Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika
Volume. 4 Nomor. 4 Juli 2026
E-ISSN .: 3031-8912; P-ISSN .: 3031-8904, Hal. 157-176
DOI: https://doi.org/10.61132/merkurius.v4i4.1713
Tersedia: https://journal.arteii.or.id/index.php/Merkurius

Klasifikasi Sentimen Pengguna Terhadap Pembayaran Digital Menggunakan Algoritma Naive Bayes

Nayla Syafiah Asmi[1*], Sinta Amellyya Sari[2], Roberto Kaban[3]
[1-2]Fakultas Sains Dan Teknologi, Program Studi Teknik Informatika, Institut Teknologi dan Bisnis Indonesia
[*]*Penulis Korespondensi: Nayla535677@gmail.com*

Abstract. *The rapid development of financial technology (fintech) in Indonesia has led to the increasing use of digital payment applications, one of which is DANA. The large number of user reviews on the Google Play Store provides textual data that can be utilized to understand users' perceptions of the application's service quality. This study aims to classify user sentiment toward the DANA application using the Naïve Bayes algorithm. The dataset used in this study was obtained from Kaggle and consists of 50,000 user reviews categorized into three sentiment classes: positive, negative, and neutral. The research process includes data collection, text preprocessing (case folding, cleaning, tokenizing, stopword removal, and stemming), feature weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method, sentiment classification using the Multinomial Naïve Bayes algorithm, and model evaluation using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The experimental results show that the proposed model achieved an accuracy of 79.26%. The positive sentiment class obtained the best performance with a precision of 0.85, recall of 0.94, and F1-score of 0.89, while the neutral sentiment class produced the lowest recall (0.19) due to the imbalance in class distribution. These findings indicate that the Naïve Bayes algorithm provides satisfactory performance for sentiment classification of DANA user reviews and can be applied as an effective approach to support the evaluation and improvement of digital payment application services.*

Keywords: *DANA; Google Play Store; Naïve Bayes; TF-IDF; Sentiment Analysis.*

Abstrak. Perkembangan teknologi finansial (fintech) di Indonesia telah mendorong meningkatnya penggunaan aplikasi pembayaran digital, salah satunya DANA. Banyaknya ulasan pengguna pada Google Play Store menghasilkan data tekstual yang dapat dimanfaatkan untuk mengetahui persepsi pengguna terhadap kualitas layanan aplikasi. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna aplikasi DANA menggunakan algoritma Naïve Bayes. Dataset yang digunakan berasal dari Kaggle dengan jumlah 50.000 ulasan yang telah dikategorikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Tahapan penelitian meliputi pengumpulan data, preprocessing teks (case folding, cleaning, tokenizing, stopword removal, dan stemming), pembobotan kata menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), proses klasifikasi menggunakan algoritma Multinomial Naïve Bayes, serta evaluasi model menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model memperoleh nilai accuracy sebesar 79,26%. Kelas sentimen positif memiliki performa terbaik dengan nilai precision sebesar 0,85, recall 0,94, dan F1-score 0,89, sedangkan kelas sentimen netral memiliki nilai recall terendah sebesar 0,19 akibat ketidakseimbangan distribusi data. Berdasarkan hasil tersebut, algoritma Naïve Bayes mampu memberikan kinerja yang cukup baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi DANA dan dapat dimanfaatkan sebagai salah satu metode analisis opini pengguna untuk mendukung peningkatan kualitas layanan aplikasi pembayaran digital.

Kata kunci: Analisis Sentimen; DANA; Google Play Store; Naïve Bayes; TF-IDF.

1. LATAR BELAKANG

Perkembangan layanan keuangan berbasis digital telah mengubah kebiasaan masyarakat dalam melakukan transaksi sehari-hari. Pembayaran yang sebelumnya banyak bergantung pada uang tunai kini dapat dilakukan melalui aplikasi dompet digital untuk berbagai kebutuhan, seperti pembayaran belanja, pembelian produk digital, pembayaran tagihan, hingga transfer dana. DANA menjadi salah satu aplikasi yang digunakan dalam ekosistem tersebut bersama sejumlah layanan pembayaran digital lainnya. Peningkatan penggunaan aplikasi tidak hanya menghasilkan aktivitas transaksi, tetapi juga mendorong

Naskah Masuk: 17 Maret 2026; Revisi: 16 April 2026; Diterima: 28 Mei 2026; Terbit: 31 Mei 2026

munculnya berbagai tanggapan pengguna mengenai pengalaman mereka selama memanfaatkan layanan. Utami dan Dandi (2026) menjelaskan bahwa keberadaan beragam platform dompet digital membentuk ruang persaingan layanan yang dapat diamati melalui respons dan sentimen pengguna.

Salah satu sumber yang merekam pengalaman tersebut adalah ulasan pada Google Play Store. Pengguna dapat menyampaikan penilaian mengenai kemudahan aplikasi, keberhasilan transaksi, akses akun, kondisi saldo, maupun kendala teknis yang mereka alami. Ramadhan dan Sugianto (2024) menunjukkan bahwa ulasan aplikasi DANA dapat dimanfaatkan sebagai sumber data untuk membaca kecenderungan sentimen pengguna. Informasi semacam ini memiliki nilai evaluatif karena komentar yang diberikan berasal dari pengalaman penggunaan aplikasi. Namun, ketika jumlah ulasan mencapai skala yang besar, pemeriksaan secara manual menjadi kurang praktis, membutuhkan waktu panjang, dan sulit menjaga konsistensi penilaian terhadap setiap komentar.

Kondisi tersebut mendorong penggunaan analisis sentimen sebagai pendekatan untuk mengolah opini pengguna secara lebih terstruktur. Analisis sentimen memungkinkan teks ulasan dipetakan berdasarkan kecenderungan penilaian yang terkandung di dalamnya. Ibrahim et al. (2024) memanfaatkan pendekatan klasifikasi untuk menganalisis sentimen pengguna layanan *e-wallet* OVO dan DANA, sedangkan Maharani et al. (2022) membandingkan Naive Bayes dan K-Nearest Neighbor dalam klasifikasi sentimen penggunaan *e-wallet*. Penelitian lain juga memperlihatkan bahwa pengolahan opini pengguna dompet digital dapat dilakukan dengan algoritma Random Forest, Naive Bayes, maupun Support Vector Machine sesuai karakteristik data dan tujuan analisis yang digunakan (Hapsari & Indriyanti, 2023; Ardiansyah et al., 2025).

Di antara berbagai metode klasifikasi tersebut, Naive Bayes tetap banyak diterapkan pada pengolahan data teks karena mekanisme probabilistiknya relatif sederhana dan proses komputasinya efisien. Umair dan Susanto (2024) menggunakan Naive Bayes untuk menganalisis sentimen ulasan pengguna aplikasi BRImo, sedangkan Riskawati et al. (2024) menerapkan metode yang sama pada ulasan aplikasi GoPay. Sejumlah penelitian tersebut memperlihatkan bahwa performa klasifikasi tidak selalu menghasilkan capaian yang seragam. Perbedaan jumlah data, tahapan praproses, representasi fitur, komposisi kelas, dan karakteristik bahasa dalam ulasan dapat memengaruhi kemampuan model ketika mengenali sentimen pengguna. Dengan demikian, keberhasilan suatu algoritma pada satu objek aplikasi belum dapat langsung dianggap mewakili kondisi pada aplikasi pembayaran digital lainnya.

Penelitian terdahulu mengenai sentimen dompet digital juga banyak diarahkan pada perbandingan antaralgoritma atau klasifikasi dua kecenderungan sentimen. Maharani et al. (2022), misalnya, membandingkan Naive Bayes dengan K-Nearest Neighbor, sementara Ardiansyah et al. (2025) membandingkan Naive Bayes dan Support Vector Machine pada layanan keuangan digital. Kajian semacam itu penting untuk melihat perbedaan performa metode, tetapi masih terdapat kebutuhan untuk menguji kemampuan model pada data ulasan berskala besar dengan tiga kategori sentimen yang tidak memiliki proporsi seimbang. Kelas netral menjadi perhatian khusus karena ekspresi pengguna pada kategori ini sering kali tidak menunjukkan penilaian positif atau negatif secara tegas, bahkan dapat berbentuk pertanyaan, permintaan bantuan, maupun pernyataan informatif.

Berdasarkan kondisi tersebut, penelitian ini menempatkan kebaruan pada penerapan Multinomial Naive Bayes terhadap 50.000 ulasan pengguna aplikasi DANA yang diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Jumlah data yang relatif besar memberikan ruang untuk menilai kemampuan model pada variasi teks pengguna yang lebih luas. Pada saat yang sama, perbedaan proporsi antar kelas menjadi aspek penting karena dapat memengaruhi kemampuan klasifikasi, terutama dalam mengenali sentimen dengan jumlah data lebih sedikit. Data teks diproses melalui tahapan praproses, direpresentasikan menggunakan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), kemudian digunakan dalam pembentukan model klasifikasi.

Penelitian ini bertujuan menerapkan algoritma Multinomial Naive Bayes untuk mengklasifikasikan sentimen pengguna terhadap aplikasi pembayaran digital DANA berdasarkan ulasan pada Google Play Store. Analisis dilakukan terhadap tiga kategori sentimen, yaitu positif, negatif, dan netral, serta dilengkapi evaluasi kinerja model untuk melihat kemampuan klasifikasi pada masing-masing kelas. Hasil penelitian diharapkan dapat memberikan gambaran yang lebih terukur mengenai pola persepsi pengguna terhadap layanan DANA dan menjadi bahan evaluasi berbasis data bagi pengembangan kualitas aplikasi pembayaran digital.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen merupakan pendekatan dalam pengolahan teks yang digunakan untuk mengenali kecenderungan opini, penilaian, atau respons yang terkandung dalam suatu dokumen. Data tekstual yang berasal dari komentar pengguna dapat dikelompokkan berdasarkan orientasi sentimennya, misalnya positif, negatif, atau netral. Dalam konteks

layanan digital, pendekatan ini membantu mengubah kumpulan opini yang tidak terstruktur menjadi informasi yang lebih mudah dianalisis. Ibrahim et al. (2024) menerapkan analisis sentimen terhadap pengguna *e-wallet* OVO dan DANA, sedangkan Maharani et al. (2022) menggunakan klasifikasi sentimen untuk mengkaji respons pengguna terhadap layanan *e-wallet* melalui perbandingan beberapa metode klasifikasi.

Penelitian ini memandang ulasan pengguna sebagai representasi pengalaman yang muncul setelah berinteraksi dengan aplikasi DANA. Isi ulasan dapat memuat apresiasi terhadap kemudahan layanan, keluhan mengenai transaksi dan saldo, pertanyaan terkait akun, maupun permintaan bantuan.

Ulasan Pengguna Aplikasi Pembayaran Digital

Ulasan pengguna merupakan bentuk umpan balik tekstual yang diberikan setelah seseorang menggunakan suatu aplikasi atau layanan digital. Pada platform seperti Google Play Store, ulasan dapat menggambarkan pengalaman pengguna terhadap fungsi aplikasi, kemudahan penggunaan, kestabilan sistem, transaksi, akses akun, maupun persoalan teknis lainnya. Ramadhan dan Sugianto (2024) menunjukkan bahwa ulasan aplikasi DANA dapat digunakan sebagai sumber data dalam proses klasifikasi sentimen menggunakan Naive Bayes. Riskawati et al. (2024) juga memanfaatkan ulasan aplikasi GoPay untuk membaca kecenderungan sentimen pengguna melalui pendekatan klasifikasi.

Karakteristik ulasan aplikasi pembayaran digital cenderung beragam karena ditulis secara bebas oleh pengguna. Satu komentar dapat berupa kalimat singkat, bahasa tidak baku, singkatan, pengulangan kata, maupun kombinasi ekspresi yang berbeda. Pada penelitian ini, ulasan pengguna DANA ditempatkan sebagai sumber utama untuk mengenali kecenderungan persepsi terhadap layanan pembayaran digital.

Text Preprocessing

Text preprocessing merupakan tahap persiapan data yang dilakukan sebelum teks digunakan dalam ekstraksi fitur dan proses klasifikasi. Ulasan pengguna pada Google Play Store umumnya memiliki bentuk yang tidak seragam karena dapat mengandung variasi huruf kapital, tanda baca, simbol, kata umum, serta bentuk kata berimbuhan. Kondisi tersebut perlu ditangani agar perbedaan penulisan yang tidak berkaitan langsung dengan orientasi sentimen tidak memperbesar keragaman fitur. Dalam penelitian mengenai sentimen layanan pembayaran digital, Al Anshari et al. (2023) menempatkan pengolahan awal teks sebagai bagian penting dalam menyiapkan ulasan pengguna sebelum proses klasifikasi. Hal ini relevan dengan penelitian ini karena data yang dianalisis juga berupa opini pengguna aplikasi pembayaran digital, khususnya DANA.

Pada penelitian ini, praproses mencakup *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. *Case folding* menyeragamkan karakter menjadi huruf kecil, sedangkan *cleaning* membersihkan unsur teks yang tidak diperlukan. *Tokenizing* memecah kalimat menjadi unit kata, kemudian *stopword removal* mengurangi kata umum yang memiliki kontribusi terbatas terhadap proses klasifikasi. Tahap *stemming* mengembalikan kata berimbuhan ke bentuk dasarnya. Rangkaian tersebut dipilih agar ulasan DANA yang semula memiliki variasi penulisan dapat direpresentasikan dalam bentuk yang lebih konsisten sebelum memasuki pembobotan fitur.

Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency atau TF-IDF merupakan metode pembobotan yang digunakan untuk mengubah data teks menjadi representasi numerik berdasarkan tingkat kepentingan suatu kata dalam dokumen. Komponen *Term Frequency* memperhitungkan frekuensi kemunculan kata pada sebuah dokumen, sedangkan *Inverse Document Frequency* mempertimbangkan penyebaran kata tersebut pada keseluruhan kumpulan dokumen. Melalui mekanisme ini, kata yang memiliki kemunculan lebih khas pada dokumen tertentu dapat memperoleh bobot yang lebih informatif dibandingkan kata yang tersebar luas pada banyak dokumen.

Penggunaan representasi fitur berbasis TF-IDF memiliki keterkaitan dengan penelitian Al Anshari et al. (2023) yang menganalisis sentimen pembayaran digital berdasarkan ulasan Google Play Store menggunakan pendekatan klasifikasi *Support Vector Machine*. Meskipun algoritma klasifikasi yang digunakan berbeda, penelitian tersebut memperlihatkan relevansi transformasi data ulasan menjadi fitur numerik dalam analisis opini pengguna layanan pembayaran digital. Pada penelitian ini, TF-IDF digunakan untuk merepresentasikan setiap ulasan DANA dalam bentuk bobot kata sebelum diproses oleh Multinomial Naive Bayes. Dengan demikian, pola kosakata yang terbentuk dari pengalaman pengguna terhadap transaksi, saldo, akun, akses aplikasi, maupun kualitas layanan dapat dipelajari secara kuantitatif oleh model klasifikasi.

Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik yang menggunakan Teorema Bayes untuk memperkirakan kemungkinan suatu data berada pada kelas tertentu. Proses klasifikasi mempertimbangkan probabilitas awal suatu kelas dan peluang kemunculan fitur berdasarkan kelas tersebut. Metode ini menggunakan asumsi independensi kondisional antarf fitur, yaitu setiap fitur diperlakukan memiliki kontribusi tersendiri terhadap kelas yang diprediksi Umair dan Susanto (2024) menerapkan Naive Bayes dalam analisis sentimen ulasan

pengguna aplikasi BRImo, sedangkan Riskawati et al. (2024) menggunakannya pada analisis sentimen aplikasi GoPay. Dalam penelitian ini, Naive Bayes digunakan untuk memperkirakan kecenderungan sebuah ulasan DANA terhadap tiga kemungkinan kelas, yaitu positif, negatif, atau netral, berdasarkan pola fitur teks yang terbentuk setelah proses praproses dan pembobotan.

Multinomial Naive Bayes

Multinomial Naive Bayes merupakan salah satu varian Naive Bayes yang banyak digunakan untuk klasifikasi dokumen teks berdasarkan pola kemunculan fitur. Berbeda dari pendekatan Gaussian yang ditujukan pada atribut numerik kontinu, model multinomial lebih sesuai untuk representasi teks berbasis frekuensi atau bobot kata. Utami dan Dandi (2026) menjelaskan pemanfaatan Naive Bayes dalam analisis sentimen dompet digital, sedangkan Umair dan Susanto (2024) menunjukkan penerapan pendekatan Naive Bayes untuk mengolah ulasan tekstual pengguna aplikasi BRImo. Kedua penelitian tersebut memperlihatkan bahwa pendekatan probabilistik dapat digunakan untuk mengenali kecenderungan kelas berdasarkan karakteristik kata yang terdapat dalam dokumen.

Pada penelitian ini, Multinomial Naive Bayes dipilih karena data yang dianalisis berbentuk ulasan pengguna dan telah ditransformasikan menjadi representasi numerik menggunakan TF-IDF. Model mempelajari distribusi fitur pada tiga kelas sentimen, yaitu positif, negatif, dan netral, kemudian menggunakan pola tersebut untuk memperkirakan kelas dari ulasan yang belum diprediksi.

Evaluasi Kinerja Model

Evaluasi kinerja diperlukan untuk mengetahui kemampuan model dalam menghasilkan prediksi yang sesuai dengan kelas aktual. Salah satu instrumen yang umum digunakan adalah *confusion matrix*, yaitu matriks yang memperlihatkan hubungan antara hasil prediksi dan label sebenarnya. Dari matriks tersebut dapat diketahui jumlah prediksi yang tepat maupun keliru pada setiap kelas. Umair dan Susanto (2024) menggunakan evaluasi klasifikasi dalam analisis sentimen berbasis Naive Bayes, sedangkan Riskawati et al. (2024) menerapkan pendekatan evaluasi pada klasifikasi sentimen aplikasi GoPay. Penggunaan evaluasi tersebut membantu menilai performa model secara lebih rinci dibandingkan hanya melihat jumlah prediksi benar secara keseluruhan.

Penggunaan beberapa metrik menjadi penting pada klasifikasi yang melibatkan distribusi kelas tidak seimbang. Nilai *accuracy* yang tinggi belum tentu menunjukkan bahwa seluruh kelas dapat dikenali dengan kemampuan yang setara, terutama ketika salah satu kategori memiliki jumlah data jauh lebih sedikit. Dalam penelitian ini, evaluasi dilakukan

melalui *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* agar kemampuan Multinomial Naive Bayes dapat diamati tidak hanya secara keseluruhan, tetapi juga berdasarkan karakteristik prediksi pada setiap kelas sentimen.

Penelitian Terdahulu

Penelitian mengenai klasifikasi sentimen pada layanan pembayaran digital telah berkembang dengan memanfaatkan beragam algoritma dan sumber data. Perbandingan penelitian yang memiliki keterkaitan dengan penelitian ini disajikan pada Tabel 1.

Tabel 1. Perbandingan Penelitian Terdahulu

Peneliti	Fokus Penelitian	Metode	Hasil/Kontribusi Utama	Keterkaitan dengan Penelitian Ini
Maharani et al. (2022)	Klasifikasi sentimen penggunaan *e-wallet*	Naive Bayes dan K-Nearest Neighbor	Membandingkan dua metode klasifikasi untuk menilai kecenderungan sentimen pengguna *e-wallet*	Menjadi dasar perbandingan penggunaan Naive Bayes pada data opini pengguna layanan pembayaran digital
Hapsari & Indriyanti (2023)	Analisis sentimen aplikasi dompet digital	Random Forest	Menunjukkan pemanfaatan algoritma klasifikasi berbasis *machine learning* untuk membaca opini pengguna dompet digital	Memberikan perspektif metode alternatif, sedangkan penelitian ini menggunakan Multinomial Naive Bayes
Ibrahim et al. (2024)	Analisis sentimen pengguna OVO dan DANA	Naive Bayes Classifier	Menerapkan klasifikasi sentimen pada dua layanan *e-wallet* berdasarkan opini pengguna	Relevan dari sisi objek DANA dan klasifikasi sentimen, tetapi penelitian ini berfokus pada ulasan Google Play Store
Ramadhan & Sugianto (2024)	Sentimen ulasan aplikasi DANA di Google Play Store	Naive Bayes	Menggunakan ulasan pengguna DANA sebagai sumber data untuk klasifikasi sentimen	Memiliki kedekatan objek dan metode, sedangkan penelitian ini menggunakan 50.000 ulasan dalam tiga kelas sentimen
Umair & Susanto (2024)	Sentimen ulasan pengguna aplikasi BRImo	Naive Bayes	Menunjukkan penerapan Naive Bayes untuk mengolah opini tekstual pada aplikasi layanan keuangan	Memperkuat penggunaan Naive Bayes pada ulasan aplikasi finansial, tetapi objek penelitian berbeda
Riskawati et al. (2024)	Analisis sentimen aplikasi GoPay	Naive Bayes Classifier	Menerapkan pendekatan probabilistik untuk mengklasifikasikan ulasan pengguna layanan pembayaran digital	Mendukung relevansi Naive Bayes pada data teks aplikasi pembayaran, sedangkan penelitian ini menggunakan objek DANA
Ardiansyah et al. (2025)	Sentimen pengguna GoPay pada layanan keuangan digital	Naive Bayes dan Support Vector Machine	Membandingkan performa dua algoritma pada klasifikasi sentimen layanan keuangan digital	Menjadi pembanding metodologis, sedangkan penelitian ini memusatkan analisis pada Multinomial Naive Bayes

Sumber: Data Penelitian, 2026

Dari Tabel 1 yang telah disajikan, penelitian terdahulu menunjukkan bahwa analisis sentimen pada layanan pembayaran digital telah dilakukan menggunakan Naive Bayes, K-Nearest Neighbor, Random Forest, dan Support Vector Machine. Objek yang dianalisis juga

beragam, mencakup DANA, OVO, GoPay, BRImo, dan layanan *e-wallet* secara umum. Sebagian penelitian menempatkan perhatian pada perbandingan performa antaralgoritma, sedangkan penelitian lainnya berfokus pada penerapan satu metode klasifikasi terhadap objek aplikasi tertentu.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan penerapan teknik *machine learning* untuk mengklasifikasikan sentimen pengguna terhadap aplikasi pembayaran digital DANA. Data penelitian berupa 50.000 ulasan pengguna yang berasal dari dataset Kaggle dan terbagi ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Seluruh data digunakan dalam proses analisis melalui tahapan *text preprocessing*, ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), pembagian data latih dan data uji, klasifikasi menggunakan Multinomial Naive Bayes, serta evaluasi kinerja model.

Tahap awal dilakukan melalui *text preprocessing* untuk membentuk data ulasan yang lebih konsisten sebelum digunakan dalam proses klasifikasi. Tahapan tersebut meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. *Case folding* digunakan untuk menyeragamkan seluruh karakter menjadi huruf kecil, sedangkan *cleaning* menghilangkan unsur teks yang tidak diperlukan. Selanjutnya, *tokenizing* memecah ulasan menjadi unit kata, *stopword removal* mengurangi kata umum yang memiliki kontribusi terbatas terhadap pembentukan sentimen, dan *stemming* mengubah kata berimbuhan ke bentuk dasarnya.

Setelah tahap praproses, data teks dikonversi menjadi representasi numerik menggunakan TF-IDF dengan jumlah maksimum 3.000 fitur. Pembobotan ini mempertimbangkan frekuensi suatu kata dalam dokumen dan tingkat penyebarannya pada keseluruhan kumpulan ulasan.

Term Frequency dihitung menggunakan Persamaan (1).

$$TF(t,d) = \frac{f(t,d)}{\sum f(t,d)} \qquad (1)$$

dengan $f(t,d)$merupakan frekuensi kemunculan kata $t$pada dokumen d, sedangkan $\sum f(t,d)$menunjukkan keseluruhan frekuensi kata dalam dokumen tersebut. Nilai *Inverse Document Frequency* dihitung menggunakan Persamaan (2).

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \qquad (2)$$

dengan Nmerupakan jumlah keseluruhan dokumen dan $df(t)$menunjukkan jumlah dokumen yang mengandung kata t. Bobot akhir setiap kata diperoleh dengan mengalikan nilai TF dan IDF sebagaimana ditunjukkan pada Persamaan (3).

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \qquad (3)$$

Melalui pembobotan tersebut, kata yang memiliki karakteristik lebih informatif terhadap suatu ulasan dapat memperoleh bobot yang berbeda dibandingkan kata yang tersebar luas pada keseluruhan dokumen. Pada penelitian ini, TF-IDF digunakan untuk membentuk vektor numerik dari ulasan pengguna DANA agar dapat diproses oleh model klasifikasi. Implementasi dilakukan menggunakan pustaka *scikit-learn* dengan parameter max_features sebanyak 3.000 kata.

Proses klasifikasi dilakukan menggunakan Multinomial Naive Bayes, yaitu varian Naive Bayes yang sesuai untuk pengolahan dokumen teks berbasis fitur kata. Algoritma ini menggunakan Teorema Bayes untuk menghitung probabilitas suatu ulasan termasuk ke dalam kelas positif, negatif, atau netral berdasarkan pola fitur yang dipelajari dari data latih. Probabilitas posterior dihitung menggunakan Persamaan (4).

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \qquad (4)$$

Kinerja model dievaluasi menggunakan *confusion matrix* untuk membandingkan kelas aktual dengan kelas hasil prediksi. Evaluasi kemudian dihitung melalui metrik *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* menunjukkan proporsi keseluruhan prediksi yang sesuai dengan kelas aktual dan dihitung menggunakan Persamaan (5).

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

Precision digunakan untuk mengukur ketepatan prediksi model terhadap kelas yang dinyatakan positif dan dihitung menggunakan Persamaan (6).

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

Recall menunjukkan kemampuan model dalam mengenali data yang benar-benar termasuk pada kelas yang dianalisis dan dihitung menggunakan Persamaan (7).

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

Selanjutnya, *F1-score* digunakan untuk mengukur keseimbangan antara *precision* dan *recall* sebagaimana ditunjukkan pada Persamaan (8).

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (8)$$

4. HASIL DAN PEMBAHASAN

Karakteristik dan Distribusi Sentimen Pengguna

Data penelitian terdiri atas 50.000 ulasan pengguna aplikasi DANA yang dikelompokkan ke dalam tiga kategori sentimen, yaitu *Positive*, *Negative*, dan *Neutral*. Hasil pengolahan menunjukkan bahwa kelas positif memiliki jumlah terbesar sebanyak 26.555 ulasan, diikuti kelas negatif sebanyak 17.073 ulasan, sedangkan kelas netral berjumlah 6.372 ulasan. Komposisi tersebut memperlihatkan adanya perbedaan proporsi yang cukup jelas antarkelas sentimen.

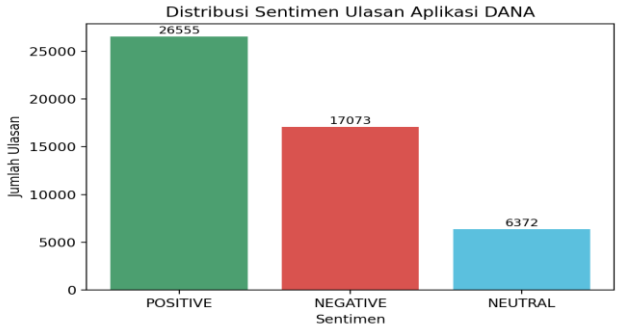
Tabel 2. Distribusi Data Berdasarkan Label Sentimen

Label Sentimen	Jumlah Data	Persentase
Positive	26.555	53,11%
Negative	17.073	34,15%
Neutral	6.372	12,74%
Total	**50.000**	**100,00%**

Sumber: Data Penelitian, 2026

Berdasarkan Tabel 2 di atas, kelas positif mencakup 53,11% dari keseluruhan data dan menjadi kategori dengan proporsi terbesar. Kelas negatif berada pada urutan kedua dengan persentase 34,15%, sementara kelas netral hanya mewakili 12,74% data. Selisih antara kelas positif dan netral mencapai 20.183 ulasan atau 40,37 poin persentase. Pada saat yang sama, jumlah ulasan negatif lebih banyak 10.701 data dibandingkan ulasan netral. Perbedaan tersebut menunjukkan bahwa distribusi kelas pada dataset bersifat tidak seimbang.

Ketimpangan distribusi menjadi aspek penting dalam proses klasifikasi karena jumlah data pada setiap kelas menentukan banyaknya pola tekstual yang tersedia selama pelatihan model. Kelas positif menyediakan contoh yang lebih banyak untuk membentuk distribusi fitur, sedangkan kelas netral memiliki keterwakilan yang jauh lebih terbatas. Situasi ini dapat memengaruhi kemampuan Multinomial Naive Bayes ketika membedakan ulasan netral dari kategori lain, terutama apabila kosakata yang digunakan memiliki kemiripan dengan ekspresi negatif atau positif.



Gambar 1. Distribusi Sentimen Ulasan Pengguna Aplikasi DANA

Dari Gambar 1 berikut, dominasi kelas positif terlihat cukup kuat dibandingkan kategori lainnya. Meskipun lebih dari separuh ulasan berada pada sentimen positif, jumlah data negatif tetap mencapai 17.073 ulasan atau sekitar sepertiga dari keseluruhan dataset. Komposisi tersebut menunjukkan bahwa persepsi pengguna terhadap aplikasi DANA tidak sepenuhnya terkonsentrasi pada pengalaman positif karena terdapat kelompok ulasan negatif dalam jumlah yang cukup besar. Sementara itu, kelas netral menempati proporsi terkecil dan memiliki jarak distribusi yang cukup lebar terhadap dua kategori lainnya.

Hasil Klasifikasi Multinomial Naive Bayes

Proses klasifikasi dilakukan menggunakan Multinomial Naive Bayes terhadap data ulasan yang telah melalui tahap *preprocessing* dan pembobotan TF-IDF. Dari total 50.000 ulasan, sebanyak 40.000 data atau 80% digunakan sebagai data latih, sedangkan 10.000 data atau 20% digunakan sebagai data uji. Pembagian dilakukan secara *stratified* agar proporsi kelas *Positive*, *Negative*, dan *Neutral* tetap terjaga pada kedua kelompok data.

Tabel 3. Pembagian Data Latih dan Data Uji

Jenis Data	Jumlah Data	Persentase
Data Latih	40.000	80%
Data Uji	10.000	20%
Total	**50.000**	**100%**

Sumber: Data Penelitian, 2026

Dari Tabel 3 yang sudah dipaparkan di atas, sebagian besar data digunakan pada tahap pelatihan agar model memperoleh informasi yang memadai untuk mempelajari distribusi fitur pada setiap kelas sentimen. Sebanyak 10.000 data lainnya dipisahkan sebagai data uji untuk mengukur kemampuan model ketika menghadapi ulasan yang tidak digunakan dalam proses pelatihan. Pemisahan tersebut penting agar penilaian performa tidak hanya menggambarkan kemampuan model mengenali data yang telah dipelajari sebelumnya.

Setelah proses pelatihan dilakukan terhadap 40.000 data, model diterapkan pada 10.000 data uji untuk menghasilkan prediksi sentimen. Hasil pengujian menunjukkan nilai *accuracy* sebesar 0,7926 atau 79,26%. Nilai tersebut berarti bahwa sekitar delapan dari setiap sepuluh ulasan pada data uji berhasil diklasifikasikan sesuai dengan label aktualnya. Dengan jumlah data uji yang mencapai 10.000 ulasan, capaian ini menunjukkan bahwa Multinomial Naive Bayes mampu mengenali sebagian besar pola sentimen yang terbentuk melalui representasi fitur TF-IDF.

Tabel 4. Ringkasan Hasil Pengujian Multinomial Naive Bayes

Komponen Pengujian	Hasil
Total Dataset	50.000
Data Latih	40.000
Data Uji	10.000
Rasio Pembagian	80:20
Metode Pembagian	*Stratified Split*
Jumlah Kelas	3
Accuracy	0,7926
Accuracy (%)	79,26%

Sumber: Data Penelitian, 2026

Dari Tabel 4 ini, model menghasilkan tingkat ketepatan klasifikasi sebesar 79,26% pada tiga kategori sentimen. Capaian tersebut menunjukkan bahwa pola kata yang terbentuk setelah tahap *preprocessing* dan pembobotan TF-IDF dapat dimanfaatkan oleh Multinomial Naive Bayes untuk membedakan kecenderungan opini pengguna. Dalam konteks penelitian ini, model tidak hanya menghadapi dua polaritas yang saling berlawanan, tetapi harus menentukan satu dari tiga kemungkinan kelas, yaitu positif, negatif, atau netral.

Kelas positif mencakup 53,11% dari keseluruhan dataset, sedangkan kelas netral hanya sebesar 12,74%. Perbedaan proporsi ini menyebabkan jumlah pola tekstual yang dipelajari model tidak sama pada setiap kategori. Model memperoleh lebih banyak contoh ulasan positif dibandingkan ulasan netral, yang dapat memengaruhi kekuatan estimasi probabilitas fitur pada masing-masing kelas.

Temuan ini sejalan dengan kecenderungan hasil penelitian terdahulu yang menunjukkan bahwa performa Naive Bayes pada data ulasan digital dapat berubah mengikuti karakteristik dataset dan objek yang dianalisis. Ramadhan dan Sugianto (2024) memperoleh akurasi 86% pada analisis sentimen aplikasi DANA menggunakan 2.000 data, sedangkan penerapan Naive Bayes pada objek dan konfigurasi data lainnya menunjukkan capaian yang berbeda. Perbedaan tersebut mengindikasikan bahwa jumlah data yang lebih besar tidak secara otomatis menghasilkan akurasi yang lebih tinggi karena variasi bahasa, komposisi kelas, proses praproses, dan tingkat kemiripan kosakata antarsentimen turut memengaruhi hasil klasifikasi.

Evaluasi Kinerja Model pada Setiap Kelas Sentimen

Evaluasi kinerja dilakukan untuk melihat kemampuan Multinomial Naive Bayes dalam mengenali masing-masing kategori sentimen. Pengukuran tidak hanya didasarkan pada *accuracy*, tetapi juga menggunakan *precision*, *recall*, dan *F1-score*. Pendekatan ini diperlukan

karena distribusi data tidak seimbang, dengan kelas positif sebagai kategori terbesar dan kelas netral sebagai kategori terkecil.

Tabel 5. Hasil Evaluasi Klasifikasi Multinomial Naive Bayes

Kelas	Precision	Recall	F1-Score	Support
Negative	0,71	0,80	0,75	3.415
Neutral	0,72	0,19	0,30	1.274
Positive	0,85	0,94	0,89	5.311
Accuracy			**0,79**	**10.000**
Macro Avg	**0,76**	**0,64**	**0,65**	**10.000**
Weighted Avg	**0,79**	**0,79**	**0,77**	**10.000**

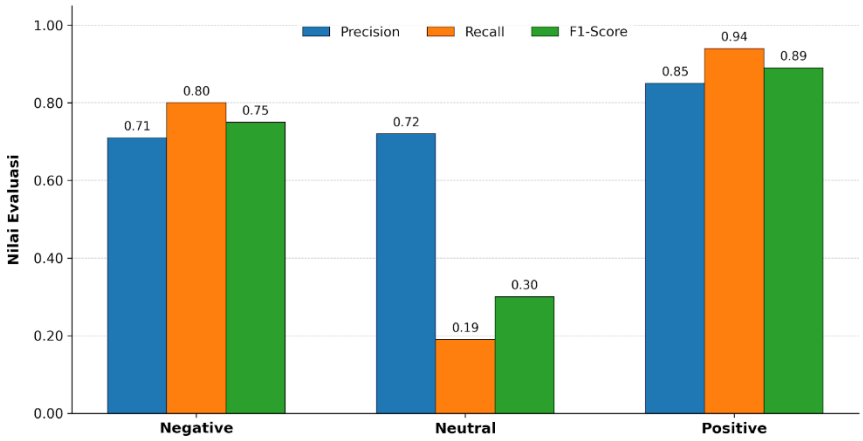
Sumber: Data Penelitian, 2026

Dari tabel di atas, kelas *Positive* menghasilkan performa terbaik dengan *precision* 0,85, *recall* 0,94, dan *F1-score* 0,89. Nilai *precision* menunjukkan bahwa prediksi positif memiliki tingkat ketepatan yang relatif tinggi, sedangkan *recall* 0,94 memperlihatkan bahwa sebagian besar ulasan yang benar-benar berlabel positif berhasil dikenali oleh model. Kombinasi keduanya menghasilkan *F1-score* tertinggi dibandingkan kelas lain. Capaian tersebut memperlihatkan bahwa pola tekstual pada ulasan positif dapat dipelajari dengan cukup kuat oleh Multinomial Naive Bayes.

Performa kelas *Negative* berada di bawah kelas positif, tetapi masih menunjukkan hasil yang relatif stabil. Nilai *precision* sebesar 0,71 menunjukkan bahwa tidak seluruh ulasan yang diprediksi negatif benar-benar berasal dari kelas negatif. Sementara itu, *recall* sebesar 0,80 memperlihatkan bahwa model berhasil mengenali sebagian besar ulasan negatif aktual. Nilai *F1-score* 0,75 menunjukkan keseimbangan performa yang lebih baik dibandingkan kelas netral, meskipun tingkat ketepatan prediksinya masih berada di bawah kategori positif.

Perbedaan paling jelas ditemukan pada kelas *Neutral*. Kategori ini memperoleh *precision* 0,72, tetapi nilai *recall* hanya 0,19 dan *F1-score* 0,30. Kombinasi tersebut menunjukkan pola yang menarik karena ketika model memberikan prediksi netral, hasilnya cukup sering tepat, tetapi sebagian besar ulasan yang sebenarnya netral justru tidak berhasil dikenali sebagai netral. Dengan kata lain, persoalan utama pada kelas ini bukan semata-mata banyaknya prediksi netral yang keliru, melainkan rendahnya kemampuan model dalam menemukan keseluruhan data netral yang sebenarnya tersedia.

Rendahnya *recall* kelas netral berkaitan dengan komposisi dataset yang tidak seimbang. Dari total 50.000 ulasan, kategori netral hanya berjumlah 6.372 data atau 12,74%, jauh di bawah kelas positif sebanyak 26.555 data dan kelas negatif sebanyak 17.073 data. Jumlah contoh yang lebih terbatas menyebabkan model memiliki representasi pola netral yang lebih sedikit selama proses pelatihan.



Gambar 2. Perbandingan Precision, Recall, dan F1-Score pada Setiap Kelas Sentimen

Dari gambar berikut, kelas positif memperlihatkan nilai yang konsisten tinggi pada ketiga metrik, sedangkan kelas negatif berada pada tingkat menengah dengan jarak *precision* dan *recall* yang tidak terlalu besar. Pola berbeda terlihat pada kelas netral. Nilai *precision* masih mencapai 0,72, tetapi *recall* turun tajam menjadi 0,19. Ketimpangan tersebut kemudian menekan *F1-score* hingga 0,30 dan menempatkan kelas netral sebagai kategori dengan performa terendah.

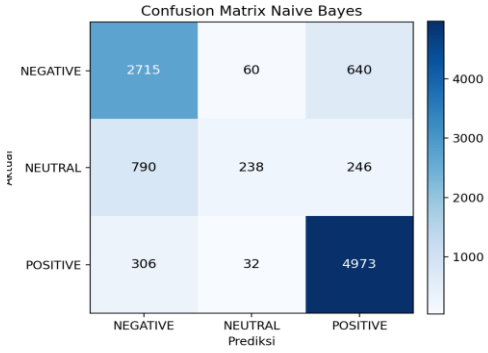
Nilai *macro average* memberikan gambaran tambahan mengenai ketimpangan tersebut. Rata-rata *precision* antarkelas mencapai 0,76, tetapi *macro recall* hanya 0,64 dan *macro F1-score* sebesar 0,65. Pada perhitungan ini, setiap kelas memperoleh bobot yang sama tanpa mempertimbangkan jumlah datanya. Nilai yang lebih rendah menunjukkan bahwa ketika seluruh kelas diperlakukan setara, kelemahan model dalam mengenali kategori netral memberikan pengaruh yang cukup besar terhadap performa keseluruhan.

Sebaliknya, *weighted average* menghasilkan *precision* 0,79, *recall* 0,79, dan *F1-score* 0,77. Nilai tersebut lebih tinggi dibandingkan *macro average* karena perhitungannya mempertimbangkan jumlah data pada masing-masing kelas. Kelas positif yang memiliki *support* terbesar sebanyak 5.311 data sekaligus menghasilkan performa terbaik, sehingga memberikan kontribusi lebih besar terhadap nilai rata-rata berbobot. Selisih antara *macro F1-score* sebesar 0,65 dan *weighted F1-score* sebesar 0,77 memperkuat indikasi bahwa performa model belum tersebar secara merata pada seluruh kategori sentimen.

Temuan tersebut menunjukkan bahwa akurasi 79,26% perlu dibaca secara hati-hati. Secara agregat, model mampu mengklasifikasikan mayoritas data uji dengan benar, tetapi keberhasilan tersebut lebih kuat pada kelas positif dan negatif daripada kelas netral. Pola ini sejalan dengan pembahasan mengenai ketidakseimbangan kelas dalam analisis sentimen, ketika kategori dengan representasi lebih kecil cenderung lebih sulit dikenali oleh model. Pada penelitian ini, kesulitan tersebut terlihat langsung dari *recall* netral sebesar 0,19, sementara kelas positif mencapai 0,94.

Analisis *Confusion Matrix*

Analisis *confusion matrix* dilakukan untuk mengidentifikasi pola prediksi benar dan kesalahan klasifikasi pada setiap kategori sentimen. Berbeda dengan nilai *accuracy* yang menggambarkan performa model secara keseluruhan, *confusion matrix* memperlihatkan arah perpindahan prediksi ketika suatu ulasan tidak berhasil dikenali sesuai kelas aktualnya. Pada penelitian ini, matriks klasifikasi dibentuk dari 10.000 data uji yang terdiri atas 3.415 ulasan *Negative*, 1.274 ulasan *Neutral*, dan 5.311 ulasan *Positive*.



Gambar 3. Confusion Matrix Hasil Klasifikasi Multinomial Naive Bayes

Tabel 6. Distribusi Hasil Prediksi Berdasarkan Confusion Matrix

Kelas Aktual	Prediksi Negative	Prediksi Neutral	Prediksi Positive	Total Aktual
Negative	**2.715**	60	640	3.415
Neutral	790	**238**	246	1.274
Positive	306	32	**4.973**	5.311
Total Prediksi	**3.811**	**330**	**5.859**	**10.000**

Sumber: Data Penelitian, 2026

Dari Tabel 6 dapat kita simpulkan bahwa sebanyak 2.715 dari 3.415 ulasan aktual *Negative* berhasil diklasifikasikan secara tepat. Kesalahan prediksi pada kelas ini terdiri atas 60 ulasan yang diarahkan ke *Neutral* dan 640 ulasan ke *Positive*. Dengan demikian, kesalahan terbesar pada sentimen negatif terjadi ketika model menganggap ulasan tersebut sebagai positif. Pola ini menunjukkan bahwa sebagian ulasan negatif memiliki kombinasi fitur yang memberikan probabilitas lebih kuat terhadap kelas positif dibandingkan kelas aktualnya.

Pada kelas *Neutral*, kemampuan model terlihat jauh lebih rendah. Dari 1.274 ulasan aktual netral, hanya 238 data yang berhasil dikenali dengan benar. Sebanyak 790 ulasan justru diprediksi sebagai *Negative*, sedangkan 246 lainnya diarahkan ke *Positive*. Artinya, 1.036 dari 1.274 data netral tidak memperoleh label yang sesuai. Temuan ini menjelaskan secara langsung rendahnya *recall* kelas netral sebesar 0,19 pada Subbab 4.3. Model hanya berhasil menemukan sebagian kecil ulasan netral, sementara mayoritas data pada kategori tersebut berpindah ke dua kelas lain.

Kesalahan paling dominan pada keseluruhan matriks adalah perpindahan *Neutral* ke *Negative* sebanyak 790 data. Jumlah ini lebih besar dibandingkan pola kesalahan lainnya, termasuk *Negative* ke *Positive* sebanyak 640 data, *Positive* ke *Negative* sebanyak 306 data, *Neutral* ke *Positive* sebanyak 246 data, *Negative* ke *Neutral* sebanyak 60 data, dan *Positive* ke *Neutral* sebanyak 32 data. Dominasi kesalahan *Neutral* → *Negative* memperlihatkan bahwa batas tekstual antara kedua kategori tersebut menjadi bagian paling sulit bagi model.

Sebaliknya, kelas *Positive* menunjukkan kemampuan pengenalan paling kuat. Dari 5.311 data aktual positif, sebanyak 4.973 ulasan berhasil diprediksi secara tepat. Kesalahan hanya terjadi pada 306 data yang diklasifikasikan sebagai *Negative* dan 32 data sebagai *Neutral*. Jumlah prediksi benar yang tinggi tersebut konsisten dengan *recall* positif sebesar 0,94 dan *F1-score* sebesar 0,89. Hasil ini memperlihatkan bahwa pola fitur pada sentimen positif lebih mudah dipisahkan dibandingkan kategori netral.

Jika dilihat dari seluruh 10.000 data uji, jumlah prediksi benar pada diagonal utama mencapai 7.926 data, yang diperoleh dari 2.715 prediksi benar *Negative*, 238 prediksi benar *Neutral*, dan 4.973 prediksi benar *Positive*. Sementara itu, 2.074 data lainnya berada di luar diagonal utama dan menunjukkan kesalahan klasifikasi. Proporsi tersebut menghasilkan *accuracy* sebesar 79,26%.

Distribusi hasil prediksi juga memperlihatkan kecenderungan model dalam menetapkan kelas. Dari 10.000 data uji, sebanyak 3.811 ulasan diprediksi sebagai *Negative*, 330 sebagai *Neutral*, dan 5.859 sebagai *Positive*. Jumlah prediksi netral jauh lebih rendah dibandingkan 1.274 data aktual pada kategori tersebut. Selisih ini menunjukkan bahwa model relatif jarang memberikan label *Neutral*, yang selaras dengan rendahnya kemampuan penemuan kelas tersebut. Sebaliknya, jumlah prediksi *Negative* dan *Positive* lebih besar daripada jumlah aktual masing-masing kelas karena sebagian ulasan netral dialihkan ke dua kategori tersebut.

Secara keseluruhan, *confusion matrix* menunjukkan bahwa kelemahan utama model tidak tersebar secara merata pada semua kelas. Multinomial Naive Bayes mampu mengenali sentimen positif dengan kuat dan sentimen negatif dengan performa yang cukup baik, tetapi

masih mengalami kesulitan dalam mempertahankan identitas kelas netral. Hasil ini memperlihatkan bahwa ketidakseimbangan distribusi data dan kemiripan kosakata antarsentimen menjadi aspek penting dalam interpretasi performa model. Dengan demikian, akurasi 79,26% perlu dipahami bersama pola kesalahan klasifikasi agar kemampuan model tidak dinilai hanya berdasarkan satu ukuran agregat.

Analisis Kata Kunci pada Setiap Kelas Sentimen

Analisis kata kunci dilakukan untuk mengidentifikasi karakteristik kosakata yang dominan pada ulasan pengguna aplikasi DANA. Tahap ini memberikan gambaran mengenai kata-kata yang paling sering muncul pada kelas *Positive*, *Negative*, dan *Neutral* setelah data melalui proses praproses teks. Frekuensi kemunculan kata digunakan sebagai dasar untuk melihat kecenderungan topik yang terbentuk pada setiap kategori sentimen.

Gambar 4. Visualisasi Kata Dominan Berdasarkan Kelas Sentimen

Dari Gambar 4, kelas positif didominasi oleh kosakata yang secara umum berkaitan dengan penilaian dan pengalaman penggunaan aplikasi, sedangkan kelas negatif lebih banyak memuat kata yang berhubungan dengan aplikasi, saldo, serta akses pengguna. Pada kelas netral, beberapa kata yang muncul juga ditemukan pada kategori negatif. Kondisi tersebut menunjukkan bahwa pemisahan sentimen tidak hanya bergantung pada keberadaan satu kata tertentu, tetapi dipengaruhi oleh kombinasi kata dan konteks ulasan secara keseluruhan.

Tabel 7. Kata dengan Frekuensi Tertinggi pada Setiap Kelas Sentimen

Peringkat	Sentimen Positive	Frekuensi	Sentimen Negative	Frekuensi	Sentimen Neutral	Frekuensi
1	sangat	7.296	dana	8.791	dana	3.158
2	bagus	5.266	bisa	5.296	bisa	2.149
3	mantap	3.683	aplikasi	3.173	akun	906
4	membantu	3.518	saldo	2.992	saldo	868
5	mudah	1.497	masuk	2.073	tolong	577

Sumber: Data Penelitian, 2026

Kelas *Positive* memperlihatkan dominasi kata “sangat” dengan frekuensi 7.296 kemunculan, diikuti “bagus” sebanyak 5.266, “mantap” sebanyak 3.683, “membantu” sebanyak 3.518, dan “mudah” sebanyak 1.497 kemunculan. Kata “bagus” dan “mantap”

menunjukkan bentuk penilaian positif yang relatif eksplisit, sedangkan "membantu" dan "mudah" berkaitan dengan manfaat serta kemudahan penggunaan aplikasi. Pola tersebut membantu menjelaskan kuatnya performa model pada kelas positif karena sebagian kosakata memiliki kecenderungan evaluatif yang cukup jelas.

Pada kelas *Negative*, kata "dana" menempati frekuensi tertinggi sebanyak 8.791 kemunculan, diikuti "bisa" sebanyak 5.296, "aplikasi" sebanyak 3.173, "saldo" sebanyak 2.992, dan "masuk" sebanyak 2.073 kemunculan. Dominasi kata "dana" dan "aplikasi" menunjukkan kuatnya penyebutan objek layanan dalam ulasan, sedangkan kemunculan "saldo" dan "masuk" memperlihatkan adanya perhatian pengguna terhadap aspek yang berkaitan dengan dana tersimpan dan akses aplikasi. Namun, kata-kata tersebut tidak selalu memiliki polaritas negatif jika berdiri sendiri. Arah sentimen baru terbentuk ketika kata muncul bersama unsur lain dalam satu ulasan.

Kelas *Neutral* menunjukkan pola yang berbeda sekaligus memiliki irisan dengan kelas negatif. Kata "dana" muncul sebanyak 3.158 kali dan "bisa" sebanyak 2.149 kali, diikuti "akun" sebanyak 906, "saldo" sebanyak 868, serta "tolong" sebanyak 577 kemunculan. Kata "saldo" terdapat pada lima kata teratas kelas negatif maupun netral, sementara kata "bisa" juga dominan pada kedua kategori. Kemiripan tersebut menunjukkan bahwa sebagian fitur tekstual tidak secara eksklusif menjadi penanda satu kelas.

Irisan kosakata tersebut membantu menjelaskan pola kesalahan pada *confusion matrix*. Sebanyak 790 ulasan aktual *Neutral* diprediksi sebagai *Negative*, yang merupakan kesalahan klasifikasi terbesar pada keseluruhan matriks. Ketika ulasan netral dan negatif sama-sama memuat kata seperti "bisa" dan "saldo", model perlu membedakan kelas berdasarkan kombinasi fitur lain yang menyertainya. Pada ulasan yang bersifat pertanyaan atau permintaan bantuan, kata "tolong" juga dapat muncul tanpa penilaian eksplisit, tetapi struktur semacam ini berpotensi memiliki kemiripan dengan keluhan pengguna. Hubungan antara kemiripan kosakata dan kesalahan klasifikasi netral ke negatif juga tampak pada hasil analisis penelitian.

Secara keseluruhan, analisis kata kunci menunjukkan bahwa performa klasifikasi tidak hanya berkaitan dengan jumlah data pada setiap kelas, tetapi juga dengan tingkat kekhasan kosakata yang membentuk sentimen. Kelas positif memiliki sejumlah kata evaluatif yang lebih eksplisit, sedangkan kelas negatif dan netral menunjukkan irisan pada kata-kata yang berkaitan dengan layanan, akun, saldo, dan akses aplikasi. Pola tersebut memperkuat hasil evaluasi sebelumnya bahwa tantangan utama Multinomial Naive Bayes terletak pada pembedaan ulasan yang tidak mengekspresikan polaritas secara tegas.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma Multinomial Naive Bayes dengan pembobotan TF-IDF mampu mengklasifikasikan 50.000 ulasan pengguna aplikasi DANA ke dalam tiga kategori sentimen dengan *accuracy* sebesar 79,26%. Kinerja terbaik diperoleh pada kelas *Positive* dengan *precision* 0,85, *recall* 0,94, dan *F1-score* 0,89, sedangkan kelas *Negative* menghasilkan *F1-score* 0,75. Keterbatasan utama ditemukan pada kelas *Neutral* dengan *recall* 0,19 dan *F1-score* 0,30. Hasil *confusion matrix* memperlihatkan bahwa dari 1.274 ulasan netral, hanya 238 data berhasil diklasifikasikan secara tepat, sementara 790 data diprediksi sebagai *Negative* dan 246 data sebagai *Positive*. Temuan tersebut menunjukkan bahwa ketidakseimbangan distribusi kelas dan kemiripan kosakata antarsentimen masih menjadi tantangan utama dalam proses klasifikasi, terutama pada kategori netral. Secara keseluruhan, Multinomial Naive Bayes cukup efektif dalam mengenali sentimen pengguna DANA, khususnya pada kelas positif dan negatif, tetapi masih memerlukan peningkatan untuk menghasilkan performa yang lebih seimbang pada seluruh kelas. Penelitian selanjutnya dapat mempertimbangkan teknik penyeimbangan data, optimasi parameter model, serta penggunaan representasi teks yang lebih kontekstual untuk meningkatkan kemampuan klasifikasi terhadap ulasan yang memiliki pola bahasa ambigu.

DAFTAR REFERENSI

Al Anshari, R., Alam, S., & Hafid T., M. (2023). Komparasi payment digital untuk analisis sentimen berdasarkan ulasan di Google Playstore menggunakan metode Support Vector Machine. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer, 2*(3), 118–128. https://doi.org/10.55123/storage.v2i3.2337

Ardiansyah, D., Aryanti, R., Fitriani, E., & Royadi, R. (2025). Analisis sentimen pengguna GoPay pada layanan keuangan digital dengan perbandingan Naïve Bayes dan SVM. *Jurnal Profitabilitas, 5*(2).

Hapsari, N. A., & Indriyanti, A. D. (2023). Analisis sentimen pada aplikasi dompet digital menggunakan algoritma Random Forest. *Journal of Emerging Information Systems and Business Intelligence, 4*(3), 186–192. https://doi.org/10.26740/jeisbi.v4i3.55696

Ibrahim, H. D., Heryana, N., & Primajaya, A. (2024). Analisis sentimen pengguna e-wallet OVO dan DANA pada Twitter dengan metode Naïve Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(5), 9858–9864.

Maharani, F. M. D., Hananto, A. L., Hilabi, S. S., Apriani, F. N., Hananto, A., & Huda, B. (2022). Perbandingan metode klasifikasi sentimen analisis penggunaan e-wallet menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. *METIK Jurnal, 6*(2), 97–103. https://doi.org/10.47002/metik.v6i2.372

Ramadhan, G. R., & Sugianto, C. A. (2024). Analisis sentimen ulasan aplikasi DANA di Google Play Store menggunakan algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(5). https://doi.org/10.36040/jati.v8i5.10732

Riskawati, R., Fatihanursari, F., Iin, I., & Rinaldi, A. R. (2024). Penerapan metode Naïve Bayes Classifier pada analisis sentimen aplikasi GoPay. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(1), 346–353. https://doi.org/10.36040/jati.v8i1.8699

Umair, M., & Susanto, E. R. (2024). Analisis sentimen ulasan pengguna pada aplikasi BRImo BRI menggunakan metode klasifikasi algoritma Naive Bayes. *Jurnal Media Informatika Budidarma, 8*(2), 1149. https://doi.org/10.30865/mib.v8i2.7381

Utami, I. W., & Dandi, M. K. (2026). Analisis sentimen pengguna aplikasi e-wallet DANA menggunakan Naive Bayes. *Journal of Data Analytics, Information, and Computer Science, 3*(2), 45–52. https://doi.org/10.70248/jdaics.v3i2.3651