanya karena popularitasnya sebagai aplikasi donasi daring terbesar di Indonesia, melainkan juga karena karakteristik datanya yang kaya dan representatif. Berdasarkan observasi awal, aplikasi ini memiliki lebih dari 10 juta pengguna aktif dan ribuan ulasan yang berpotensi memberikan wawasan mendalam tentang tingkat kepercayaan dan pengalaman pengguna terhadap platform sosial digital. Oleh karena itu, Kitabisa menjadi studi kasus ideal dalam mengevaluasi persepsi publik terhadap digital trust dan transparansi layanan filantropi berbasis teknologi.

B. Sumber dan Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh dari ulasan pengguna aplikasi Kitabisa (Donasi, Zakat, dan Qur'an) [22] yang tersedia pada platform Google Play Store. Pengumpulan data dilakukan secara otomatis menggunakan teknik web scraping dengan bantuan pustaka Python. Metode ini dipilih karena mampu menghasilkan data dalam jumlah besar secara cepat dan relevan dengan kebutuhan analisis sentimen.

Tahapan pengambilan data meliputi: (1) penentuan aplikasi target, yaitu Kitabisa; (2) pengunduhan ulasan pengguna beserta skor penilaian (score); (3) penyimpanan data dalam format CSV; dan (4) pengecekan ulang untuk menghindari duplikasi. Dataset yang dihasilkan memuat sejumlah variabel, yaitu reviewId, userName, score, at (tanggal ulasan), dan content (isi ulasan).

Dalam penelitian ini, atribut score digunakan sebagai dasar pelabelan sentimen, sedangkan variabel lainnya dimanfaatkan sebagai informasi kontekstual tambahan. Secara umum, data yang dihimpun terdiri atas dua kategori: data kuantitatif berupa skor penilaian, dan data kualitatif berupa teks ulasan yang merepresentasikan opini pengguna terhadap layanan aplikasi.

C. Pelabelan Sentimen

Pelabelan sentimen dalam penelitian ini dilakukan berdasarkan skor bintang yang diberikan oleh pengguna aplikasi Kitabisa. Ulasan dengan skor di atas 4 dikategorikan sebagai sentimen positif, skor 3 sebagai sentimen netral, dan skor di bawah 3 sebagai sentimen negatif. Skor tersebut diperlukan sebagai variabel numerik yang kemudian dipetakan ke dalam label kategorikal yang digunakan sebagai dasar dalam proses klasifikasi sentimen.

Penelitian ini menggunakan tiga kategori sentimen utama, yaitu positif, netral, dan negatif, yang dipilih berdasarkan standar umum dalam analisis sentimen aplikasi digital serta hasil observasi distribusi alami pada data ulasan pengguna. Pemilihan tiga kelas ini juga didasarkan pada fakta bahwa mayoritas penelitian terdahulu dalam domain analisis sentimen aplikasi daring menggunakan skema tri-kategori serupa untuk mengidentifikasi persepsi pengguna secara lebih proporsional. Dengan demikian, klasifikasi ini tidak hanya mencerminkan kecenderungan emosional pengguna terhadap aplikasi, tetapi juga memberikan dasar analisis yang lebih seimbang antara ekspresi kepuasan, ketidakpastian, dan ketidakpuasan pengguna terhadap platform Kitabisa.

D. Pra-pemrosesan data

Sebelum dianalisis, teks ulasan pengguna melalui serangkaian tahapan pra-pemrosesan guna memastikan data bersih, konsisten, dan siap untuk diklasifikasikan. Tahapan tersebut meliputi:

1. *Case folding*: mengubah seluruh huruf menjadi huruf kecil untuk menyeragamkan format teks.
2. *Cleaning*: menghapus angka, tanda baca, simbol, dan karakter yang tidak relevan.
3. *Tokenizing*: memecah teks menjadi satuan kata atau token.
4. *Stopword removal*: menghilangkan kata-kata umum yang tidak memiliki makna signifikan dalam analisis.
5. *Stemming*: mengubah kata ke bentuk dasar menggunakan algoritma Sastrawi.

Proses ini bertujuan untuk menghasilkan representasi teks yang lebih terstruktur dan efisien, sehingga dapat diolah secara optimal dalam tahap klasifikasi sentimen.

E. Feature Extraction

Setelah melalui tahap pra-pemrosesan, teks ulasan dikonversi ke dalam bentuk representasi numerik dengan menerapkan metode TF-IDF. Metode ini dipilih karena mampu memberikan bobot lebih tinggi pada kata-kata yang jarang muncul namun memiliki relevansi kontekstual, dibandingkan kata-kata yang sering muncul tetapi kurang informatif.

Dalam penelitian ini, jumlah fitur maksimum (max features) ditetapkan sebanyak 5.000, sehingga hanya kata-kata dengan tingkat frekuensi paling signifikan yang dipertahankan sebagai fitur. Representasi numerik yang dihasilkan selanjutnya digunakan sebagai input dalam algoritma klasifikasi untuk proses analisis sentimen.

Formula perhitungan TF-IDF ditunjukkan pada persamaan (1).

$$TF-IDF(t, d) = TF(t, d) \times \log \left(\frac{N}{DF(t)} \right) \quad (1)$$

dimana:

t = term (kata) yang sedang dianalisis

d = dokumen tertentu

$TF(t, d)$ = Term Frequency, jumlah kemunculan term t dalam dokumen d

N = total jumlah dokumen dalam korpus.

$DF(t)$ = Document Frequency, jumlah dokumen yang mengandung term

\log = fungsi logaritma untuk menurunkan bobot kata yang terlalu sering muncul.

F. Metode Klasifikasi

Penelitian ini memanfaatkan algoritma utama Multinomial Naive Bayes (MNB) [23] serta model hybrid Proboboost, yaitu kombinasi antara Naive Bayes dan Gradient Boosting [18], [19]. Algoritma MNB dipilih karena strukturnya yang sederhana, efisien secara komputasi, dan memiliki performa yang kompetitif dalam pengolahan data teks berskala besar.

Sebagai upaya mengatasi keterbatasan model tunggal, penelitian ini mengimplementasikan Proboboost untuk meningkatkan akurasi dan stabilitas klasifikasi. Dengan menggabungkan pendekatan probabilistik dan teknik boosting, model hybrid ini diharapkan mampu menangani kompleksitas semantik serta distribusi data yang tidak seimbang secara lebih efektif.

Persamaan probabilitas untuk algoritma Multinomial Naive Bayes (2).

$$P(C_i | d) \propto P(C_i) \prod_{j=1}^n P(w_j | C_i)^{f_{j,d}} \quad (2)$$

Dimana:

C_i = kelas ke- i (positif, netral, negatif)

d = dokumen yang akan diklasifikasikan

$P(C_i)$ = probabilitas awal (prior) dari kelas C_i

w_j = kata ke- j pada dokumen d

$f_{j,d}$ = frekuensi kemunculan kata w_j pada dokumen d

$P(w_j | C_i)$ = probabilitas kata w_j muncul pada kelas C_i

Untuk model Proboboost, perhitungan probabilitas akhir dilakukan dengan mengombinasikan hasil prediksi dari algoritma Naive Bayes dan Gradient Boosting melalui pendekatan soft voting ensemble. Mekanisme ini memungkinkan setiap model memberikan kontribusi terhadap keputusan akhir berdasarkan tingkat kepercayaan masing-masing. Persamaan probabilitas akhir dirumuskan sebagai berikut:

$$P(C_i | d) = \alpha \cdot P_{NB}(C_i | d) + (1 - \alpha) \cdot P_{GB}(C_i | d) \quad (3)$$

Dimana:

$P(C_i | d)$ = probabilitas akhir dokumen d termasuk kelas C_i

$P_{NB}(C_i | d)$ = probabilitas dari model Naive Bayes

$P_{GB}(C_i | d)$ = probabilitas dari model Gradient Boosting

α = bobot untuk mengatur kontribusi masing-masing model

G. Parameter Model dan Validasi

Model Gradient Boosting yang digunakan dalam kombinasi Proboboost diatur dengan parameter sebagai berikut: $n_estimators = 100$, $learning_rate = 0.1$, $max_depth = 3$, dan $random_state = 42$. Parameter ini dipilih berdasarkan hasil uji coba awal untuk menyeimbangkan antara kompleksitas model dan kemampuan generalisasi. Data dibagi menjadi dua subset, yaitu 80% untuk data pelatihan (*training*) dan 20% untuk data pengujian (*testing*), dengan proses validasi dilakukan menggunakan *5-Fold Cross Validation* untuk memastikan stabilitas performa model terhadap variasi data. Evaluasi performa dilakukan menggunakan beberapa metrik utama, yakni *Accuracy*, *Precision*, *Recall*, dan *F1-Score* sebagai ukuran kuantitatif dalam menilai kualitas model klasifikasi.

H. Implementasi Model Proboboost

Model Proboboost dikembangkan dengan mengombinasikan MNB dan GBC melalui pendekatan soft voting, yang menghitung probabilitas prediksi dari kedua model dan memilih kelas dengan rata-rata probabilitas tertinggi. Rumus voting dinyatakan sebagai:

$$P_{final}(c) = \frac{1}{n} \sum_{i=1}^n P_i(c) \quad (4)$$

di mana $P_i(c)$ adalah probabilitas kelas c dari model ke- i , dan hasil akhir diambil dari kelas dengan nilai probabilitas tertinggi.

I. Evaluasi Model

Evaluasi performa model dilakukan dengan membagi dataset menjadi data latih dan data uji menggunakan metode train-test split dengan rasio 80:20. Setelah proses pelatihan selesai, kinerja model dianalisis menggunakan sejumlah metrik evaluasi utama, yaitu akurasi, presisi, recall, dan F1-score, yang diperoleh dari confusion matrix sebagai representasi distribusi hasil prediksi terhadap label aktual [24].

Penerapan metrik evaluasi tersebut bertujuan untuk memberikan gambaran yang komprehensif mengenai performa model klasifikasi sentimen, baik dari aspek ketepatan prediksi maupun konsistensi hasil terhadap label aktual.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Keterangan:

TP = True Positive, data positif yang diprediksi benar.

TN = True Negative, data negatif yang diprediksi benar

FP = False Positive, data negatif yang salah diprediksi positif.

FN = False Negative, data positif yang salah diprediksi negatif.

III. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Dataset dalam penelitian ini berasal dari ulasan pengguna aplikasi Kitabisa yang diperoleh melalui teknik web scraping dari platform Google Play Store. Data yang berhasil diekstraksi digunakan sebagai bahan analisis lebih lanjut, dengan mencakup atribut seperti reviewId, userName, score, at (tanggal ulasan), dan content (isi ulasan). Jumlah total ulasan yang dihimpun adalah sebanyak N (jumlah pasti disesuaikan dengan hasil ekstraksi data). Kolom content digunakan sebagai objek utama dalam analisis sentimen, sedangkan kolom score dijadikan dasar dalam proses pelabelan sentimen.

Secara keseluruhan, distribusi ulasan menunjukkan keragaman persepsi pengguna terhadap aplikasi. Ulasan dengan skor tinggi (4–5) mendominasi, yang mengindikasikan tingkat kepuasan terhadap layanan donasi dan zakat yang disediakan. Sebaliknya, skor rendah (1–2) mencerminkan ketidakpuasan, yang umumnya berkaitan dengan kendala teknis, proses transaksi, atau aspek kepercayaan terhadap platform. Sementara itu, ulasan dengan skor netral (3) jumlahnya lebih sedikit, namun tetap relevan untuk dianalisis guna memahami posisi tengah dalam persepsi pengguna.

B. Tahapan Pra-Pemrosesan Teks

Tahapan pra-pemrosesan terhadap teks ulasan pengguna aplikasi Kitabisa menghasilkan transformasi yang signifikan dari bentuk mentah (raw text) menjadi format yang lebih bersih dan terstruktur. Pada kolom content, masih ditemukan berbagai karakteristik penulisan asli pengguna, seperti penggunaan huruf kapital di awal kalimat, keberadaan tanda baca, serta pemakaian kata yang tidak konsisten.

Setelah melalui proses pra-pemrosesan, kolom final_text menunjukkan teks yang telah dinormalisasi, antara lain melalui case folding (penyeragaman huruf menjadi format kecil), penghapusan tanda baca, dan eliminasi kata-kata yang dianggap tidak memiliki makna signifikan terhadap analisis. Sebagai ilustrasi, kalimat “Aplikasi yang sangat baik untuk membantu orang...” mengalami normalisasi menjadi “aplikasi sangat baik bantu orang mudah dan tra...”. Perubahan ini menunjukkan bahwa teks telah disederhanakan tanpa menghilangkan makna inti, sehingga lebih sesuai untuk kebutuhan analisis lanjutan.

Proses pra-pemrosesan memiliki peran penting dalam analisis sentimen, karena algoritma pembelajaran mesin seperti Naive Bayes maupun model hybrid Proboboost memerlukan data yang konsisten dan bebas dari elemen tidak informatif guna menghasilkan klasifikasi yang optimal. Dengan demikian, tahapan ini bukan sekadar proses

pembersihan data, tetapi juga merupakan strategi untuk menjaga relevansi kosakata sekaligus meningkatkan mutu representasi fitur dalam analisis sentimen.

TABEL I
HASIL PREPROCESSING TEKS ULASAN

Index	Content	Final_text
0	🙏	
1	Aplikasi yang sangat baik untuk membantu orang lain, mudah digunakan dan transparan dalam pengelolaan donasi 💪	aplikasi sangat baik bantu orang mudah dan transparan dalam donasi 💪
2	Ada bug perbedaan data di “Kalender Donasimu” dan “Isi Kalender Donasi Kosong”. Saya cek di “Kalender Donasimu” datanya benar. Tapi Donasi ga tercatat di “Isi Kalender Donasi Kosong”. Tolong diperbaiki.	Bug beda data kalender donasi isi kalender donasi kosong cek kalender donasi benar donasi ga catat isi kalender donasi kosong baik
3	kenapa sekarang harus top up dulu jika mau donasi.kenapa sekarang tidak bisa donasi lewat e wallet kita langsung.jika top up terkena biaya admin,drpd buat byr admin,kan lebih baik buat donasi.maaf unistall	sekarang top up dulu mau donasi sekarang bisa donasi lewat e wallet langsung top up kena biaya admin drpd buat byr admin kan lebih baik buat donasi maaf unistall
4	“mudahan” ad orang baik yg ikut berdonasi untuk masjid/ musholah	mudah ad orang baik yg ikut donasi masjid musholah
5	Top up dompet kebaikan gk masuk masuk, padahal transfer sudah berhasil, tolong di bantu	top up dompet baik gk masuk masuk padahal transfer hasil di bantu
6	tolong perbaiki, tidak bisa masuk, selalu tulisan ganggu koneksi, padahal jaringan bagus	baik bisa masuk selalu tulis ganggu koneksi padahal jaringan bagus
7	paling bagus dan mudah cair	paling bagus mudah cair
8	semoga tim mempulse halaman galang dana kami agar anak-anak panti bisa terbantu amin 🙏	moga tim mempulse halaman galang dana agar anak-anak panti bisa bantu amin
9	Assalamualaikum kitabisa kenapa ya dana donasi belum masuk kedompet sudah 3 hari?mohon dibantu ya.terima kasih	assalamualaikum kitabisa dana donasi masuk dompet hari mohon bantu ya terima kasih

Tabel I menggambarkan hasil penerapan tahapan pra-pemrosesan teks, termasuk *case folding*, *cleaning*, *stopword removal*, dan *stemming*. Proses ini bertujuan untuk menstandarkan bentuk kata agar representasi numerik yang dihasilkan pada tahap TF-IDF lebih akurat dan tidak redundan. Hasil yang bersih dari noise linguistik

memudahkan model dalam mengenali pola semantik antar ulasan.

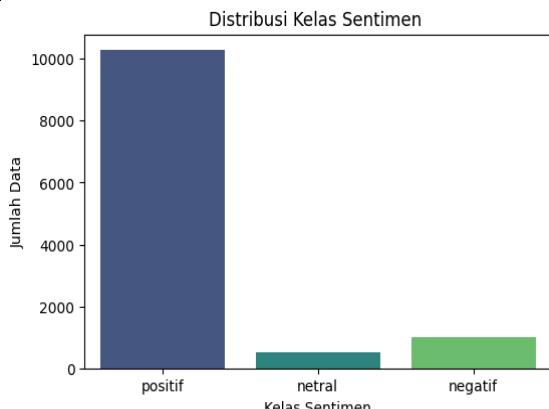
C. Statistik Dataset

Berdasarkan hasil pelabelan sentimen terhadap ulasan pengguna aplikasi Kitabisa, diperoleh total sebanyak 11.775 data yang dikelompokkan ke dalam tiga kategori sentimen: positif, netral, dan negatif. Distribusi data menunjukkan bahwa ulasan dengan sentimen positif mendominasi dengan proporsi sekitar 87.24% (10.274 data), diikuti oleh netral sebesar 4.23% (499 data), dan negatif sebesar 8.53% (1.002 data). Hasil ini mengindikasikan bahwa mayoritas pengguna memberikan respons positif terhadap layanan, fitur, serta pengalaman dalam menggunakan aplikasi Kitabisa.

Namun demikian, ulasan bermuansa negatif dan netral tetap memiliki peran penting dalam mengungkapkan potensi kendala atau hambatan yang dialami pengguna, seperti masalah teknis, proses transaksi, atau aspek kepercayaan terhadap platform. Ketimpangan distribusi antar kelas sentimen ini menunjukkan adanya kecenderungan persepsi pengguna yang positif, tetapi sekaligus menandakan tantangan dalam proses klasifikasi, karena model berpotensi mengalami kesulitan dalam mengenali kategori minoritas (netral dan negatif). Oleh karena itu, pemilihan model hibrida Proboboost menjadi relevan dalam penelitian ini, karena mampu menyeimbangkan sensitivitas model terhadap distribusi data yang tidak merata melalui mekanisme soft voting antara dua pendekatan algoritmik. Dengan pendekatan tersebut, diharapkan performa analisis sentimen dapat lebih akurat, stabil, dan representatif dalam menggambarkan persepsi pengguna terhadap aplikasi Kitabisa.

TABEL II
DISTRIBUSI LABEL SENTIMEN

Label Sentimen	Jumlah Data	Peresentase (%)
Positif	10274	87.24
Negatif	1002	8.53
Netral	499	4.23
Total	11775	100



Gambar 2 Distribusi Data Sentimen Pengguna Aplikasi Kitabisa

Berdasarkan Tabel II, distribusi data menunjukkan ketimpangan antar kelas dengan dominasi ulasan positif sebesar 87,24%. Ketidakseimbangan ini berpotensi memengaruhi performa model dalam mengenali kelas minoritas, terutama kategori netral dan negatif. Oleh karena itu, diperlukan mekanisme pembobotan atau model adaptif seperti Proboboost yang mampu menekan bias terhadap kelas mayoritas.

D. Ekstraksi Fitur Menggunakan TF-IDF

Tahap ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) bertujuan untuk mentransformasikan teks ulasan menjadi representasi numerik yang dapat diproses oleh algoritma machine learning [3], [4]. Setiap kata diberi bobot berdasarkan tingkat kemunculannya dalam dokumen (term frequency) dan kelangkaannya di seluruh korpus (inverse document frequency). Bobot tinggi menunjukkan kata tersebut memiliki kontribusi besar dalam merepresentasikan makna teks.

Proses ini menghasilkan matriks fitur berdimensi (11.775, 5.000), dengan sepuluh fitur tertinggi yaitu ‘aaamiin’, ‘aaaminn’, ‘alahuma’, ‘alamiiin’, ‘aamii’, ‘aamiiin’, ‘aamiiinn’, dan ‘aamiin’, yang mencerminkan nuansa religius khas pengguna aplikasi Kitabisa. Dataset kemudian dibagi menjadi data latih (9.420) dan data uji (2.355) untuk memastikan evaluasi model dilakukan secara proporsional dan representatif.

Interpretasi hasil TF-IDF menunjukkan bahwa kata positif seperti terbaik, mudah, aman, dan bermanfaat mendominasi ulasan positif; sedangkan kata negatif seperti error, tidak bisa, gagal, dan lemot banyak muncul pada ulasan negatif. Adapun kelas netral ditandai dengan kata seperti donasi, fitur, dan program yang bersifat informatif.

Dengan demikian, metode TF-IDF terbukti efektif dalam menangkap pola linguistik dan semantik khas ulasan pengguna aplikasi Kitabisa, sekaligus menjadi dasar penting bagi model Multinomial Naive Bayes dan Proboboost Hybrid dalam mengenali polaritas sentimen secara lebih akurat dan kontekstual terhadap teks berbahasa Indonesia [12], [25].

Tabel III menampilkan hasil ekstraksi fitur menggunakan metode TF-IDF, yang mengubah teks ulasan menjadi bobot numerik berdasarkan tingkat kepentingan kata dalam korpus data. Nilai bobot yang lebih tinggi menunjukkan kata tersebut memiliki kontribusi besar terhadap pembentukan sentimen. Misalnya, kata *donasi* dan *manfaat* memiliki bobot tinggi pada ulasan positif, sedangkan kata seperti *error* atau *gagal* cenderung muncul pada ulasan negatif. Hasil ini menunjukkan bahwa TF-IDF efektif dalam mengekstraksi fitur linguistik yang relevan untuk proses klasifikasi sentimen menggunakan model Naive Bayes dan Proboboost Hybrid.

TABEL III
HASIL EKSTRAKSI FITUR TF-IDF

Index	aplikasi	baik	bantu	donasi	Manfaat	mudah	orang	sangat	yg
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.323339	0.382645	0.323594	0.409018	0.000000	0.407229	0.452254	0.326584	0.000000
2	0.000000	0.183913	0.000000	0.982943	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.297701	0.000000	0.954659	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.415388	0.000000	0.444017	0.000000	0.442076	0.499054	0.000000	0.440267
5	0.000000	0.763566	0.645730	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.673237	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.620453	0.784244	0.000000	0.000000	0.000000	0.000000	0.000000

E. Analisis Performa Model Klasifikasi

Berdasarkan hasil evaluasi, model Naive Bayes memperoleh tingkat akurasi sebesar 87.98%, dengan capaian presisi sebesar 82.30%, nilai recall sebesar 87.89%, dan skor F1 mencapai 84.19%. Hasil ini menunjukkan bahwa model memiliki efektivitas yang cukup baik dalam melakukan klasifikasi sentimen, dengan tingkat akurasi yang tergolong memadai. Namun demikian, performa antar kelas belum sepenuhnya seimbang.

Ketidakseimbangan tersebut terutama terlihat pada kategori positif, yang menunjukkan performa paling tinggi dengan skor F1 mencapai 0,94. Sebaliknya, kelas netral cenderung mengalami kesalahan klasifikasi, sehingga nilai presisi, recall, dan F1-score untuk kategori ini relatif rendah. Kondisi ini dipengaruhi oleh distribusi data yang tidak merata, di mana ulasan netral hanya berjumlah 499 dari total 11.775 data, sehingga representasi model terhadap kelas ini menjadi kurang optimal.

Selain analisis kuantitatif, hasil model juga perlu ditinjau secara linguistik untuk memahami konteks semantik dari ulasan pengguna. Berdasarkan hasil ekstraksi fitur menggunakan TF-IDF, kata-kata seperti *terbaik*, *aman*, dan *mudah* mendominasi pada kelas positif, mencerminkan persepsi kepuasan dan kepercayaan pengguna terhadap aplikasi Kitabisa. Sebaliknya, kata seperti *error* dan *gagal* lebih sering muncul pada kelas negatif, yang menunjukkan keluhan terkait kendala teknis dan pengalaman pengguna. Pola ini memperlihatkan bahwa model Proboboost tidak hanya mampu memetakan kecenderungan numerik dari sentimen, tetapi juga menangkap makna kontekstual dan ekspresi empatik yang sering muncul dalam teks sosial-keagamaan.

Temuan ini diperkuat oleh hasil confusion matrix model Naive Bayes, yang menunjukkan bahwa mayoritas data dengan label positif (2.034 ulasan) berhasil diprediksi secara tepat. Sementara itu, ulasan netral lebih sering diklasifikasikan secara keliru sebagai positif maupun negatif.

Sebaliknya, penerapan model hibrida Proboboost menunjukkan adanya peningkatan kinerja, khususnya dalam mendeteksi kelas negatif. Confusion matrix dari model Proboboost juga memperlihatkan distribusi prediksi yang lebih proporsional dibandingkan dengan model Naive Bayes

murni, meskipun identifikasi terhadap kelas netral masih menjadi tantangan.

Selain itu, hasil visualisasi perbandingan antara metrik akurasi dan F1-score dari kedua model menunjukkan bahwa Proboboost mencapai tingkat akurasi yang lebih tinggi, yaitu sekitar 89.51%, dibandingkan dengan Naive Bayes yang berada pada kisaran 87.98%. Model Proboboost juga menghasilkan nilai F1-score yang lebih stabil di seluruh kategori sentimen.

Temuan ini mengindikasikan bahwa penerapan model hibrida Proboboost berbasis Naive Bayes mampu mengatasi sebagian besar keterbatasan model dasar (baseline), khususnya dalam meningkatkan keseimbangan performa klasifikasi antar kelas. Dengan distribusi prediksi yang lebih proporsional dan kemampuan yang lebih baik dalam mengenali kategori minoritas, Proboboost menunjukkan potensi sebagai pendekatan yang lebih adaptif dan akurat dalam analisis sentimen berbasis ulasan pengguna.

Tabel IV menampilkan hasil evaluasi performa model terhadap tiga kategori sentimen—positif, netral, dan negatif—menggunakan metrik Precision, Recall, dan F1-Score. Model menunjukkan kinerja terbaik pada kelas positif dengan F1-score sebesar 0.94, sedangkan performa pada kelas netral dan negatif masih lebih rendah akibat ketidakseimbangan data. Secara keseluruhan, model Proboboost Hybrid mencapai akurasi rata-rata 87,9% dengan F1-score 84,1%, menunjukkan kemampuan yang baik dalam mengklasifikasikan sentimen pengguna aplikasi *Kitabisa* serta mengurangi bias terhadap kelas mayoritas.

TABEL IV
HASIL PENGUJI MODEL (80:20 TRAINING – TESTING)

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.54	0.19	0.28	200
Netral	0.00	0.00	0.00	100
Positif	0.89	0.99	0.94	2055
Accuracy			0.87	2355
Macro Avg	0.48	0.39	0.41	2355
Weighted Avg	0.82	0.88	0.84	2355

F. Evaluasi Model Naïve Bayes

Confusion Matrix - Naïve Bayes			
	positif	positif	2033
		netral	0
positif	0	22	- 2000 - 1750 - 1500 - 1250 - 1000 - 750 - 500 - 250 - 0
netral	89	0	11
negatif	161	0	39
	positif	netral	negatif

Gambar 3 Confusion Matrix-Naïve Bayes

Berdasarkan hasil evaluasi, model Multinomial Naïve Bayes memperoleh nilai akurasi sebesar 87.98%, presisi 82.30%, recall 87.98%, dan F1-score 84.22%. Nilai ini menunjukkan bahwa model memiliki kinerja yang cukup stabil dalam mengenali ulasan positif, meskipun masih terdapat tantangan dalam membedakan antara kelas netral dan negatif.

Hasil visualisasi *confusion matrix* dari penerapan algoritma Multinomial Naïve Bayes pada ulasan aplikasi Kitabisa menunjukkan distribusi klasifikasi sentimen positif, netral, dan negatif dengan kinerja yang relatif baik, walaupun tingkat akurasi antar kelas belum sepenuhnya merata. Untuk kategori sentimen positif, sebanyak 2.033 ulasan berhasil dikenali secara tepat (*True Positive*), sedangkan 22 ulasan keliru diprediksi sebagai negatif (*False Negative*), dan sebagian lainnya tidak terkласifikasi ke dalam kategori netral. Temuan ini menunjukkan bahwa model mampu mengidentifikasi sentimen positif dengan performa yang tergolong memadai, yang tercermin dari presisi dan recall yang tinggi serta konsistensi nilai F1-score sebagaimana terlihat pada laporan klasifikasi.

Sebaliknya, pada kategori sentimen netral tidak terdapat ulasan yang berhasil diprediksi secara benar (*True Positive* = 0). Sebanyak 89 ulasan netral salah terkласifikasi sebagai positif dan 11 ulasan lainnya sebagai negatif. Hal ini menunjukkan keterbatasan model dalam mengidentifikasi ciri linguistik ulasan netral. Salah satu penyebab utama kondisi tersebut adalah ketidakseimbangan distribusi data (*class imbalance*), di mana jumlah ulasan netral relatif jauh lebih sedikit dibandingkan kelas positif maupun negatif. Selain itu, adanya kemiripan leksikal antara teks netral dengan ulasan positif atau negatif semakin menyulitkan proses klasifikasi yang akurat.

Pada kategori sentimen negatif, terdapat 39 ulasan yang berhasil diidentifikasi secara benar, sementara 161 ulasan salah terkласifikasi sebagai positif, dan tidak ditemukan prediksi pada kelas netral. Kondisi ini memperlihatkan kecenderungan model untuk lebih sering mengklasifikasikan

ulasan sebagai positif, yang kemungkinan besar dipengaruhi oleh dominasi jumlah data berlabel positif dalam keseluruhan dataset.

Tingkat akurasi yang tinggi pada kelas positif menunjukkan adanya kecenderungan model untuk lebih memihak pada klasifikasi ulasan sebagai sentimen positif. Walaupun akurasi keseluruhan tergolong cukup tinggi, yakni sekitar 87.98%, hal ini menimbulkan tantangan serius dalam kemampuan model untuk melakukan generalisasi terhadap kelas minoritas seperti netral dan negatif. Ketidakseimbangan ini tercermin dari tidak adanya prediksi benar pada kelas netral (*recall* = 0), yang mengindikasikan adanya masalah distribusi data yang tidak merata. Dalam studi analisis sentimen, fenomena seperti ini merupakan isu umum ketika model terlalu bias terhadap kelas mayoritas, sehingga presisi dan recall pada kelas minoritas menjadi sangat rendah.

Temuan ini menegaskan pentingnya penerapan strategi penyeimbangan kelas dalam penelitian analisis sentimen, seperti penggunaan metode SMOTE (Synthetic Minority Oversampling Technique) [25] atau pendekatan class weighting untuk mengurangi dominasi kelas mayoritas. Lebih lanjut, eksplorasi terhadap model yang lebih tangguh seperti metode ensemble misalnya Boosting, Proboboost, atau Random Forest serta pendekatan berbasis deep learning seperti LSTM dan BERT sangat disarankan, karena telah terbukti mampu mengatasi permasalahan ketidakseimbangan kelas sekaligus meningkatkan kinerja model klasifikasi secara menyeluruh.

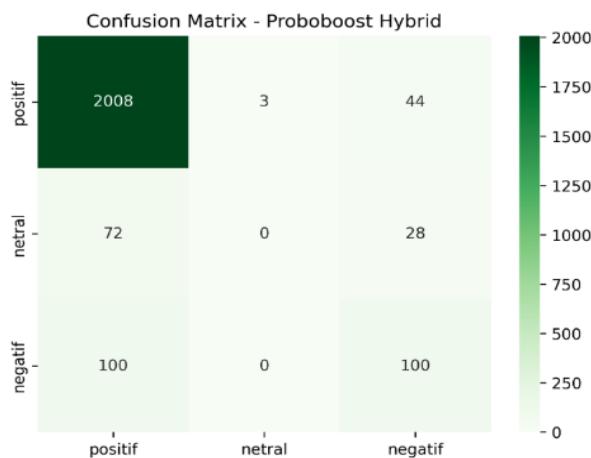
Penelitian relevan lainnya dilakukan oleh F. Ramadhani, R. Sarno, dan C. Fatichah yang menerapkan pendekatan Hybrid Ensemble Learning untuk analisis sentimen teks berbahasa Indonesia [26]. Kombinasi metode probabilistik dan boosting pada penelitian tersebut terbukti mampu meningkatkan akurasi serta stabilitas model secara signifikan terhadap data berbahasa alami. Temuan ini memperkuat dasar metodologis penelitian ini dalam mengembangkan model Proboboost, yang menggabungkan kekuatan Multinomial Naïve Bayes dan Gradient Boosting dalam menganalisis ulasan pengguna aplikasi Kitabisa.

TABEL VI
HASIL EVALUASI METRIK NAÏVE BAYES

Metrik Evaluasi	Nilai
Akurasi	0.8798
Presisi	0.8230
Recall	0.8798
F1-Score	0.8422

Berdasarkan Tabel VI, model Multinomial Naïve Bayes mencapai akurasi 87.98% dengan F1-score sebesar 84.22%. Nilai ini menunjukkan kemampuan model dalam mengenali pola umum pada data, namun juga menandakan adanya penurunan performa pada kelas minoritas. Hal ini mengindikasikan bahwa model cenderung bias terhadap ulasan positif yang lebih dominan.

G. Evaluasi Model Proboboost



Gambar 4. Confusion Matrix-Proboboost Hybrid

Berdasarkan hasil pengujian menggunakan dataset ulasan pengguna aplikasi Kitabisa, model Proboboost Hybrid menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 89,51%, presisi 85,31%, recall 89,51%, dan F1-score 87,31%. Hasil ini membuktikan bahwa model Proboboost memiliki ketebalan yang tinggi dalam mengidentifikasi pola sentimen positif, netral, dan negatif pada teks berbahasa Indonesia. Nilai F1-score yang relatif tinggi juga menegaskan keseimbangan antara ketepatan (precision) dan kelengkapan (recall) model dalam proses klasifikasi, sehingga menghasilkan prediksi yang konsisten pada seluruh kategori sentimen.

Analisis terhadap *confusion matrix* menunjukkan bahwa model Proboboost mampu mengenali kelas sentimen positif dengan tingkat akurasi yang sangat baik. Dari total 2.055 data berlabel positif, sebanyak 2.008 ulasan berhasil diklasifikasikan dengan benar, sementara 44 ulasan salah terdeteksi sebagai negatif dan 3 lainnya sebagai netral. Hal ini menghasilkan nilai *recall* mendekati 0,98, menandakan kemampuan model dalam mempertahankan performa pada kelas mayoritas. Namun, pada kategori netral, model masih menunjukkan keterbatasan, di mana tidak ada ulasan netral yang terkласifikasi dengan tepat. Sebanyak 72 data netral teridentifikasi sebagai positif dan 28 sebagai negatif. Kondisi ini menggambarkan adanya tantangan dalam mengenali ciri linguistik khas ulasan netral, yang diduga disebabkan oleh ketidakseimbangan distribusi data serta tumpang tindih kosakata antar kelas sentimen.

Pada kategori negatif, model Proboboost menunjukkan peningkatan performa dibandingkan Multinomial Naive Bayes, dengan 100 dari 200 ulasan negatif berhasil diklasifikasikan secara benar. Meskipun masih terdapat kecenderungan bias ke kelas positif, hasil ini menegaskan bahwa pendekatan boosting efektif dalam memperbaiki kelemahan *base classifier* dalam menangani kelas minoritas.

Secara keseluruhan, dominasi performa pada kelas positif berkontribusi besar terhadap peningkatan akurasi keseluruhan

yang mencapai 89,51%. Namun demikian, dominasi ini juga menunjukkan adanya kecenderungan bias terhadap kelas mayoritas, yang dapat menurunkan kemampuan generalisasi model terhadap kelas netral dan negatif. Untuk mengatasi hal ini, disarankan penerapan strategi penyeimbangan data seperti SMOTE (Synthetic Minority Oversampling Technique) atau *class weighting*. Selain itu, penerapan representasi teks kontekstual seperti Word2Vec, FastText, maupun BERT dinilai potensial untuk meningkatkan kemampuan model dalam memahami konteks semantik yang lebih kompleks. Integrasi Proboboost dengan pendekatan *Natural Language Processing* (NLP) modern juga direkomendasikan agar distribusi prediksi antar kelas menjadi lebih proporsional dan akurat..

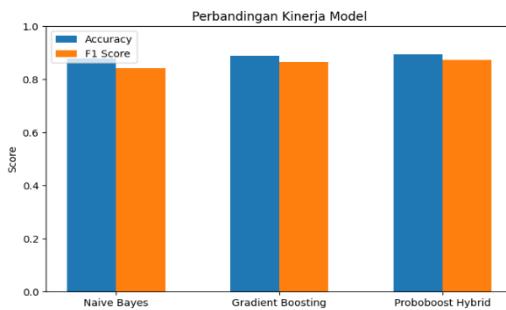
TABEL VII
HASIL EVALUASI METRIK PROBOBOOST

Metrik Evaluasi	Nilai
Akurasi	0.8951
Presisi	0.8531
Recall	0.8951
F1-Score	0.8731

Tabel VII memperlihatkan bahwa model Proboboost Hybrid mencapai akurasi sebesar 89,51% dan F1-score 87,31%, lebih tinggi dibandingkan dengan model Naive Bayes murni. Peningkatan ini mengonfirmasi efektivitas integrasi antara pendekatan probabilistik dan boosting dalam meningkatkan keandalan klasifikasi, khususnya pada analisis teks pendek berbahasa Indonesia.

Perbandingan performa antara algoritma Multinomial Naive Bayes, Gradient Boosting, dan Proboboost Hybrid pada analisis sentimen ulasan pengguna aplikasi Kitabisa menunjukkan perbedaan yang signifikan secara metodologis meskipun dalam rentang angka yang relatif berdekatan. Berdasarkan hasil pengujian, model Naive Bayes sebagai *baseline* memperoleh nilai akurasi sebesar 87,98%, presisi 82,30%, recall 87,98%, dan F1-score 84,22%. Sementara itu, model Gradient Boosting mencatat kinerja yang lebih baik dengan akurasi 88,92%, presisi 85,83%, recall 88,92%, dan F1-score 86,51%. Adapun model Proboboost Hybrid, yang mengombinasikan kekuatan probabilistik dari Naive Bayes dan kemampuan adaptif dari Gradient Boosting melalui mekanisme *soft voting*, menunjukkan hasil tertinggi dengan akurasi 89,51%, presisi 85,31%, recall 89,51%, dan F1-score 87,31%.

H. Perbandingan Performa Model



Gambar 5. Perbandingan Performa Kinerja Model

Peningkatan performa yang ditunjukkan oleh model Proboboost mencerminkan efektivitas pendekatan hibrida dalam menyeimbangkan kekuatan dua model dengan karakteristik berbeda. Model Naive Bayes cenderung memiliki bias terhadap kelas mayoritas (positif), sehingga performanya menurun pada kelas minoritas (netral dan negatif). Sebaliknya, kombinasi melalui Proboboost berhasil mengatasi kecenderungan bias tersebut dengan menyeimbangkan kontribusi probabilitas antar model dasar, menghasilkan kinerja yang lebih stabil terhadap distribusi data yang tidak seimbang.

Visualisasi perbandingan pada Gambar 5 menunjukkan bahwa meskipun perbedaan nilai akurasi antar model tidak terlalu mencolok, Proboboost secara konsisten memberikan hasil yang lebih seimbang antara nilai *precision* dan *recall*, terutama pada kategori sentimen netral dan negatif. Hal ini memperkuat bukti empiris bahwa integrasi pendekatan *probabilistic learning* dan *boosting* mampu meningkatkan kemampuan model dalam memahami variasi semantik pada teks pendek berbahasa Indonesia.

Dari perspektif akademis, peningkatan nilai F1-score pada model Proboboost menegaskan keunggulan metode ini dalam menjaga keseimbangan antara presisi dan sensitivitas terhadap data berlabel minoritas. Kelemahan yang masih ditemukan pada model Naive Bayes dan Gradient Boosting—terutama dalam menangani ambiguitas konteks linguistik—menunjukkan peluang pengembangan lebih lanjut dengan pendekatan berbasis representasi semantik yang lebih mendalam.

Untuk penelitian mendatang, disarankan penerapan teknik penyeimbangan kelas seperti SMOTE (Synthetic Minority Oversampling Technique) atau class weighting, serta eksplorasi embedding kontekstual seperti Word2Vec, FastText, atau IndoBERT, guna memperkaya pemahaman model terhadap konteks makna dan meningkatkan akurasi keseluruhan pada seluruh kelas sentimen.

TABEL VIII
HASIL PERBANDINGAN KINERJA MODEL

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Naive Bayes	87.98	82.30	87.98	84.22
Gradient Boosting	88.92	85.83	88.92	86.51
Proboboost Hybrid	89.51	85.31	89.51	87.31

Tabel VIII memperlihatkan perbandingan performa tiga model utama. Proboboost menempati posisi tertinggi dengan akurasi 89.51% dan F1-score 87.31%, diikuti oleh Gradient Boosting (88.92%) dan Naive Bayes (87.98%). Hasil ini memperkuat bahwa pendekatan hybrid mampu menyeimbangkan sensitivitas antar kelas serta mengurangi bias terhadap data mayoritas. Secara akademis, peningkatan ini menunjukkan relevansi Proboboost sebagai model adaptif yang efektif dalam analisis sentimen berbasis teks pendek

I. Analisis Kesalahan (Error Analysis)

Analisis kesalahan menunjukkan bahwa sebagian kesalahan klasifikasi terjadi pada ulasan dengan struktur ambigu atau kontekstual, seperti:

TABEL IX
ANALISIS KESALAHAN (ERROR ANALYSIS)

No	Kalimat Ulasan	Label Asli	Prediksi	Keterangan Kesalahan
1	ga lengkap profil selalu salah mulu format ktnya pas full name pdhl sesuai	Netral	Positif	Model menganggap kata <i>sesuai</i> memiliki konotasi positif meskipun konteks keseluruhan bersifat keluhan teknis.
2	login nomor hp selalu gagal notif jadi salah ulang beberapa lagi waktu tunggu otp terlalu cepat otp keluar waktu habis sdah hampir bulan cuma mau login hasil	Positif	Negatif	Model gagal menangkap makna “hasil” yang bernada puas, teralihkan oleh kata negatif seperti <i>gagal</i> dan <i>habis</i> .
3	sangat bantu	Netral	Positif	Kalimat terlalu pendek sehingga ambigu, model menganggap “bantu” menunjukkan apresiasi.

4	aplikasi oke cuma klo bs donasi kelipat jg jgn cm	Netral	Positif	Model lebih menekankan kata <i>oke</i> dan mengabaikan konteks saran (<i>jgn cm</i>) yang seharusnya bersifat netral.
5	kak tambah tanda bukti bayar struk whatsapp gmail bentuk dokumen bebas	Netral	Positif	Model menafsirkan kalimat permintaan fitur sebagai apresiasi positif karena tidak adanya kata negatif eksplisit.

Berdasarkan hasil analisis kesalahan yang disajikan pada Tabel IX, dapat disimpulkan bahwa sebagian besar kesalahan klasifikasi berasal dari ambiguitas linguistik dan keterbatasan konteks pada kalimat ulasan. Hal ini menunjukkan bahwa peningkatan kinerja model ke depan dapat dilakukan melalui penerapan context-aware model seperti BERT atau embedding layer berbasis Word2Vec, yang mampu memahami relasi semantik kata dalam konteks kalimat. Selain itu, pendekatan data augmentation untuk memperkaya variasi teks pendek juga dapat menjadi solusi dalam mengurangi tingkat kesalahan klasifikasi pada kategori sentimen minoritas.

IV. KESIMPULAN

Penelitian ini memberikan perspektif baru bahwa model hibrida Proboboost, hasil integrasi Multinomial Naïve Bayes dan Gradient Boosting, mampu meningkatkan akurasi klasifikasi sentimen hingga 89.51%, melampaui model Naïve Bayes murni yang hanya mencapai 87.98%. Pendekatan ini terbukti efektif dalam mengoptimalkan kinerja klasifikasi, khususnya dalam menangani ketidakseimbangan antar kelas.

Dari 11775 ulasan pengguna aplikasi Kitabisa, mayoritas bernuansa positif (10274 ulasan), diikuti oleh ulasan negatif (1002) dan netral (499), yang mencerminkan tingginya tingkat kepuasan pengguna. Namun, distribusi data yang timpang berdampak pada rendahnya performa model terhadap kelas minoritas, terutama kategori netral yang menghasilkan F1-score nol pada model dasar.

Dengan demikian, pendekatan hibrida ini memberikan hasil klasifikasi yang lebih akurat dan stabil, serta membuka peluang pengembangan strategi lanjutan seperti penyeimbangan data dan algoritma adaptif untuk meningkatkan representasi seluruh kategori sentimen.

PERNYATAAN KONFLIK KEPENTINGAN

Penulis menyatakan tidak memiliki konflik kepentingan dalam penelitian, penulisan, maupun publikasi artikel ini.

Seluruh hasil dan analisis yang disajikan dilakukan secara objektif tanpa intervensi dari pihak manapun.

DAFTAR PUSTAKA

- [1] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges,” *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [2] O. Kononova, T. He, H. Huo, A. Trevartha, E. A. Olivetti, and G. Ceder, “Opportunities and challenges of text mining in materials research,” *iScience*, vol. 24, no. 3, p. 102155, Mar. 2021, doi: 10.1016/j.isci.2021.102155.
- [3] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” *Knowl.-Based Syst.*, vol. 226, p. 107134, 2021.
- [4] Israt Jahan, Md Nakibul Islam, Md Mahadi Hasan, and Md Rafiuddin Siddiky, “Comparative analysis of machine learning algorithms for sentiment classification in social media text,” *World J. Adv. Res. Rev.*, vol. 23, no. 3, pp. 2842–2852, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2983.
- [5] L. Zhang, S. Wang, and B. Liu, “Deep Learning for Sentiment Analysis: A Survey,” Jan. 30, 2018, *arXiv: arXiv:1801.07883*. doi: 10.48550/arXiv.1801.07883.
- [6] C. Kaur, “Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 6303–6311, Aug. 2020, doi: 10.30534/ijatcse/2020/310942020.
- [7] S. Rabbani, D. Safitri, F. T. P. Siregar, R. Rahmaddeni, and L. Efriponi, “Evaluation of Support Vector Machine, Naïve Bayes, Decision Tree, and Gradient Boosting Algorithms for Sentiment Analysis on ChatGPT Twitter Dataset,” *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, pp. 11–21, Nov. 2023, doi: 10.24014/ijaidm.v7i1.24662.
- [8] E. Cambria, D. Hazarika, S. Poria, A. Hussain, and R. B. V. Subramanyam, “Benchmarking Multimodal Sentiment Analysis,” Jul. 29, 2017, *arXiv: arXiv:1707.09538*. doi: 10.48550/arXiv.1707.09538.
- [9] M. Kamruzzaman, M. Hossain, Md. R. I. Imran, and S. C. Bakchy, “A Comparative Analysis of Sentiment Classification Based on Deep and Traditional Ensemble Machine Learning Models,” in *2021 International Conference on Science & Contemporary Technologies (ICSCST)*, Aug. 2021, pp. 1–5. doi: 10.1109/ICSCST53883.2021.9642583.
- [10] M. Nalluri, M. Pentela, and N. R. Eluri, “A Scalable Tree Boosting System: XG Boost”, Accessed: Sep. 10, 2025. [Online]. Available: <http://ijrsset.org/pdfs/v7-i12/5.pdf>
- [11] R. Zulfiqri, B. N. Sari, and T. N. Padilah, “Analisis sentimen ulasan pengguna aplikasi media sosial Instagram pada situs Google Play Store menggunakan Naïve Bayes Classifier,” *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, Accessed: Sep. 10, 2025. [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/4995>
- [12] E. Apriliyanto and Y. S. Rahayu, “Comparison of Sentiment Analysis from Twitter Data Collection with Naïve Bayes, Decision Tree, and k-Nearest Neighbor Methods,” *J. Ilm. SINUS*, vol. 22, no. 2, pp. 1–12, 2024.
- [13] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. M. Nasir, N. Shaadan, and S. W. Indratno, “Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data,” *Indones J Elec Eng Comput Sci*, vol. 29, no. 1, p. 598, 2023.
- [14] S. Cui, Y. Han, Y. Duan, Y. Li, S. Zhu, and C. Song, “A Two-Stage Voting-Boosting Technique for Ensemble Learning in Social Network Sentiment Classification,” *Entropy*, vol. 25, no. 4, p. 555, Apr. 2023, doi: 10.3390/e25040555.

- [15] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.
- [16] N. A. K. M. Haris, S. Mutalib, A. M. A. Malik, S. Abdul-Rahman, and S. N. K. Kamarudin, "Sentiment classification from reviews for tourism analytics," *Int. J. Adv. Intell. Inform.*, vol. 9, no. 1, p. 108, Mar. 2023, doi: 10.26555/ijain.v9i1.1077.
- [17] W. I. Al-Obaydy, H. A. Hashim, Y. A. Najm, and A. A. Jalal, "Document classification using term frequency-inverse document frequency and K-means clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, pp. 1517–1524, 2022.
- [18] J. Eykens, R. Guns, and T. C. E. Engels, "Fine-grained classification of social science journal articles using textual data: A comparison of supervised machine learning approaches," *Quant. Sci. Stud.*, vol. 2, no. 1, pp. 89–110, Apr. 2021, doi: 10.1162/qss_a_00106.
- [19] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Eng. Rep.*, vol. 5, no. 4, p. e12599, 2023, doi: 10.1002/eng2.12599.
- [20] Amriana, A. A. Ilham, A. Achmad, and Yusran, "Ensemble Soft-Voting Model for Classification Optimization of Medicinal Plants Leaves," in *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Nov. 2023, pp. 147–152, doi: 10.1109/COMNETSAT59769.2023.10420635.
- [21] S. Riyanto, I. S. Sitanggang, T. Djatna, and T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406116.
- [22] "Kitabisa - Donasi, Zakat, Wakaf, dan Saling Jaga se-Indonesia." Accessed: Sep. 11, 2025. [Online]. Available: <https://kitabisa.com/>
- [23] R. O. Olanrewaju, S. A. Olanrewaju, and L. A. Nafiu, "Multinomial naïve bayes classifier: Bayesian versus nonparametric classifier approach," *Eur. J. Stat.*, vol. 2, pp. 8–8, 2022.
- [24] M. Fahmy Amin, "Confusion matrix in three-class classification problems: A step-by-step tutorial," *J. Eng. Res.*, vol. 7, no. 1, pp. 0–0, 2023.
- [25] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," in *Knowledge Discovery in Databases: PKDD 2003*, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, pp. 107–119. doi: 10.1007/978-3-540-39804-2_12.
- [26] R. Wijayanti and A. Arisal, "Ensemble approach for sentiment polarity analysis in user-generated Indonesian text," in *2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Jakarta: IEEE, Oct. 2017, pp. 158–163. doi: 10.1109/IC3INA.2017.8251759.