Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Pengguna Dompet Digital pada Google Play Store

M Adhe Akbar*, Fenty Ariany

Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1,*]m_adhe_akbar@teknokrat.ac.id, [2]fentyariany@teknokrat.ac.id
Email Penulis Korespondensi: m_adhe_akbar@teknokrat.ac.id
Submitted: **24/01/2026**; Accepted: **05/03/2026**; Published**: 06/03/2026**

Abstrak−Pertumbuhan pesat pengguna dompet digital di Indonesia yang mencapai jutaan pengguna aktif menghasilkan volume ulasan yang masif di Google Play Store. Data tekstual ini mengandung wawasan krusial terkait kepuasan pelanggan, namun seringkali belum dimanfaatkan secara optimal karena tantangan dalam pengolahan data tidak terstruktur. Penelitian ini bertujuan melakukan analisis komparasi performa antara algoritma probabilistik Naive Bayes dan algoritma berbasis jarak K-Nearest Neighbor (KNN) dalam mengklasifikasikan sentimen pengguna aplikasi DANA, OVO, DOKU, dan LinkAja. Studi ini menggunakan dataset sebanyak 18.869 ulasan yang menunjukkan ketidakseimbangan kelas ringan (mild imbalance) dengan dominasi sentimen negatif sebesar 57,54%. Untuk menjaga representasi data asli yang besar, penelitian ini menerapkan teknik Stratified Sampling tanpa penyeimbangan data sintetis (seperti SMOTE), yang dilanjutkan dengan tahapan preprocessing komprehensif berbantuan pustaka Sastrawi dan ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF). Optimasi model dilakukan secara sistematis menggunakan GridSearchCV untuk Naive Bayes dan Elbow Method untuk penentuan nilai k optimal pada KNN. Hasil pengujian empiris menunjukkan bahwa algoritma Naive Bayes dengan parameter smoothing alpha 0,1 mencapai performa terbaik dengan akurasi 88,5% dan AUC 0,9237, mengungguli KNN pada k=27 yang memperoleh akurasi 87,4%. Validitas perbedaan kinerja ini dikonfirmasi signifikan melalui uji statistik McNemar dengan p-value 0,0045. Temuan krusial lainnya adalah efisiensi komputasi, di mana Naive Bayes terbukti 129 kali lebih cepat dalam proses prediksi dibandingkan KNN. Berdasarkan keunggulan akurasi dan efisiensi waktu yang signifikan, Naive Bayes direkomendasikan sebagai metode yang lebih superior untuk analisis sentimen real-time pada ekosistem teknologi finansial.

Kata Kunci: Analisis Sentimen; Dompet Digital; Naive Bayes; K-Nearest Neighbor; Uji McNemar

Abstract−The rapid growth of digital wallet users in Indonesia, reaching millions of active users, has generated a massive volume of reviews on the Google Play Store. This textual data contains crucial insights regarding customer satisfaction but is often underutilized due to challenges in processing unstructured data. This study aims to perform a comparative performance analysis between the probabilistic Naive Bayes algorithm and the distance-based K-Nearest Neighbor (KNN) in classifying user sentiment for DANA, OVO, DOKU, and LinkAja applications. This study utilizes a dataset of 18,869 reviews which exhibits a mild class imbalance with a negative sentiment dominance of 57.54%. To preserve the representation of the large original data, this research applies Stratified Sampling without synthetic data balancing techniques (such as SMOTE), followed by comprehensive preprocessing stages aided by the Sastrawi library and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction. Model optimization was systematically conducted using GridSearchCV for Naive Bayes and the Elbow Method to determine the optimal k value for KNN. Empirical test results show that the Naive Bayes algorithm with a smoothing parameter alpha of 0.1 achieved the best performance with an accuracy of 88.5% and an AUC of 0.9237, outperforming KNN at k=27 which obtained an accuracy of 87.4%. The validity of this performance difference was confirmed to be significant through the McNemar statistical test with a p-value of 0.0045. Another crucial finding is computational efficiency, where Naive Bayes proved to be 129 times faster in the prediction process compared to KNN. Based on the significant advantages in accuracy and time efficiency, Naive Bayes is recommended as the superior method for real-time sentiment analysis in the financial technology ecosystem.

Keywords: Sentiment Analysis; Digital Wallet; Naive Bayes; K-Nearest Neighbor; McNemar Test

1. PENDAHULUAN

Transformasi digital di Indonesia kini bukan sekadar wacana, melainkan telah mengubah fundamental cara masyarakat bertransaksi. Uang tunai yang dulunya mendominasi kantong masyarakat, perlahan namun pasti mulai tergeser oleh kehadiran dompet digital (e-wallet) yang menawarkan kecepatan dan efisiensi. Berdasarkan data Bank Indonesia, nilai transaksi uang elektronik di Indonesia mencapai sekitar Rp 1.600 triliun di bulan Januari–Agustus tahun 2024, meningkat 35,76% dibandingkan tahun sebelumnya [1]. Pertumbuhan ini didorong oleh penggunaan smartphone yang tinggi di Indonesia, dimana 78% penduduk Indonesia telah memiliki akses terhadap perangkat mobile pintar yang memungkinkan penggunaan berbagai aplikasi pembayaran digital [2]. Lonjakan ini menjadi bukti tak terbantahkan bahwa kepercayaan publik terhadap ekosistem pembayaran digital telah menebal dan menjadi elemen vital ekonomi nasional.

Wajah pembayaran di Indonesia kini mengalami pergeseran. Dompet digital (e-wallet), yang dulunya hanya dianggap sebagai opsi cadangan, kini telah berubah menjadi instrumen vital dalam ekonomi harian masyarakat di indonesia. Melalui platform raksasa seperti DANA, OVO, DOKU, dan LinkAja, batasan transaksi konvensional berhasil didobrak, mulai dari pembayaran gerai ritel, transfer dana instan, hingga pelunasan tagihan utilitas dan pembelian pulsa, semuanya kini serba digital. Fenomena ini bukan sekadar klaim tanpa dasar. Data menunjukkan bahwa jumlah pengguna dompet digital di Indonesia mencapai sekitar 63,6 juta pengguna dan diprediksi meningkat pesat hingga lebih dari 200 juta pengguna pada 2025 [3]. Ledakan jumlah pengguna ini lantas memicu rivalitas sengit

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

di antara penyedia layanan, yang kini berlomba-lomba menyajikan fitur terlengkap dan reliabilitas sistem terbaik demi memenangkan hati konsumen.

Menariknya, peningkatan jumlah pengguna dompet digital tidak selalu berarti mereka puas. Dalam kehidupan nyata, pengguna masih sering kebingungan dengan hal-hal seperti transaksi yang sering gagal, pelanggaran keamanan akun, respons layanan lambat, dan aplikasi yang sering crash saat digunakan. Keluhan-keluhan jujur ini bertebaran di kolom ulasan platform aplikasi seperti Google Play Store. Kita dapat menemukan keluhan-keluhan jujur ini di bagian ulasan toko aplikasi macam Google Play Store. Penyedia layanan tidak lagi bisa mengabaikan ulasan-ulasan ini, mereka harus membacanya untuk meningkatkan kualitas layanan dan mencegah pelanggan beralih ke pesaing. Analisis sentimen berbasis machine learning menawarkan solusi otomatis, skalabel untuk mengekstrak insight yang berharga dari volume data ulasan yang sangat besar dan terus bertambah setiap harinya [4].

Analisis sentimen, yang juga dikenal sebagai Opinion Mining, adalah metode Pemrosesan Bahasa Alami (NLP) yang secara otomatis menemukan dan menganalisis emosi, sikap, atau opini (positif, negatif, atau netral) dalam data teks seperti ulasan produk, posting media sosial, atau umpan balik pelanggan. Ini membantu kita memahami apa yang dipikirkan orang tentang suatu topik, produk, atau layanan social [5]. Analisis sentimen untuk ulasan produk dan layanan digital berusaha mengelompokkan pendapat pengguna berdasarkan isi ulasan, seperti positif, negatif, atau netral. Kemampuan teknik Natural Language Processing untuk dengan cepat menganalisis ribuan atau bahkan jutaan ulasan. Machine Learning telah digunakan secara sukses untuk mengklasifikasikan sentimen di berbagai bidang, seperti Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, dan arsitektur deep learning seperti CNN, RNN, LSTM, serta model hibrida yang menggabungkan berbagai arsitektur [6].

Eksplorasi terhadap metode analisis sentimen pada ekosistem aplikasi seluler di Indonesia telah melahirkan ragam pendekatan teknis dengan capaian yang dinamis. Dalam konteks e-commerce, Wibowo et al. [7] menggunakan algoritma Naive Bayes yang dikombinasikan dengan feature selection menggunakan metode Chi-Square untuk analisis sentimen ulasan aplikasi e-commerce di Indonesia dan berhasil mendapat akurasi 86,2%. Penelitian serupa yang dilakukan oleh Rahmawati dan Kusuma pada tahun 2023 membandingkan performa algoritma Naive Bayes dan Support Vector Machine (SVM) pada data transportasi daring; hasilnya mengungkap bahwa meskipun SVM unggul tipis dalam ketepatan prediksi, Naive Bayes justru jauh lebih superior dari sisi efisiensi komputasi, menjadikannya opsi paling rasional untuk sistem real-time [8]. Pada penelitian Santoso tahun 2022 sektor perhotelan menyoroti perilaku algoritma KNN yang sangat sensitif, menegaskan bahwa penentuan parameter k yang presisi adalah kunci mutlak untuk mendongkrak performa model [9].

Penelitian oleh Fattahila pada tahun 2021 menggunakan CNN–LSTM pada review DANA, OVO, LinkAja, Sakuku (Google Play); akurasi validasi 83% dan menunjukkan kecenderungan sentimen negatif pada kategori transaksi, akses, layanan, akun, dan performa [10]. Pendekatan Deep Learning dengan arsitektur LSTM mampu menghasilkan akurasi prediksi yang tinggi, namun pengorbanan dari sisi waktu komputasi dan kebutuhan sumber daya cukup besar, sebagaimana umum terjadi pada model deep learning berbasis LSTM [11]. Penelitian ini membandingkan sebagian algoritma klasifikasi klasik berikut: Naive Bayes, KNN, dan Decision Tree pada dataset ulasan aplikasi DANA dan menemukan bahwa Naive Bayes menciptakan penyeimbang terbaik antara akurasi prediksi serta kemampuan efisiensi komputasi, sejalan dengan temuan berbagai studi yang menunjukkan bahwa Naive Bayes sering memberikan kombinasi akurasi yang baik dengan waktu komputasi dan penggunaan memori yang sangat efisien dibandingkan algoritma lain seperti KNN dan Decision Tree [12].

Meski cukup banyak penelitian mengenai analisis sentimen pada ulasan aplikasi mobile di Indonesia, terdapat beberapa gap penelitian yang penting untuk diisi penelitian ini. Pertama, pada umumnya penelitian yang ada hanya fokus pada satu aplikasi dompet digital saja, sehingga kemampuan generalisasi hasil penelitian menjadi terbatas serta sulit diterapkan pada kondisi yang lebih besar. Kedua, perbandingan menyeluruh antara algoritma Naive Bayes dan K-Nearest Neighbor dengan optimasi parameter yang sistematis dan metodis masih sangat jarang dilakukan, sementara itu kedua algoritma ini memiliki karakter yang berbeda dan cocok untuk skenario penggunaan yang berbeda pula. Ketiga, analisis statistik untuk mengkonfirmasi signifikansi perbedaan performa antar algoritma menggunakan uji statistik formal seperti McNemar test belum banyak diaplikasikan dalam penelitian-penelitian serupa di Indonesia.

Bersumber gap penelitian tersebut, penelitian ini memiliki 4 tujuan penting untuk dicapai: (1) menerapkan serta membandingkan performa algoritma Naive Bayes dan K-Nearest Neighbor dalam mengklasifikasi sentimen ulasan pengguna dompet digital dengan dataset yang representatif; (2) melakukan optimasi parameter dengan cara sistematis menggunakan teknik GridSearchCV dan cross-validation untuk Naive Bayes dan Elbow Method untuk KNN; (3) menganalisa perbandingan antara kedua performa algoritma secara statistik menggunakan uji McNemar untuk membuktikan signifikansi temuan serta (4) memberikan saran algoritma yang paling cocok untuk aplikasi analisis sentimen real-time pada domain fintech. Dataset yang dipakai dalam penelitian ini mencakup ulasan dari empat aplikasi dompet digital paling populer di Indonesia, yaitu DANA, OVO, DOKU, dan LinkAja, sehingga hasil penelitian diharapkan dapat digeneralisasi secara lebih luas untuk dompet digital di Indonesia.

Kontribusi penting serta orisinalitas penelitian ini terdapat pada penyediaan analogi empiris yang menyeluruh serta rigorus antara algoritma Naive Bayes dan KNN untuk analisis sentimen dompet digital dengan metodologi yang sistematis, mencakup optimasi parameter yang terstruktur, validasi silang untuk memastikan reliabilitas hasil, dan pengujian signifikansi statistik untuk membuktikan keunggulan satu algoritma atas yang lain. Hasil penelitian ini diharapkan dapat menjadi referensi ilmiah yang berharga bagi peneliti akademis dan praktisi industri dalam memilih-

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

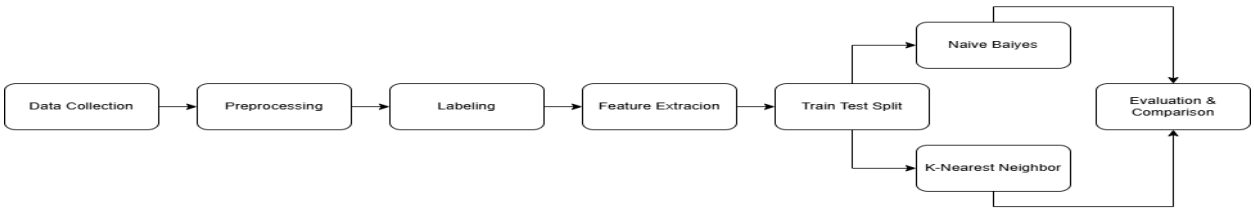
milih algoritma yang paling tepat untuk mengembangkan sistem analisis sentimen pada domain financial technology di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini memakai pendekatan kuantitatif dengan metode eksperimental komparatif untuk membandingkan performa dua algoritma klasifikasi machine learning dalam tugas analisis sentimen. Desain penelitian dirancang secara sistematis untuk memastikan validitas dan reliabilitas hasil yang diperoleh. Konsep penelitian terdiri dari sebagian fase utama yang saling berkaitan: pengumpulan data melalui web scraping, preprocessing data teks untuk mensterilkan serta menstandardisasi data, ekstraksi fitur menggunakan metode TF-IDF, pembagian data menjadi training dan testing set, pelatihan model dengan optimasi parameter, evaluasi performa menggunakan berbagai metrik, dan analisis statistik untuk menguji signifikansi perbedaan.

Pengumpulan data dicoba menggunakan teknik web scraping dengan memanfaatkan library google-play-scraper yang berjalan pada bahasa pemrograman Python versi 3.9. Library ini memungkinkan ekstraksi data ulasan dengan cara otomatis dari Google Play Store dengan berbagai parameter yang dapat dikonfigurasi [13]. Data yang dididapatkan berupa ulasan pengguna lengkap dengan metadata seperti rating bintang, tanggal posting, dan identifikasi pengguna dari empat aplikasi dompet digital utama di Indonesia yaitu DANA, OVO, DOKU, dan LinkAja. Pengumpulan data diambil pada rentang waktu Januari 2025 dengan batas 5.000 ulasan per aplikasi untuk memperoleh ilustrasi yang representatif. Keseluruhan data yang berhasil dikumpulkan sebanyak 23.412 ulasan. Setelah proses filtering untuk menghilangkan ulasan dengan rating 3 yang dianggap netral dan tidak memberikan sinyal perbaikan layanan yang jelas maupun polaritas sentimen yang tegas bagi penyedia dompet digital, total ulasan final yang digunakan sebanyak 18.869 ulasan. Terkait proporsi data, dataset menunjukkan adanya ketidakseimbangan kelas ringan (mild imbalance) dengan komposisi 57,54% sentimen negatif (10.857 ulasan) dan 42,46% sentimen positif (8.012 ulasan). Penelitian ini tidak menerapkan teknik oversampling sintetis seperti SMOTE karena volume data yang besar (18.869 ulasan) dinilai cukup representatif untuk melatih model mengenali pola kedua kelas tanpa memicu bias noise buatan. Sebagai gantinya, mitigasi ketimpangan dilakukan menggunakan teknik Stratified Sampling pada tahap pembagian data (train-test split) untuk memastikan rasio kelas pada data uji konsisten dengan populasi aslinya. Penggunaan metrik AUC-ROC dipilih sebagai indikator performa utama karena sifatnya yang invarian terhadap ketidakseimbangan kelas dibandingkan akurasi murni.



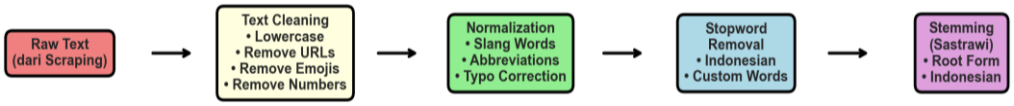
Gambar 1. Alur Metodologi Penelitian

Berdasarkan Gambar 1, penelitian ini diawali dengan tahap pengumpulan data, di mana data berupa ulasan pengguna dikumpulkan dari Google Play Store. Aplikasi yang menjadi objek penelitian meliputi DANA, OVO, GoPay, dan LinkAja, dengan target sebanyak 5.000 ulasan untuk setiap aplikasi. Data ulasan yang telah dikumpulkan kemudian diproses pada tahap preprocessing untuk meningkatkan kualitas data teks. Tahap ini mencakup proses pembersihan teks (*text cleaning*) untuk menghilangkan karakter yang tidak relevan, normalisasi teks untuk menyeragamkan penulisan kata, penghapusan *stopword* guna menghilangkan kata-kata umum yang tidak memiliki makna signifikan, serta proses *stemming* menggunakan algoritma Sastrawi untuk mengubah kata ke bentuk dasarnya. Setelah preprocessing selesai, data ulasan kemudian diberi label sentimen berdasarkan nilai rating yang diberikan pengguna. Ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan ulasan dengan rating 1 hingga 3 dikategorikan sebagai sentimen negatif, sementara ulasan dengan rating netral dikecualikan dari proses analisis. Data yang telah berlabel selanjutnya memasuki tahap ekstraksi fitur menggunakan metode TF-IDF, dengan pengaturan *n-gram* unigram, bigram, dan trigram (n = 1, 2, dan 3). Selain itu, dilakukan optimasi jumlah fitur untuk memperoleh representasi teks yang paling optimal. Tahap berikutnya adalah pembagian data menjadi data latih dan data uji dengan rasio 80% untuk pelatihan dan 20% untuk pengujian, menggunakan metode *stratified sampling* agar distribusi kelas tetap seimbang. Data latih kemudian digunakan untuk membangun dua model klasifikasi, yaitu Naive Bayes dan K-Nearest Neighbor (KNN). Pada model Naive Bayes dilakukan proses *alpha tuning* menggunakan GridSearchCV, sedangkan pada model KNN dilakukan optimasi nilai K dan pemilihan metrik jarak yang paling sesuai. Tahap akhir penelitian adalah evaluasi dan perbandingan kinerja kedua model klasifikasi. Evaluasi dilakukan menggunakan beberapa metrik, yaitu akurasi, precision, recall, F1-score, serta analisis ROC-AUC. Selain itu, dilakukan uji statistik McNemar untuk mengetahui perbedaan performa kedua model secara signifikan, serta analisis biaya komputasi guna menilai efisiensi masing-masing metode.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

2.2 Preprocessing Data

Tahap preprocessing data ialah tahap kritis yang bermaksud buat membersihkan, menstandardisasi, serta mentransformasi data bacaan anonim saat sebelum diproses lebih lanjut oleh algoritma machine learning, alhasil menciptakan representasi bacaan yang lebih teratur serta bermutu besar [14]. Mutu preprocessing sangat mempengaruhi performa model akhir. Dalam penelitian ini, daftar stopword juga diperluas untuk mencakup nama-nama aplikasi (dana, ovo, doku, linkaja) yang sering muncul namun tidak berkontribusi pada polaritas sentimen. Kelima, stemming dilakukan memakai library Sastrawi [15] yang secara khusus didesain untuk memproses teks berbahasa Indonesia, untuk mengubah perkata imbuhan menjadi bentuk kata dasarnya alhasil variasi morfologis tidak mempengaruhi representasi fitur. Dalam implementasi teknisnya, proses stemming ini dijalankan secara sekuensial menggunakan fungsi pandas.apply() tanpa menerapkan teknik optimasi tambahan seperti dictionary-based caching atau parallel processing. Eksekusi tahap ini membutuhkan waktu komputasi selama 30 menit untuk memproses 18.869 ulasan. Meskipun durasi tersebut masih dapat ditoleransi untuk penelitian ini karena sifatnya yang hanya dieksekusi offline preprocessing sebelum pelatihan model, kami menyadari bahwa kompleksitas waktu algoritma Sastrawi cukup tinggi pada data berskala besar. Oleh karena itu, bagi penelitian selanjutnya dengan volume dataset yang lebih masif, kami sangat merekomendasikan penerapan teknik optimasi dictionary-based caching guna meningkatkan efisiensi waktu pemrosesan secara signifikan.



Gambar 2. Pipeline Preprocessing Teks

Berdasarkan Gambar 2, alur preprocessing dalam penelitian ini terdiri dari empat tahapan utama yang dilakukan secara berurutan. Pertama, Text Cleaning dilakukan secara komprehensif yang mencakup pengubahan huruf menjadi kecil (lowercase), serta penghapusan elemen noise seperti URL, emoji, angka, dan karakter khusus yang tidak relevan. Kedua, Normalization diterapkan untuk menstandarisasi kata-kata tidak baku (slang words), singkatan, dan memperbaiki kesalahan penulisan (typo correction) agar sesuai dengan kaidah bahasa Indonesia, sehingga memaksimalkan pengenalan pola sentimen. Ketiga, Stopword Removal diaplikasikan untuk menghilangkan kata-kata umum (seperti kata penghubung dan preposisi) serta kata khusus (seperti nama aplikasi) yang tidak memiliki kontribusi signifikan terhadap polaritas sentimen. Keempat, Stemming dilakukan menggunakan pustaka Sastrawi untuk mentransformasi kata berimbuhan menjadi bentuk kata dasarnya (root form).

2.3 Ekstraksi Fitur TF-IDF

Ekstraksi fitur yakni proses transformasi data teks yang telah melewati preprocessing menjadi representasi numerik dapat diproses oleh algoritma machine learning. Ekstraksi fitur dicoba menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik pembobotan kata yang digunakan dalam information retrieval serta text mining [16]. Metode TF-IDF membagi bobot setiap kata berdasarkan 2 bagian: frekuensi kemunculan kata dalam dokumen individual (term frequency) serta kebalikan frekuensi dokumen yang mengandung kata tersebut dalam seluruh korpus (inverse document frequency). Gabungan kedua bagian ini memberikan bobot yang lebih tinggi untuk kata yang kerap timbul dalam dokumen tertentu tetapi jarang timbul di dokumen lain, mengindikasikan kata tersebut memiliki nilai diskriminatif yang tinggi untuk membedakan antar dokumen. Rumus matematis TF-IDF dinyatakan sebagai berikut:

$$TF - IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (1)$$

Dimana TF(t,d) merepresentasikan frekuensi term t dalam dokumen d, dan IDF(t,D) dihitung sebagai log(N/df(t)) dengan N adalah jumlah total dokumen dalam korpus dan df(t) adalah jumlah dokumen yang mengandung term t dalam korpus. Implementasi TF-IDF penelitian ini memakai TfidfVectorizer dari library scikit-learn untuk mentransformasi koleksi dokumen menjadi matriks fitur berbobot TF-IDF [17] dengan konfigurasi parameter telah dioptimasi: max_features=5000 untuk membatasi jumlah fitur penting, ngram_range=(1,2) untuk mendapatkan unigram dan bigram, min_df=2 untuk mengabaikan kata yang sangat jarang muncul, dan max_df=0.85 untuk mengabaikan kata yang sangat umum.

2.4 Algoritma Naive Bayes

Naive Bayes adalah salah satu metode paling sederhana dan klasik untuk klasifikasi teks dan sering dijadikan baseline karena kesederhanaan, efisiensi, dan interpretabilitasnya [18]. Meskipun asumsi independensi ini naif (naive), algoritma ini telah terbukti sangat efisien untuk berbagai tugas klasifikasi teks karena sifatnya yang robust terhadap fitur yang redundan dan efisien secara komputasi. Untuk pengelompokan teks dengan fitur berbasis frekuensi seperti TF-IDF, varian Multinomial Naive Bayes adalah opsi yang paling sesuai karena didesain untuk menangani data count atau frekuensi. Probabilitas posterior untuk kelas tertentu diberikan dokumen dihitung menggunakan formula:

$$P(c|d) \propto P(c) \times \prod P(wi|c) \quad (2)$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

P(c|d) merupakan probabilitas posterior kelas c diberikan dokumen d yang ingin diprediksi, P(c) merupakan probabilitas prior kelas yang diestimasi dari proporsi kelas dalam training data, dan P(wi|c) merupakan probabilitas likelihood kata wi muncul dalam dokumen yang termasuk kelas c. Kategori dengan probabilitas posterior tertinggi diseleksi sebagai prediksi. Untuk mengatasi masalah zero probability yang terjadi ketika terdapat kata dalam testing data yang tidak pernah muncul dalam training data untuk kelas tertentu, digunakan teknik Laplace smoothing dengan parameter alpha [19]. Parameter smoothing alpha dioptimasi menggunakan teknik GridSearchCV dengan 5-fold cross-validation pada rentang nilai kandidat [0.01, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0] untuk menemukan nilai optimal.

2.5 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) adalah jenis *lazy learning* atau pembelajaran berbasis contoh yang tidak menggunakan parameter apa pun. Model ini tidak terlibat dalam proses pembelajaran yang berbeda dari data pelatihan sebaliknya, ia menyimpan semua contoh pelatihan dan melakukan perhitungan selama prediksi [20]. KNN mengurutkan data baru (titik kueri) ke dalam ruang fitur berdasarkan apa yang dikatakan sebagian besar dari k tetangga terdekat. Kesamaan kosinus adalah pilihan yang lebih baik untuk data teks yang memiliki representasi TF-IDF yang jarang dan berdimensi tinggi dibandingkan dengan jarak Euclidean. Hal ini karena kesamaan kosinus membandingkan kesamaan berdasarkan sudut antara vektor, bukan magnitude [21]. Formula cosine similarity adalah:

$$cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} \tag{3}$$

A dan B adalah vektor representasi dua dokumen, (A · B) adalah dot product antar vektor, dan ||A||, ||B|| adalah magnitude (norm) masing-masing vektor. Nilai cosine similarity berkisar antara 0 (tidak ada kesamaan) hingga 1 (identik). Pemilihan nilai k yang optimal sangat mempengaruhi performa KNN: nilai k terlalu kecil menyebabkan model sensitif terhadap noise (overfitting), sedangkan nilai k terlalu besar menyebabkan model kehilangan kemampuan diskriminasi (underfitting). Dalam penelitian ini, nilai k optimal ditentukan menggunakan Elbow Method dengan mengevaluasi akurasi validasi pada rentang nilai k=1 hingga k=29 (hanya nilai ganjil untuk menghindari tie dalam voting) menggunakan 5-fold cross-validation.

2.6 Evaluasi dan Validasi

Kami menemukan cara untuk membagi data dan mengevaluasi model agar kami bisa mendapatkan perkiraan kinerja yang akurat dan adil. Kami menggunakan pembagian bertingkat untuk membagi kumpulan data menjadi 80% data pelatihan (15.095 ulasan) dan 20% data pengujian (3.774 ulasan). Kami memastikan bahwa rasio kelas sama di kedua subkumpulan sehingga kedua kelas sentimen terwakili secara merata [22]. Uji data disimpan terpisah dan tidak digunakan sama sekali selama pelatihan dan optimasi parameter untuk memastikan tidak ada data yang hilang. Evaluasi performa model dilakukan menggunakan multiple metrics untuk mendapatkan gambaran yang komprehensif: accuracy mengukur proporsi prediksi benar secara keseluruhan; precision mengukur ketepatan prediksi positif; recall mengukur kelengkapan deteksi kelas positif; F1-score memberikan harmonic mean dari precision dan recall; dan Area Under ROC Curve (AUC-ROC) mengukur kemampuan diskriminasi model secara independen dari threshold keputusan. Validasi tambahan dilakukan menggunakan teknik 10-fold cross-validation untuk mengestimasi stabilitas dan variance performa model. Terakhir, uji statistik McNemar digunakan untuk menguji signifikansi perbedaan performa antara Naive Bayes dan KNN pada tingkat signifikansi α = 0,05, untuk memberikan bukti statistik apakah satu algoritma secara konsisten lebih unggul dari yang lain [23].

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Data

Hasil distribusi data ulasan untuk setiap aplikasi dompet digital setelah proses pembersihan dan penyaringan disajikan pada Tabel 1. Analisis distribusi sentimen per aplikasi menunjukkan variasi yang signifikan antar layanan dompet digital.Ketika melihat distribusi sentimen untuk setiap aplikasi, jelas bahwa itu sangat berbeda dari satu aplikasi dompet digital ke aplikasi lainnya. OVO memiliki ulasan paling negatif, dengan 86,78% di antaranya (4.083 dari 4.705), yang menunjukkan bahwa pengguna sangat tidak senang dengan layanan tersebut. Peringkat rata-ratanya hanya 1,59 dari 5. LinkAja, di sisi lain, memiliki reputasi yang jauh lebih baik di kalangan pengguna. Ini memiliki persentase ulasan positif tertinggi (68,63%) dan peringkat rata-rata 3,76. DANA memiliki pangsa sentimen negatif sebesar 66,22% (2.997 ulasan) dan peringkat rata-rata 2,40. DOKU, di sisi lain, memiliki pembagian yang lebih merata, dengan 47,04% sentimen negatif (2.262 ulasan) dan peringkat rata-rata 3,15.

Tabel 1. Distribusi Data Ulasan per Aplikasi

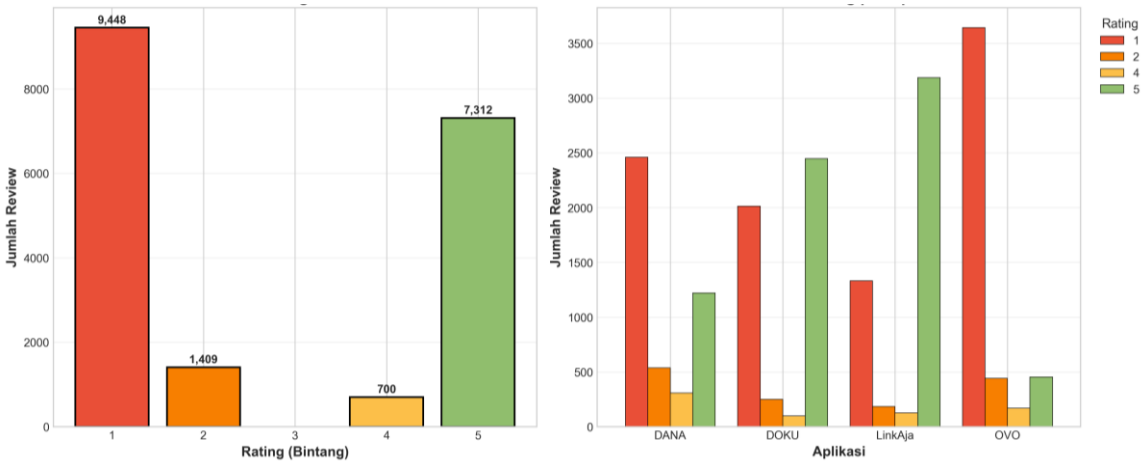
Aplikasi	Total	Positif	Negatif	% Positif	% Negatif	Rating
DANA	4.526	1.529	2.997	33,78%	66,22%	2,40
OVO	4.705	622	4.083	13,22%	86,78%	1,59
DOKU	4.809	2.547	2.262	52,96%	47,04%	3,15
LinkAja	4.829	3.314	1.515	68,63%	31,37%	3,76

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

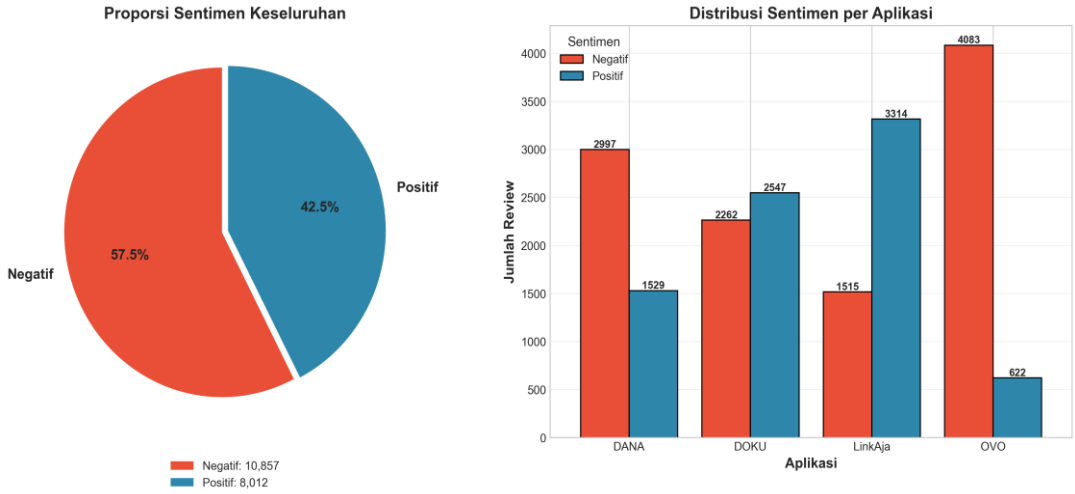
Aplikasi	Total	Positif	Negatif	% Positif	% Negatif	Rating
TOTAL	18.869	8.012	10.857	42,46%	57,54%	2,74

Berdasarkan Tabel 1, studi ini menggunakan kumpulan data yang terdiri dari 18.869 ulasan yang cukup mewakili keempat aplikasi. DANA memiliki 4.526 ulasan (24,0% dari total), OVO memiliki 4.705 ulasan (24,9%), DOKU memiliki 4.809 ulasan (25,5%), dan LinkAja memiliki ulasan terbanyak dengan 4.829 ulasan (25,6%). Distribusi yang relatif merata ini penting untuk menghindari bias hasil analisis oleh karakteristik aplikasi tertentu.

Distribusi rating keseluruhan dan per aplikasi divisualisasikan pada Gambar 3, sedangkan distribusi sentimen disajikan pada Gambar 4.



Gambar 3. Distribusi Rating Keseluruhan dan per Aplikasi



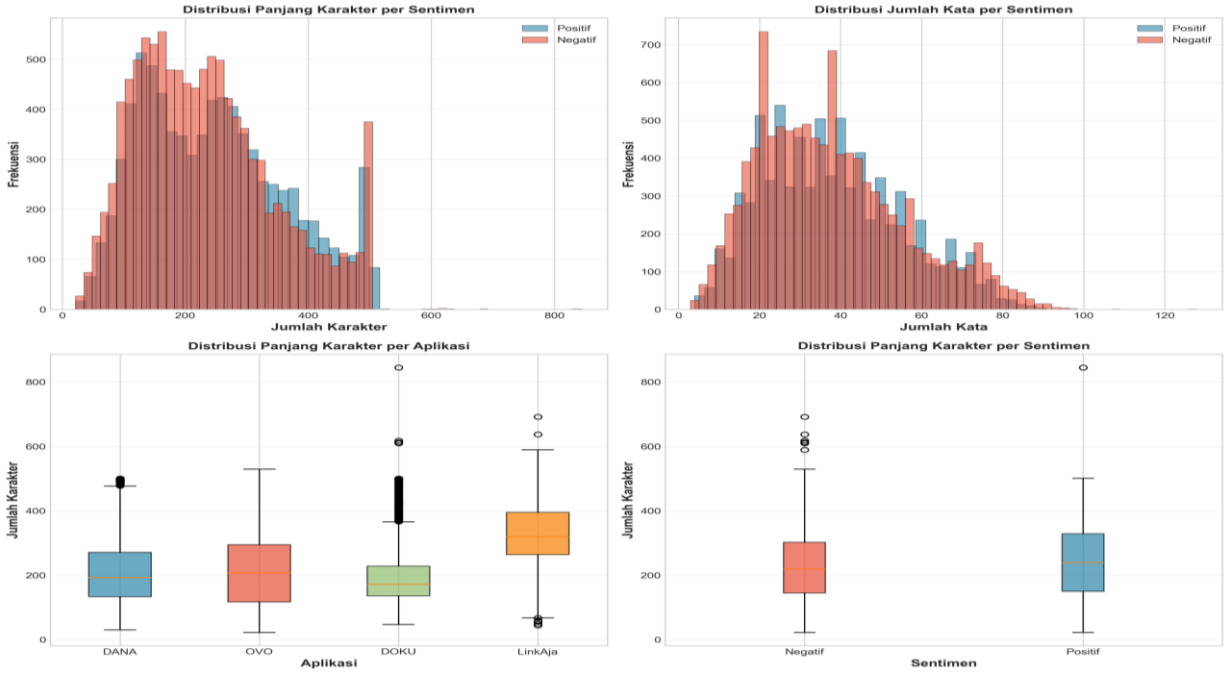
Gambar 4. Distribusi Sentimen Keseluruhan dan per Aplikasi

Berdasarkan Gambar 3, distribusi rating menunjukkan pola bimodal dengan 9.448 ulasan bintang satu dan 7.312 ulasan bintang lima, mengindikasikan polarisasi ekstrem dalam penilaian pengguna. Gambar 4 mengkonfirmasi bahwa sentimen negatif mendominasi dengan 57,5% (10.857 ulasan), sementara sentimen positif mencapai 42,5% (8.012 ulasan). OVO memiliki proporsi ulasan negatif tertinggi (86,78%), sedangkan LinkAja memiliki reputasi terbaik dengan 68,63% ulasan positif dan rating rata-rata 3,76.

3.2 Analisis Karakteristik Teks

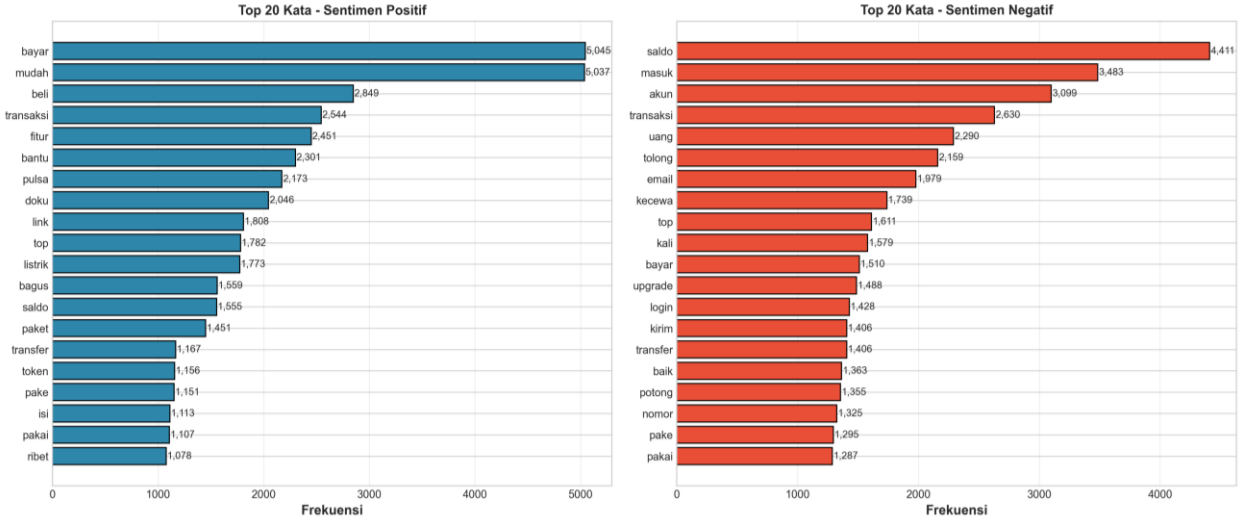
Karakteristik teks ulasan dianalisis untuk mengidentifikasi pola bahasa yang membedakan sentimen positif dan negatif. Distribusi panjang karakter dan jumlah kata per sentimen disajikan pada Gambar 5.Analisis frekuensi kata untuk setiap jenis sentimen disajikan pada Gambar 6, sedangkan visualisasi word cloud sentimen positif dan negatif ditampilkan pada Gambar 7 dan Gambar 8.Visualisasi word cloud pada Gambar 7 dan Gambar 8 memberikan gambaran intuitif mengenai kata-kata yang dominan dalam masing-masing kelas sentimen. Word cloud sentimen positif didominasi oleh kata-kata seperti "mudah", "bayar", "transaksi", "pulsa", "listrik", "bagus", dan "layan", mencerminkan kepuasan pengguna terhadap kemudahan transaksi dan kualitas layanan. Sementara itu, word cloud sentimen negatif menunjukkan dominasi kata-kata seperti "saldo", "masuk", "akun", "verifikasi", "gagal", "tolong", dan "kecewa", mengindikasikan masalah-masalah utama yang dihadapi pengguna terkait dengan akses akun, verifikasi identitas, dan kegagalan transaksi.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

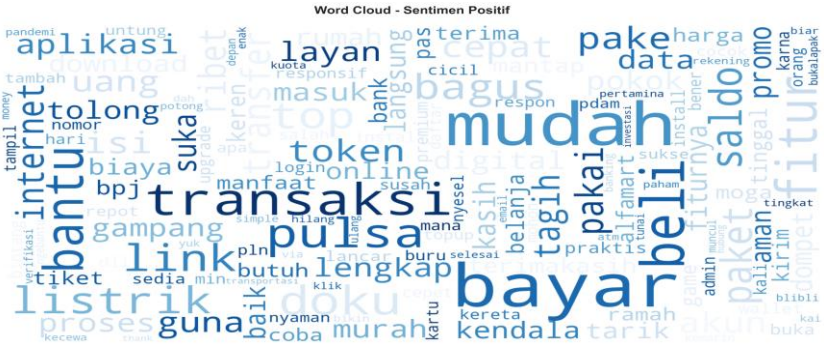


Gambar 5. Analisis Karakteristik Panjang Teks Review

Berdasarkan Gambar 5, ulasan negatif memiliki distribusi panjang yang lebih luas dengan rata-rata sekitar 200 karakter, sedangkan ulasan positif cenderung lebih pendek dengan rata-rata 180 karakter. Pola ini mengindikasikan bahwa pengguna yang tidak puas cenderung menulis ulasan lebih panjang untuk merinci keluhan mereka. Analisis per aplikasi menunjukkan LinkAja memiliki panjang karakter median terpanjang, sedangkan DOKU memiliki terpendek.



Gambar 6. Analisis Frekuensi Kata per Sentimen



Gambar 7. Word Cloud Sentimen Positif

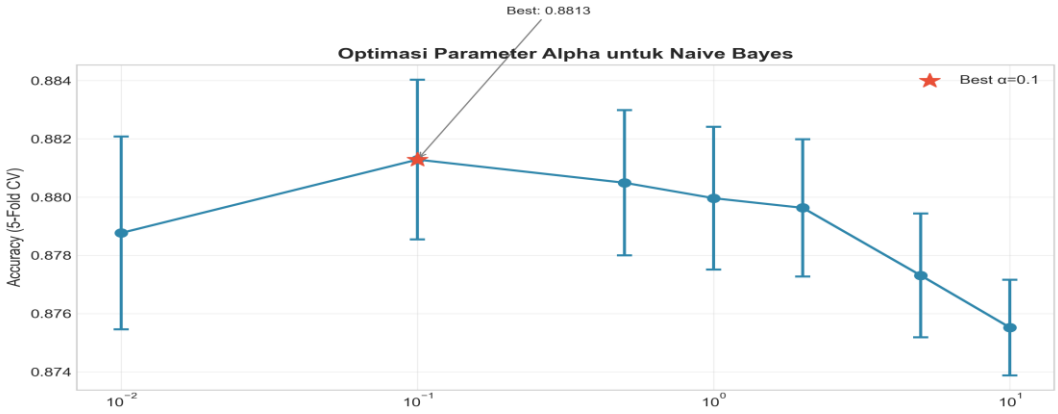
Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

Gambar 8. Word Cloud Sentimen Negatif

Berdasarkan Gambar 6 sampai Gambar 8, terlihat perbedaan kosakata yang signifikan antara sentimen positif dan negatif. Sentimen positif didominasi oleh kata "bayar" (5.045 kali), "mudah" (5.037), "beli" (2.849), "transaksi" (2.544), dan "fitur" (2.451), mencerminkan kepuasan pengguna terhadap kemudahan pembayaran dan kelengkapan fitur. Sebaliknya, sentimen negatif didominasi oleh kata "saldo" (4.411), "masuk" (3.483), "akun" (3.099), "transaksi" (2.630), dan "uang" (2.290), mengindikasikan masalah utama terkait akses akun, verifikasi identitas, dan kegagalan transaksi. Kehadiran kata "kecewa" (1.739) dalam daftar kata negatif mengkonfirmasi ekspresi ketidakpuasan pengguna secara eksplisit.

3.3 Hasil Optimasi Parameter Naive Bayes

Optimasi parameter alpha untuk algoritma Multinomial Naive Bayes dilakukan secara sistematis menggunakan teknik GridSearchCV dengan 5-fold cross-validation untuk memastikan hasil yang robust dan tidak overfitting terhadap satu split data tertentu. Parameter α pada algoritma Multinomial Naive Bayes berfungsi sebagai Laplace (additive) smoothing factor yang sangat penting untuk menangani masalah zero probability pada fitur kata yang tidak muncul di data latih, sehingga seluruh fitur tetap memiliki probabilitas non-nol dan model menjadi lebih stabil pada data teks berdimensi tinggi [24]. Nilai alpha yang optimal (0.1) memberikan keseimbangan yang baik antara model yang terlalu fit terhadap training data (alpha terlalu kecil) dan model yang terlalu general sehingga kehilangan informasi discriminative (alpha terlalu besar). Temuan ini konsisten dengan rekomendasi dari penelitian terdahulu yang merekomendasikan nilai alpha dalam rentang 0.1 hingga 1.0 untuk tugas klasifikasi teks.



Gambar 9. Optimasi Parameter Alpha untuk Naive Bayes

Berdasarkan Gambar 9 di atas hasil evaluasi performa untuk berbagai nilai alpha yang diuji dalam skala logaritmik. Hasil eksperimen 5-fold cross-validation menunjukkan bahwa nilai alpha=0.1 memberikan rata-rata akurasi validasi (mean validation accuracy) tertinggi sebesar 88,13% (Best: 0.8813). Angka ini mencerminkan kinerja model selama fase pelatihan untuk pemilihan parameter terbaik. Nilai alpha yang lebih kecil (0.01) menghasilkan akurasi validasi sekitar 87,85%, menunjukkan indikasi overfitting. Sebaliknya, nilai alpha yang lebih besar (>1.0) menyebabkan penurunan performa secara gradual hingga mencapai sekitar 87,55% pada alpha=10.0.

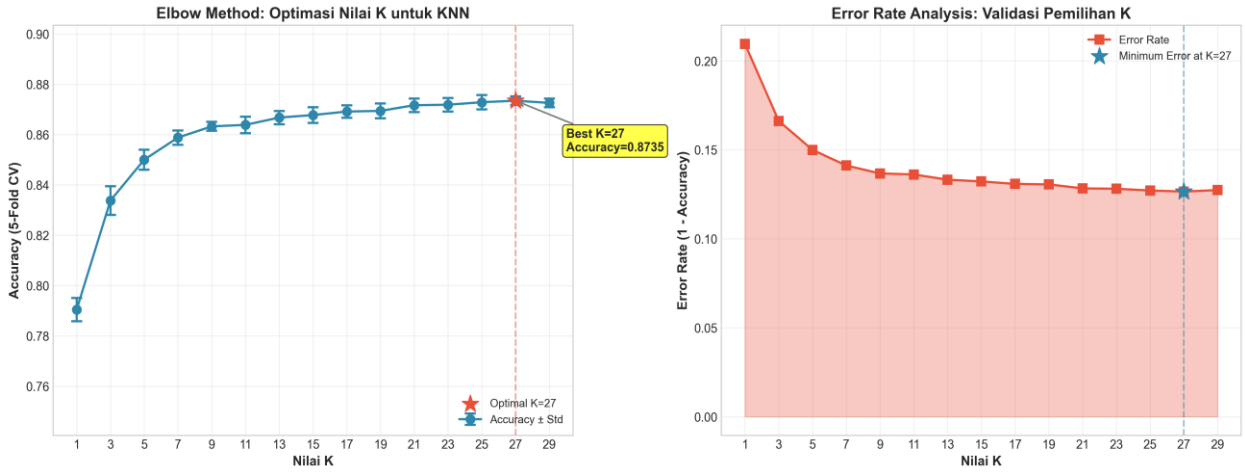
3.4 Hasil Optimasi Parameter KNN

Optimasi nilai k untuk algoritma K-Nearest Neighbor dilakukan menggunakan pendekatan Elbow Method dengan evaluasi komprehensif pada nilai k=1 hingga k=29 (hanya nilai ganjil).

Berdasarkan analisis Elbow Method, nilai k=27 dipilih sebagai parameter optimal karena memberikan akurasi tertinggi dengan standar deviasi yang sangat rendah (±0,0018), mengindikasikan stabilitas performa yang excellent antar fold validation. Pemilihan nilai k yang relatif besar ini dapat dijelaskan oleh beberapa faktor: dataset yang besar (18.869 sampel) memungkinkan penggunaan neighborhood yang lebih besar tanpa risiko kehilangan informasi lokal,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

representasi TF-IDF yang sparse dan high-dimensional (5.000 fitur) membuat perbedaan jarak antar dokumen relatif kecil, dan adanya noise dalam data ulasan lebih baik ditangani dengan voting dari banyak neighbor.



Gambar 10. Elbow Method: Optimasi Nilai K untuk KNN

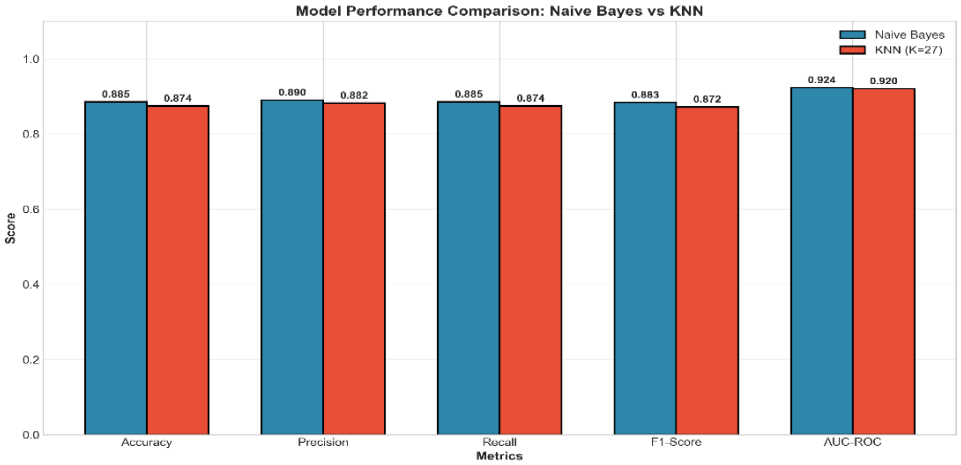
Berdasarkan Gambar 10 menunjukkan grafik hubungan antara nilai k dengan akurasi validasi beserta interval standar deviasi. Hasil eksperimen menunjukkan pola peningkatan akurasi yang signifikan dari k=1 (sekitar 79%) hingga k=7 (sekitar 86%), kemudian peningkatan yang lebih landai hingga mencapai puncak pada k=27 dengan akurasi 87,35% dan standar deviasi ±0,18%. Error rate analysis pada panel kanan menunjukkan penurunan error rate dari sekitar 21% pada k=1 menjadi sekitar 13% pada k=27.

3.5 Perbandingan Performa Model

Tabel 2 menyajikan perbandingan komprehensif performa kedua algoritma setelah optimasi parameter pada data testing yang terdiri dari 3.774 sampel.

Tabel 2. Perbandingan Performa Naive Bayes dan KNN

Metrik	Naive Bayes (α=0,1)	KNN (k=27)	Selisih
Accuracy	88,5%	87,4%	+1,1%
Precision	89,0%	88,2%	+0,8%
Recall	88,5%	87,4%	+1,1%
F1-Score	88,3%	87,2%	+1,1%
AUC-ROC	0,9237	0,9203	+0,0034

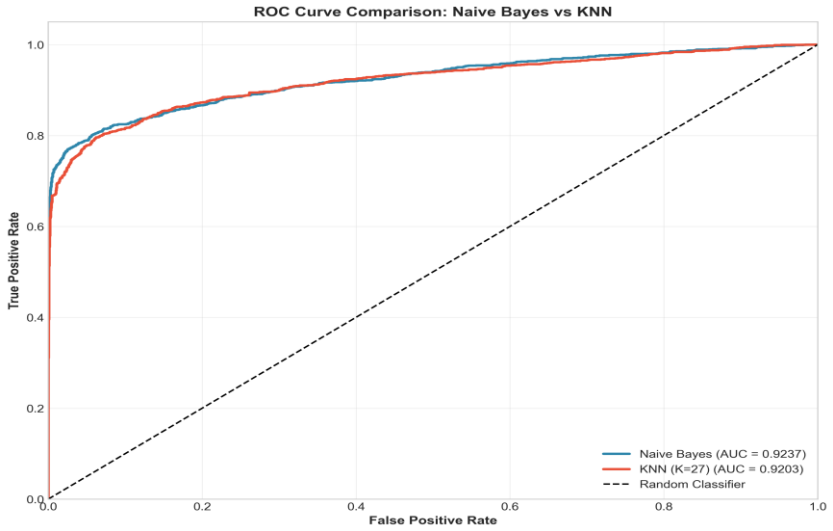


Gambar 11. Perbandingan Performa Model: Naive Bayes vs KNN

Berdasarkan Tabel 2 dan Visualisasi perbandingan performa antara kedua model secara grafis ditunjukkan pada Gambar 11, Evaluasi ini menggunakan metrik akurasi pengujian (testing accuracy) pada data yang belum pernah dilihat model sebelumnya (unseen data). Hasil pengujian menunjukkan bahwa Naive Bayes dengan parameter optimal alpha=0.1 mencapai akurasi sebesar 88,5%. Nilai ini sedikit lebih tinggi dibandingkan rata-rata akurasi validasi (88,13%), yang mengindikasikan bahwa model memiliki kemampuan generalisasi yang sangat baik dan tidak mengalami overfitting. Hasil evaluasi menunjukkan bahwa Naive Bayes dengan parameter optimal alpha=0.1

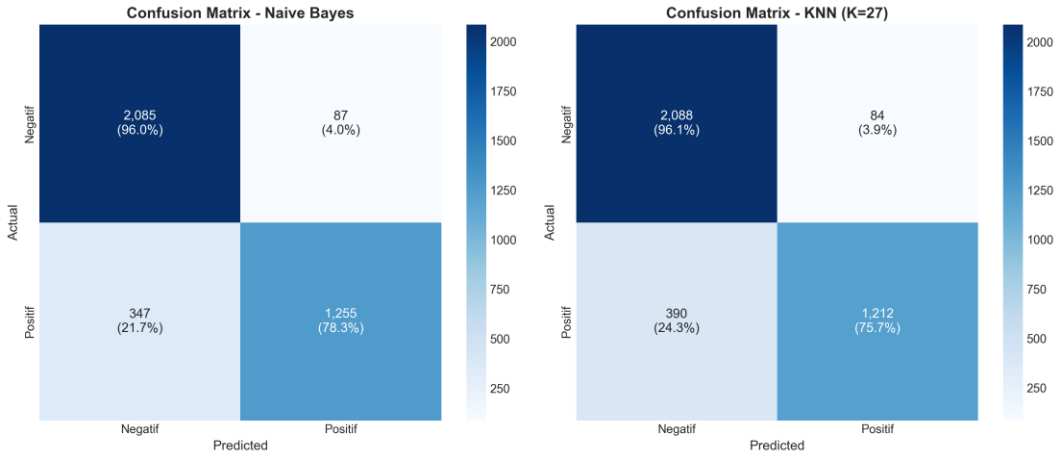
Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

mencapai performa yang superior: akurasi 88,5%, precision 89,0%, recall 88,5%, F1-score 88,3%, dan AUC-ROC 0,9237. Sebagai perbandingan, KNN dengan parameter optimal k=27 mencapai: akurasi 87,4%, precision 88,2%, recall 87,4%, F1-score 87,2%, dan AUC-ROC 0,9203. Naive Bayes secara konsisten mengungguli KNN pada semua metrik evaluasi dengan margin 0,8% hingga 1,1%. Kedua algoritma mencapai nilai AUC-ROC di atas 0,90 yang termasuk kategori klasifikasi "Excellent", menunjukkan kemampuan diskriminasi yang sangat baik untuk membedakan antara ulasan positif dan negatif.



Gambar 12. Kurva ROC: Perbandingan Naive Bayes dan KNN

Kurva ROC pada Gambar 12 memvisualisasikan trade-off antara True Positive Rate dan False Positive Rate untuk kedua algoritma pada berbagai threshold klasifikasi. Kedua kurva berada jauh di atas garis diagonal yang merepresentasikan random classifier, mengkonfirmasi kemampuan diskriminasi yang baik. Naive Bayes menunjukkan area under curve yang sedikit lebih besar, terutama pada region dengan False Positive Rate rendah yang penting untuk aplikasi dimana false alarm perlu diminimalisir.



Gambar 13. Confusion Matrix Naive Bayes dan KNN

Berdasarkan Gambar 13, analisis confusion matrix memberikan insight lebih detail mengenai pola klasifikasi kedua algoritma. Untuk Naive Bayes, confusion matrix menunjukkan True Negative (TN) sebesar 2.085 sampel (96,0% dari total negatif aktual), False Positive (FP) sebesar 87 sampel (4,0%), False Negative (FN) sebesar 347 sampel (21,7% dari total positif aktual), dan True Positive (TP) sebesar 1.255 sampel (78,3%). Sementara untuk KNN, confusion matrix menunjukkan TN sebesar 2.088 sampel (96,1%), FP sebesar 84 sampel (3,9%), FN sebesar 390 sampel (24,3%), dan TP sebesar 1.212 sampel (75,7%). Perbedaan paling mencolok terletak pada False Negative, dimana KNN menghasilkan 43 kesalahan lebih banyak dalam mengklasifikasikan ulasan positif sebagai negatif. Fenomena ini dapat dijelaskan oleh karakteristik algoritma KNN yang sangat bergantung pada kemiripan vektor dalam ruang TF-IDF berdimensi tinggi; ulasan positif yang menggunakan kosakata mirip dengan ulasan negatif (misalnya mengandung kata "transaksi" atau "saldo" dalam konteks positif) cenderung salah diklasifikasikan karena tetangga terdekatnya didominasi oleh ulasan negatif yang lebih banyak dalam dataset. Sebaliknya, Naive Bayes yang menghitung probabilitas independen per kata lebih mampu menangkap pola kombinasi kata positif secara keseluruhan tanpa terpengaruh oleh dominasi kelas negatif dalam neighborhood.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

3.6 Hasil Cross-Validation

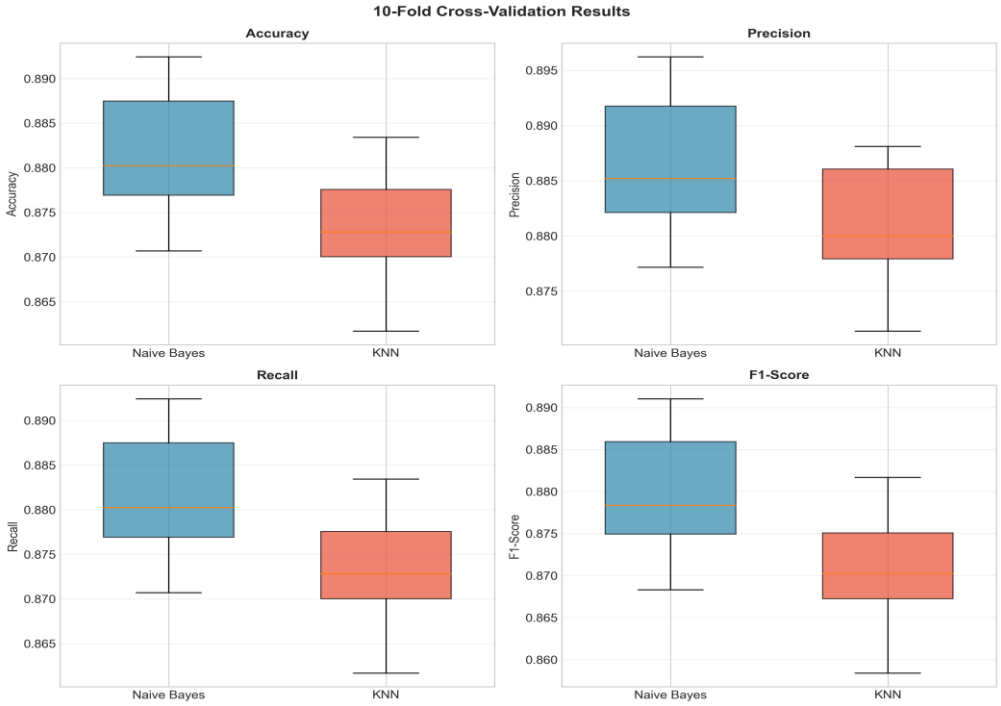
Validasi menggunakan teknik 10-fold cross-validation dilakukan sebagai evaluasi tambahan untuk mengestimasi stabilitas dan generalisasi model pada berbagai partisi data yang berbeda. Teknik cross-validation sangat penting karena memastikan kinerja model tidak bergantung pada pilihan data pelatihan dan pengujian tertentu dan membantu menemukan overfitting.

Tabel 3. Hasil 10-Fold Cross-Validation

Algoritma	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0,8808 ± 0,0070	0,8905 ± 0,0065	0,8808 ± 0,0070	0,8792 ± 0,0074
KNN	0,8727 ± 0,0064	0,8800 ± 0,0058	0,8727 ± 0,0064	0,8696 ± 0,0069

Berdasarkan Tabel 3, Naive Bayes mencapai rata-rata akurasi 0,8808 dengan standar deviasi 0,0070, sementara KNN mencapai rata-rata akurasi 0,8727 dengan standar deviasi 0,0064. Boxplot pada Gambar 14 menunjukkan distribusi performa pada setiap fold, dimana Naive Bayes menunjukkan median yang lebih tinggi dan distribusi yang lebih kompak. Hasil ini mengkonfirmasi bahwa Naive Bayes tidak hanya lebih akurat tetapi juga menunjukkan performa yang konsisten dan stabil di berbagai subset data.

Kedua algoritma menunjukkan kinerja yang konsisten selama cross-validation memberikan jaminan bahwa model tidak mengalami overfitting yang serius terhadap training data. Coefficient of variation untuk kedua algoritma berada di bawah 1%, mengindikasikan variance yang sangat rendah relatif terhadap mean performa. Interquartile range yang sempit pada boxplot mengkonfirmasi bahwa performa tidak bervariasi secara signifikan antar fold. Temuan ini penting karena menunjukkan bahwa hasil evaluasi pada test set dapat diandalkan sebagai estimator yang valid untuk performa model pada data baru yang belum pernah dilihat, memberikan confidence dalam deployment model untuk aplikasi production.



Gambar 14. Hasil 10-Fold Cross-Validation

Pada Gambar 14 menunjukkan distribusi performa pada setiap fold, dimana Naive Bayes menunjukkan median yang lebih tinggi dan distribusi yang lebih kompak. Hasil ini mengkonfirmasi bahwa Naive Bayes tidak hanya lebih akurat tetapi juga menunjukkan performa yang konsisten dan stabil di berbagai subset data.

3.7 Hasil Uji McNemar

Uji McNemar digunakan sebagai metode statistik formal untuk menguji apakah perbedaan performa yang diamati antara Naive Bayes dan KNN secara statistik signifikan. Konstruksi tabel kontingensi menunjukkan: 3.226 sampel diklasifikasikan benar oleh kedua algoritma (85,5%), 74 sampel diklasifikasikan benar hanya oleh KNN (2,0%), 114 sampel diklasifikasikan benar hanya oleh Naive Bayes (3,0%), dan 360 sampel diklasifikasikan salah oleh kedua algoritma (9,5%).

Statistik chi-square untuk uji McNemar dihitung sebagai $\chi^2 = (|114-74|-1)^2/(114+74) = 8,05$ dengan p-value = 0,0045. Karena p-value (0,0045) jauh lebih kecil dari tingkat signifikansi α=0,05, hipotesis nol yang menyatakan tidak

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

ada perbedaan performa ditolak. Hasil ini memberikan bukti statistik yang kuat bahwa Naive Bayes memang secara genuine dan konsisten lebih unggul dibandingkan KNN untuk tugas analisis sentimen pada dataset dompet digital ini.

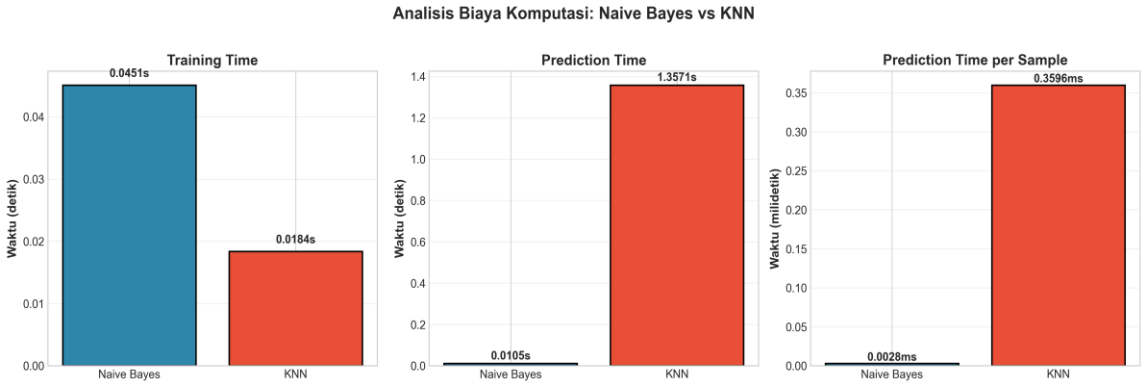
3.8 Analisis Efisiensi Komputasi

Selain metrik akurasi, efisiensi komputasi merupakan faktor penting dalam pemilihan algoritma untuk aplikasi real-time.

Tabel 4. Perbandingan Efisiensi Komputasi

Algoritma	Waktu Training	Waktu Prediksi	Waktu/Sampel
Naive Bayes	0,0451 detik	0,0105 detik	0,0028 ms
KNN	0,0184 detik	1,3571 detik	0,3596 ms

Berdasarkan Tabel 4, Naive Bayes membutuhkan waktu training 0,0451 detik dan waktu prediksi hanya 0,0105 detik untuk 3.774 sampel (0,0028 ms/sampel). KNN membutuhkan waktu training lebih singkat yaitu 0,0184 detik karena sifat lazy learning-nya, namun waktu prediksi jauh lebih lama yaitu 1,3571 detik (0,3596 ms/sampel) [25].

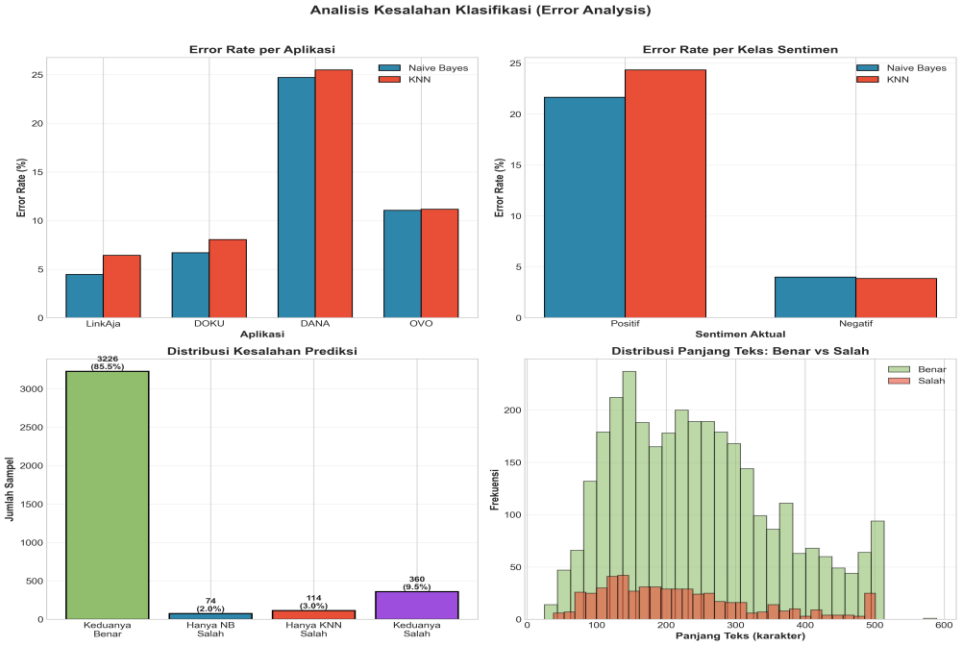


Gambar 15. Analisis Biaya Komputasi: Naive Bayes vs KNN

Merujuk Gambar 15 perbedaan waktu prediksi dapat dijelaskan oleh kompleksitas algoritmik yang berbeda. KNN memiliki kompleksitas waktu O(n×d) untuk setiap prediksi individual, sedangkan Naive Bayes memiliki kompleksitas O(c×d). Dengan Naive Bayes 129 kali lebih cepat dalam prediksi, algoritma ini jelas merupakan pilihan yang superior untuk aplikasi yang membutuhkan klasifikasi real-time pada volume data yang besar.

3.9 Analisis Error

Analisis error dilakukan untuk memahami pola kesalahan klasifikasi kedua algoritma. Tingkat error tertinggi ditemukan pada ulasan aplikasi DANA dengan error rate sekitar 24,5% untuk Naive Bayes dan 25,5% untuk KNN. Aplikasi lain menunjukkan error rate yang lebih rendah: OVO (11%), DOKU (7-8%), dan LinkAja (4,5-6,5%).



Gambar 16. Analisis Kesalahan Klasifikasi

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

Pada Gambar 16 kedua algoritma menunjukkan error rate yang lebih tinggi pada kelas positif (22-24,5%) dibandingkan kelas negatif (sekitar 4%), yang dapat dijelaskan oleh ketidakseimbangan kelas dalam dataset dimana kelas negatif mendominasi dengan proporsi 57,54%. Tingginya error rate pada DANA dapat dijelaskan oleh karakteristik ulasan yang lebih ambigu dan penggunaan bahasa campuran (code-mixing) yang lebih intensif. Selain faktor linguistik, tingginya tingkat kesalahan pada klasifikasi ulasan DANA juga dipengaruhi oleh dua faktor struktural data yang signifikan. Pertama, analisis distribusi kelas menunjukkan adanya ketimpangan yang lebih tajam pada dataset DANA, di mana sentimen negatif mendominasi sebesar 66,22% (2.997 ulasan) berbanding 33,78% sentimen positif (1.529 ulasan). Ketidakseimbangan ini cenderung membiaskan model untuk lebih mengenali pola kelas negatif, sehingga meningkatkan False Negative rate. Kedua, analisis kosakata (vocabulary analysis) mengungkap bahwa ulasan DANA memiliki variasi terminologi teknis yang lebih kompleks dan spesifik ekosistem, seperti kemunculan fitur unik "Dana Cicil", "Dana Goals", "Paylater", dan "Dana Viz" (verifikasi wajah). Istilah-istilah fitur ini sering kali muncul dalam ulasan bersentimen campuran (misalnya: pengguna memuji aplikasi secara umum, namun kecewa karena fitur 'Dana Cicil' tidak muncul), yang menciptakan ambiguitas fitur dalam ruang vektor TF-IDF dan menyulitkan model menarik garis batas keputusan yang tegas.Analisis kualitatif menunjukkan bahwa banyak ulasan mengandung sentimen campuran atau menggunakan sarkasme yang sulit dideteksi oleh algoritma bag-of-words, seperti contoh ulasan "Wah bagus banget ya, saldo saya hilang tapi CS-nya santai aja" yang secara leksikal mengandung kata positif namun bermakna negatif. Ketiadaan penerapan teknik penyeimbangan data sintetis (seperti SMOTE) dalam eksperimen ini teridentifikasi sebagai keterbatasan utama penelitian. Ketimpangan distribusi data sebesar 57,54% (negatif) berbanding 42,46% (positif) menyebabkan bias model yang lebih condong mengenali pola kelas mayoritas. Hal ini dikonfirmasi oleh tingginya tingkat kesalahan (error rate) pada kelas positif (22-24,5%) dibandingkan kelas negatif (sekitar 4%), yang secara langsung berdampak pada rendahnya nilai Recall spesifik untuk sentimen positif. Oleh karena itu, penerapan teknik oversampling seperti SMOTE direkomendasikan bukan hanya sebagai opsi, melainkan sebagai kebutuhan krusial dalam penelitian masa depan untuk memperbaiki sensitivitas model terhadap kelas minoritas [26]

3.10 Perbandingan dengan Penelitian Terdahulu

Perbandingan dengan penelitian terdahulu menunjukkan bahwa hasil penelitian ini kompetitif. Akurasi Naive Bayes sebesar 88,5% lebih tinggi dibandingkan penelitian Wibowo et al. [7] yang mencapai 86,2% pada dataset e-commerce. Penelitian dengan Deep Learning [11] mencapai akurasi 89,3% yang sedikit lebih tinggi, namun dengan waktu training dalam hitungan jam dibandingkan milidetik untuk Naive Bayes. Keunggulan penelitian ini adalah penerapan uji signifikansi statistik McNemar yang memberikan bukti empiris kuat bahwa perbedaan performa bukan hasil kebetulan.

4. KESIMPULAN

Penelitian ini telah berhasil membandingkan performa algoritma Naive Bayes dan K-Nearest Neighbor untuk klasifikasi sentimen ulasan pengguna dompet digital di Indonesia. Berdasarkan eksperimen pada dataset 18.869 ulasan dari empat aplikasi dompet digital (DANA, OVO, DOKU, dan LinkAja), algoritma Naive Bayes dengan parameter alpha=0.1 mencapai performa terbaik dengan akurasi 88,5%, precision 89,0%, recall 88,5%, F1-score 88,3%, dan AUC-ROC 0,9237, mengungguli KNN dengan k=27 yang mencapai akurasi 87,4% dan AUC-ROC 0,9203. Perbedaan performa ini dikonfirmasi signifikan secara statistik melalui uji McNemar dengan p-value=0,0045 pada tingkat signifikansi α=0,05. Selain keunggulan akurasi, Naive Bayes juga terbukti 129 kali lebih cepat dalam fase prediksi dibandingkan KNN, menjadikannya pilihan yang lebih cocok untuk implementasi sistem analisis sentimen real-time. Temuan lain yang menarik adalah perbedaan vocabulary yang signifikan antara sentimen positif dan negatif, dimana sentimen positif didominasi kata-kata terkait kemudahan transaksi ("bayar", "mudah", "beli"), sedangkan sentimen negatif didominasi kata-kata terkait masalah akun dan saldo ("saldo", "masuk", "akun", "gagal"). Penelitian ini memiliki keterbatasan dalam penggunaan pendekatan binary classification dan ketidakseimbangan kelas yang belum ditangani secara khusus. Penelitian selanjutnya dapat mengeksplorasi pendekatan multiclass classification, penerapan teknik deep learning seperti IndoBERT untuk menangani konteks bahasa Indonesia yang lebih kompleks termasuk sarkasme dan code-mixing, serta penerapan teknik balancing seperti SMOTE untuk mengatasi ketidakseimbangan kelas.

REFERENCES

[1] A. Ramadhani, D. Putu, and Y. Pardita, “Manajemen transformasi digital pembayaran: Faktor-faktor adopsi e-money dan implikasinya pada velocity of money,” El-Mal: *Jurnal Kajian Ekonomi & Bisnis Islam*, vol. 6, no. 9, 2025.
[2] Asosiasi Penyelenggara Jasa Internet Indonesia, “APJII jumlah pengguna internet Indonesia tembus 221 juta orang,” 2023. [Online]. Available: https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang
[3] A. Ciptarianto, “E-wallet application penetration for financial inclusion in Indonesia,” *International Journal of Current Science Research and Review*, vol. 5, no. 2, 2022, doi: 10.47191/ijcsrr/v5-i2-03.
[4] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” *Knowledge-Based Systems*, vol. 226, 2021, doi: 10.1016/j.knosys.2021.107134.
[5] K. Naithani and Y. Raiwani, “Realization of natural language processing and machine learning approaches for text-based

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2335−2348
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9285

sentiment analysis," Expert Systems, vol. 40, 2022, doi: 10.1111/exsy.13114.
[6] L. Bharadwaj, "Sentiment analysis in online product reviews: Mining customer opinions for sentiment classification," *International Journal for Multidisciplinary Research*, vol. 5, no. 5, 2023, doi: 10.36948/ijfmr.2023.v05i05.6090.
[7] A. Wibowo, S. Rahayu, and D. Kusuma, "Analisis sentimen ulasan aplikasi e-commerce menggunakan Naive Bayes dengan feature selection chi-square," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 3, 2023.
[8] D. Rahmawati and B. Kusuma, "Perbandingan algoritma Naive Bayes dan SVM untuk klasifikasi sentimen ulasan transportasi online," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 17, no. 2, 2023.
[9] H. Santoso, M. Firdaus, and A. Pramono, "Optimasi parameter K pada algoritma K-nearest neighbor untuk klasifikasi sentimen ulasan hotel," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 8, no. 4, 2022.
[10] A. Fattahila et al., "Indonesian digital wallet sentiment analysis using CNN and LSTM method," in Proc. Int. Conf. Artificial Intelligence and Big Data Analytics, 2021, doi: 10.1109/ICAIBDA53487.2021.9689712.
[11] A. Pandey, A. Kajla, D. Shrivastava, and C. Samadiya, "Spam email detection using machine learning," IMRJR, 2025, doi: 10.17148/imrjr.2025.020403.
[12] M. Romano, G. Contu, F. Mola, and C. Conversano, "Threshold-based Naïve Bayes classifier," Advances in Data Analysis and Classification, 2023, doi: 10.1007/s11634-023-00536-8.
[13] J. W. Fachezi, "Google-play-scraper: Python library for scraping Google Play Store," GitHub Repository. [Online]. Available: https://github.com/JoMingyu/google-play-scraper
[14] S. Mola, D. Polly, and N. Rumlaklak, "Sentiment analysis on user reviews of the Edlink application using the random forest classifier method," *Jurnal Sisfotek Global*, vol. 15, no. 1, 2025, doi: 10.38101/sisfotek.v15i1.15788.
[15] A. F. Aji, "Sastrawi: High quality Indonesian stemmer," GitHub Repository. [Online]. Available: https://github.com/sastrawi/sastrawi
[16] M. Habibi and P. W. Cahyo, "Clustering user characteristics based on the influence of hashtags on the Instagram platform," *Indonesian Journal of Computing and Cybernetics Systems,* vol. 13, no. 4, 2021, doi: 10.22146/ijccs.50574.
[17] N. Semary et al., "Enhancing machine learning-based sentiment analysis through feature extraction techniques," *PLOS ONE*, vol. 19, 2024, doi: 10.1371/journal.pone.0294968.
[18] H. Xu and Y. Li, "Classification of news texts based on Bayes algorithm," in Proc. Int. Conf. Electronic Information Technology and Computer Engineering, 2021, doi: 10.1145/3501409.3501636.
[19] M. Aditya, A. Helen, and I. Suryana, "Naïve Bayes and maximum entropy comparison for translated novel's genre classification," *Journal of Physics: Conference Series,* vol. 1722, 2021, doi: 10.1088/1742-6596/1722/1/012007.
[20] R. Halder et al., "Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, 2024, doi: 10.1186/s40537-024-00973-y.
[21] A. Jalal and B. Ali, "Text documents clustering using data mining techniques," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 1, 2021, doi: 10.11591/ijece.v11i1.pp664-670.
[22] J. Brownlee, "How to configure k-fold cross-validation," Machine Learning Mastery, 2021. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/
[23] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," Scientific Reports, vol. 14, 2024, doi: 10.1038/s41598-024-56706-x.
[24] F. Sabiq et al., "Performance comparison of multinomial and Bernoulli Naïve Bayes algorithms with Laplace smoothing optimization in fake news classification," *in Proc. Int. Conf. Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics*, 2024, doi: 10.1109/ICOABCD63526.2024.10704399.
[25] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Stanford University Press, 2024.
[26] A. Sharma, P. Singh, and R. Chandra, "SMOTified-GAN for class imbalanced pattern classification problems," *IEEE Access,* vol. 10, 2022, doi: 10.1109/ACCESS.2022.3158977.