Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Perbandingan Naïve Bayes, SVM, dan XGBoost dengan SMOTE untuk Analisis Sentimen Ulasan DANA dan OVO

Eko Wardianto[*], Putry Wahyu Setyaningsih

Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia
Email: [1,*]wardiant.eko@gmail.com, [2]putryws@mercubuana-yogya.ac.id
Email Penulis Korespondensi: wardiant.eko@gmail.com
Submitted: 19/05/2026; Accepted: 31/05/2026; Published: 31/05/2026

Abstrak–Ulasan pengguna aplikasi dompet digital DANA dan OVO mengandung informasi penting mengenai kepuasan dan permasalahan layanan, namun volume data yang besar dan ketidakseimbangan distribusi kelas sentimen menjadi tantangan utama dalam analisis otomatis. Penelitian ini bertujuan membandingkan performa Naïve Bayes, Support Vector Machine (SVM), dan XGBoost dalam klasifikasi sentimen ulasan pengguna, sekaligus mengevaluasi efektivitas SMOTE dalam menangani ketidakseimbangan data pada masing-masing algoritma. Dataset terdiri dari 4.987 ulasan yang diklasifikasikan ke dalam tiga kelas sentimen: positif, negatif, dan netral. Tahapan penelitian meliputi pra-pemrosesan teks, pembobotan fitur TF-IDF, serta evaluasi menggunakan *confusion matrix*. Penerapan SMOTE menunjukkan dampak signifikan pada Naïve Bayes, sementara SVM dan XGBoost tidak menunjukkan perubahan performa yang berarti sehingga model finalnya menggunakan data asli. Hasil pengujian menunjukkan SVM sebagai algoritma terbaik dengan akurasi 85,14% dan F1-score makro 77,60%, diikuti XGBoost dengan akurasi 82,32%. Analisis per aplikasi mengungkapkan bahwa model pada DANA mencapai akurasi lebih tinggi (87,55%) dibandingkan OVO (82,13%), yang mengindikasikan perbedaan karakteristik linguistik dan distribusi sentimen antar platform. Penelitian ini berkontribusi dalam memberikan perbandingan sistematis antar algoritma dengan mempertimbangkan penanganan imbalance data secara selektif pada konteks ulasan aplikasi keuangan digital.

Kata Kunci: Analisis Sentimen; Naïve Bayes; Support Vector Machine; XGBoost; TF-IDF; SMOTE; DANA; OVO

Abstract–User reviews of the DANA and OVO digital wallet applications provide valuable insights into user satisfaction and service-related issues. However, the large volume of review data and the imbalanced distribution of sentiment classes pose significant challenges for automated sentiment analysis. This study aims to compare the performance of Naïve Bayes, Support Vector Machine (SVM), and XGBoost for sentiment classification of user reviews, while evaluating the effectiveness of the Synthetic Minority Over-sampling Technique (SMOTE) in handling class imbalance for each algorithm. The dataset consists of 4,987 user reviews classified into three sentiment categories: positive, negative, and neutral. The research process includes text preprocessing, TF-IDF feature extraction, model training, and performance evaluation using a confusion matrix and standard classification metrics. The experimental results indicate that SMOTE significantly improves the performance of the Naïve Bayes classifier, whereas it provides little to no performance improvement for SVM and XGBoost; therefore, the final SVM and XGBoost models are trained using the original dataset. Among the evaluated algorithms, SVM achieves the best overall performance with an accuracy of 85.14% and a macro F1-score of 77.60%, followed by XGBoost with an accuracy of 82.32%. Furthermore, application-specific analysis reveals that the model achieves higher accuracy on DANA reviews (87.55%) than on OVO reviews (82.13%), suggesting differences in linguistic characteristics and sentiment distributions across the two platforms. This study provides a systematic comparison of machine learning algorithms for Indonesian digital wallet sentiment analysis and demonstrates that selective application of data balancing techniques can improve classification performance without necessarily benefiting all algorithms.

Keywords: Sentiment Analysis; Naïve Bayes; Support Vector Machine; XGBoost; TF-IDF; SMOTE; DANA; OVO

1. PENDAHULUAN

Aplikasi dompet digital seperti DANA dan OVO telah menjadi bagian integral dalam transaksi keuangan sehari-hari masyarakat Indonesia. DANA tercatat berhasil menarik 20 juta pengguna sejak diperkenalkan pada November 2018 hingga pertengahan 2019, dengan rata-rata 1,5 juta transaksi setiap harinya [1], sementara OVO sebagai salah satu platform dompet digital terkemuka di Indonesia juga terus mengalami pertumbuhan pengguna yang signifikan [2]. Pertumbuhan ini mendorong meningkatnya volume ulasan pengguna pada platform distribusi aplikasi seperti Google Play Store dan App Store, yang berisi berbagai tanggapan terhadap layanan aplikasi, baik berupa sentimen positif, negatif, maupun netral. Namun, tingginya volume ulasan menjadikan proses analisis secara manual tidak efektif dan memakan waktu, sehingga diperlukan pendekatan otomatis untuk memperoleh informasi yang relevan dari data tersebut.

Analisis sentimen berbasis *Natural Language Processing* (NLP) dan *machine learning* menjadi solusi untuk mengklasifikasikan opini pengguna secara sistematis. Melalui pendekatan ini, data teks yang tidak terstruktur dapat diolah untuk mengidentifikasi persepsi pengguna serta menemukan pola permasalahan yang sering muncul. Selain itu, analisis sentimen juga dapat dikembangkan menjadi analisis berbasis aspek untuk memberikan pemahaman yang lebih spesifik terhadap bagian layanan yang menjadi perhatian pengguna, seperti kecepatan sistem, keamanan, dan kemudahan penggunaan.

Sejumlah penelitian terdahulu telah menerapkan pendekatan komparasi algoritma untuk analisis sentimen ulasan pengguna. Penelitian pada ulasan aplikasi OVO menunjukkan bahwa metode Support Vector Machine

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

(SVM) memiliki performa lebih baik dibandingkan Naïve Bayes dengan akurasi masing-masing sekitar 89% dan 86% [3]. Namun, penelitian tersebut hanya membandingkan dua algoritma dan belum mempertimbangkan keseimbangan performa antar kelas sentimen, khususnya pada kelas minoritas.

Penelitian lain yang membandingkan Naïve Bayes, SVM, dan Random Forest menunjukkan bahwa SVM dan Random Forest cenderung memberikan akurasi lebih tinggi dibandingkan Naïve Bayes pada aplikasi Shoppee di Google Play Store [4]. Meskipun demikian, penelitian tersebut lebih berfokus pada perbandingan nilai akurasi tanpa analisis lebih lanjut terkait distribusi kelas atau potensi ketidakseimbangan data yang dapat mempengaruhi hasil klasifikasi.

Selanjutnya, penelitian pada aplikasi DANA yang membandingkan XGBoost dan SVM menunjukkan bahwa kedua metode mampu mencapai akurasi tinggi hingga sekitar 93% setelah dilakukan penanganan ketidakseimbangan data menggunakan teknik SMOTE [5]. Temuan ini menegaskan bahwa permasalahan *imbalanced data* merupakan faktor penting dalam analisis sentimen. Namun, penelitian tersebut hanya berfokus pada satu aplikasi dan belum melakukan perbandingan lintas aplikasi dalam satu kerangka analisis yang sama.

Pendekatan analisis sentimen berbasis aspek juga telah diterapkan dalam penelitian lain, seperti pada aplikasi Flip yang menggunakan Naïve Bayes untuk mengklasifikasikan sentimen berdasarkan aspek layanan [6]. Meskipun memberikan wawasan yang lebih spesifik, penelitian ini belum mengombinasikan analisis aspek dengan perbandingan algoritma secara komprehensif.

Selain itu, penelitian yang membandingkan sentimen pengguna DANA dan OVO menunjukkan adanya perbedaan karakteristik distribusi sentimen pada masing-masing aplikasi [7]. Namun, penelitian tersebut belum mengkaji performa model klasifikasi secara terpisah untuk masing-masing aplikasi maupun pada data gabungan, sehingga belum memberikan gambaran menyeluruh mengenai perbedaan performa model.

Berdasarkan kajian penelitian terdahulu, masih terdapat beberapa *research gap*. Pertama, sebagian besar penelitian hanya berfokus pada satu aplikasi sehingga belum memberikan perbandingan langsung antar aplikasi dalam satu eksperimen yang konsisten. Kedua, penelitian komparasi algoritma umumnya masih berfokus pada nilai akurasi tanpa mempertimbangkan keseimbangan klasifikasi antar kelas sentimen. Ketiga, evaluasi efektivitas SMOTE pada penelitian sebelumnya umumnya diterapkan secara seragam pada semua model tanpa menganalisis konsistensi dampaknya secara komparatif antar algoritma, sehingga belum memberikan gambaran apakah SMOTE memberikan manfaat yang konsisten pada semua jenis algoritma klasifikasi. Keempat, penelitian sebelumnya masih terbatas dalam menggabungkan analisis sentimen berbasis aspek, penanganan imbalanced data, dan evaluasi performa model secara terpisah maupun gabungan dalam satu studi.

Oleh karena itu, penelitian ini menganalisis sentimen ulasan pengguna aplikasi DANA dan OVO dengan membandingkan algoritma Naïve Bayes, Support Vector Machine (SVM), dan XGBoost. Pemilihan XGBoost sebagai algoritma pembanding didasarkan pada kemampuannya sebagai model ensemble berbasis boosting yang terbukti efektif dalam menangani data tidak seimbang dan menghasilkan performa tinggi pada berbagai tugas klasifikasi teks [5]. Evaluasi tidak hanya dilakukan berdasarkan akurasi, tetapi juga mempertimbangkan keseimbangan performa antar kelas sentimen melalui *precision*, *recall*, dan *F1-score*. Selain itu, penelitian ini mengevaluasi efektivitas SMOTE secara komparatif pada ketiga algoritma untuk mengidentifikasi konsistensi dampaknya dalam menangani ketidakseimbangan data.

Penelitian juga dilakukan pada tiga skenario dataset, yaitu data DANA, OVO, dan data gabungan, guna menganalisis perbedaan karakteristik sentimen pada masing-masing aplikasi. Selain klasifikasi sentimen positif, negatif, dan netral, penelitian ini menerapkan analisis berbasis aspek menggunakan metode *Latent Dirichlet Allocation* (LDA) guna mengidentifikasi topik layanan yang paling sering dibahas pengguna. sekaligus mengintegrasikannya dengan distribusi sentimen per aspek untuk menghasilkan wawasan yang lebih spesifik dan dapat ditindaklanjuti.

Kontribusi utama penelitian ini terletak pada integrasi beberapa pendekatan dalam satu kerangka eksperimen yang konsisten, yaitu: (1) perbandingan tiga algoritma klasifikasi, (2) evaluasi dampak SMOTE secara komparatif pada ketiga algoritma, (3) analisis performa model secara terpisah pada dataset DANA, OVO, dan gabungan, serta (4) analisis sentimen berbasis aspek menggunakan LDA yang diintegrasikan dengan distribusi sentimen per aspek. Dengan pendekatan tersebut, penelitian ini tidak hanya menghasilkan klasifikasi sentimen secara umum, tetapi juga memberikan pemahaman yang lebih spesifik mengenai aspek layanan yang mempengaruhi persepsi pengguna. Hasil penelitian diharapkan dapat menjadi referensi dalam pemilihan algoritma analisis sentimen serta membantu pengembang aplikasi DANA dan OVO dalam memahami kebutuhan pengguna berdasarkan ulasan aplikasi.

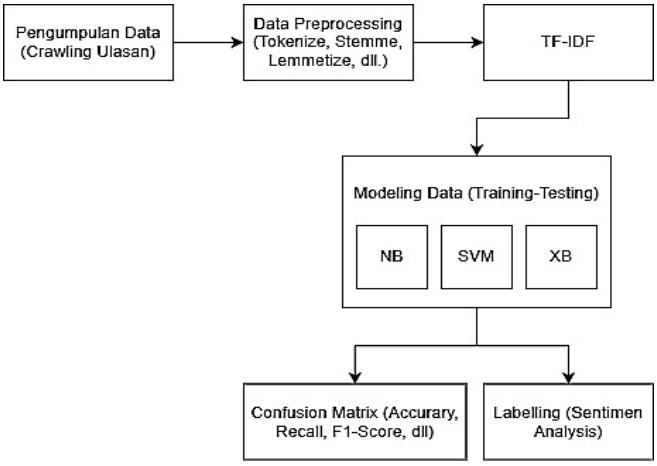
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna aplikasi DANA dan OVO serta mengevaluasi performa beberapa algoritma klasifikasi dalam kondisi data yang tidak seimbang. Tahapan penelitian disusun secara terstruktur agar proses analisis sesuai dengan tujuan penelitian. Alur penelitian pada

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Gambar 1 meliputi pengumpulan data, pra-pemrosesan, representasi fitur, penanganan ketidakseimbangan data, pemodelan, dan evaluasi model[2], [8].



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, tahapan penelitian dimulai dari pengumpulan data ulasan pengguna aplikasi DANA dan OVO dari Google Play Store dan App Store menggunakan teknik *web scraping*. Total dataset yang digunakan sebanyak 4.987 ulasan yang mencerminkan pengalaman pengguna secara langsung terhadap kedua aplikasi [1], [2]. Penggunaan dua platform bertujuan untuk meningkatkan keberagaman data serta mengurangi bias sumber tunggal.

Setelah data terkumpul, dilakukan tahap pelabelan sentimen menggunakan pendekatan *lexicon-based* dengan memanfaatkan kamus kata positif dan negatif berbahasa Indonesia. Pendekatan ini dipilih karena tidak memerlukan data berlabel sebelumnya dan telah terbukti efektif menghasilkan label yang akurat pada penelitian analisis sentimen ulasan aplikasi berbahasa Indonesia, bahkan mengungguli metode pelabelan berbasis rating dalam beberapa konteks [9]. Penentuan label dilakukan berdasarkan akumulasi skor kata pada setiap ulasan, di mana ulasan dengan skor positif dikategorikan sebagai sentimen positif, skor negatif sebagai sentimen negatif, dan skor nol sebagai sentimen netral.

Data hasil crawling umumnya masih mengandung *noise* berupa karakter khusus, tanda baca, angka, dan kata tidak baku, sehingga perlu dilakukan tahap pra-pemrosesan. Tahap ini mencakup *case folding*, tokenisasi, penghapusan *stopwords*, *stemming*, dan normalisasi kata untuk meningkatkan kualitas data sebelum tahap klasifikasi [10], [11]. Teks ulasan yang telah dibersihkan kemudian diubah ke dalam bentuk representasi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency), yang memberikan bobot lebih tinggi pada kata yang dianggap penting dalam suatu dokumen sehingga dapat membantu meningkatkan performa klasifikasi sentimen [6].

Dataset hasil representasi TF-IDF selanjutnya dibagi menjadi data latih dan data uji dengan rasio 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi performa model pada data baru. Teknik pembagian data ini umum digunakan dalam penelitian klasifikasi teks [12], [13]. Sebelum tahap pemodelan, dilakukan eksperimen penanganan ketidakseimbangan data menggunakan teknik *oversampling* SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan representasi kelas minoritas pada data latih [5], [14]. Eksperimen SMOTE dilakukan secara terpisah pada masing-masing algoritma untuk mengevaluasi konsistensi dampaknya secara komparatif.

Tahap pemodelan menggunakan tiga algoritma, yaitu Naïve Bayes, *Support Vector Machine* (SVM), dan XGBoost. Pemodelan dilakukan untuk membandingkan performa masing-masing algoritma dalam mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral[5]. Eksperimen pada penelitian ini dilakukan menggunakan tiga skenario dataset, yaitu data gabungan DANA dan OVO, data DANA, serta data OVO. Pengujian pada data gabungan bertujuan untuk mengevaluasi kemampuan model dalam mengenali pola sentimen dari dua aplikasi secara bersamaan, sementara pengujian pada masing-masing aplikasi dilakukan untuk mengidentifikasi perbedaan karakteristik sentimen pengguna dan melihat konsistensi performa model pada setiap dataset. Model yang telah dibangun dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk memungkinkan analisis trade-off antara akurasi dan keseimbangan klasifikasi antar kelas [14], [15].

2.2 Metode yang Digunakan

2.2.1 Analisis Sentimen

Analisis sentimen merupakan proses komputasional untuk mengidentifikasi dan mengekstraksi opini subjektif dari teks, yang bertujuan mengelompokkan kecenderungan ulasan pengguna ke dalam kelas positif, negatif, dan netral[7], [16]. Dalam penelitian ini, analisis sentimen diterapkan menggunakan pendekatan teks untuk

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

menangkap makna aktual dari ulasan pengguna, bukan hanya berdasarkan rating numerik yang diberikan pengguna[17]. Pendekatan ini memungkinkan identifikasi sentimen yang lebih akurat karena rating tidak selalu mencerminkan isi ulasan secara konsisten.

2.2.2 Naïve Bayes

Naïve Bayes digunakan sebagai model baseline karena memiliki proses komputasi yang sederhana dan cukup efektif pada klasifikasi teks berdimensi tinggi. Algoritma ini memiliki keunggulan dalam kecepatan komputasi dan efisiensi pada data teks berdimensi tinggi[18], [19]. Namun, asumsi independensi yang digunakan tidak selalu sesuai dengan karakteristik data teks, sehingga performanya dapat lebih rendah dibandingkan metode lain [11]. Meskipun demikian, Naïve Bayes tetap banyak digunakan sebagai baseline dalam analisis sentimen.

2.2.3 Support Vector Machine

SVM dipilih karena mampu memisahkan data antar kelas menggunakan hyperplane optimal sehingga sering memberikan performa baik pada analisis teks. SVM sangat cocok menangani data berdimensi tinggi termasuk teks. Beberapa penelitian menunjukkan bahwa SVM mampu menghasilkan performa klasifikasi yang lebih stabil dibandingkan Naïve Bayes [3], [14].

2.2.4 XGBoost

XGBoost digunakan untuk mengevaluasi kemampuan metode ensemble yang menggabungkan beberapa model untuk meningkatkan akurasi prediksi. XGBoost dikenal memiliki performa tinggi karena mampu menangani pola data yang kompleks sehingga sering digunakan dalam berbagai tugas klasifikasi. Penelitian sebelumnya menunjukkan bahwa XGBoost dapat memberikan tingkat akurasi yang baik, terutama pada dataset berskala besar [5], [14].

2.2.5 Latent Dirichlet Allocation (LDA)

LDA digunakan dalam penelitian ini sebagai metode topic modeling untuk mengidentifikasi aspek-aspek layanan yang paling sering dibahas dalam ulasan pengguna. LDA bekerja dengan mengelompokkan kata-kata yang sering muncul bersamaan ke dalam topik-topik laten, sehingga memungkinkan ekstraksi informasi yang lebih spesifik dari data teks yang tidak terstruktur [20]. Penerapan LDA dalam analisis sentimen berbasis aspek pada ulasan aplikasi terbukti efektif dalam mengungkap pola opini pengguna dan memberikan dasar bagi pengembang dalam memprioritaskan perbaikan layanan [20], [21].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data penelitian berasal dari ulasan pengguna aplikasi dompet digital DANA dan OVO yang diambil dari Google Play Store serta App Store, yang merupakan sumber utama ulasan pengguna terhadap aplikasi mobile. Pengumpulan data dilakukan menggunakan teknik web scraping berbasis Python secara mandiri, karena dataset yang tersedia pada sumber publik umumnya memiliki jumlah data yang terbatas atau tidak mencerminkan kondisi ulasan terbaru.

Data dikumpulkan selama periode tahun 2025 dengan rentang waktu yang berbeda pada masing-masing platform. Ulasan aplikasi DANA pada App Store diperoleh dari rentang 17 September 2025 hingga 27 Oktober 2025, sedangkan ulasan OVO pada App Store mencakup periode 26 Maret 2025 hingga 21 Oktober 2025. Pada Google Play Store, data OVO diperoleh dari periode 1 November 2025 hingga 25 November 2025, sedangkan data DANA berasal dari periode 23 November 2025 hingga 25 November 2025. Dari proses tersebut diperoleh 4.987 ulasan, dan setelah tahap pembersihan data meliputi penghapusan duplikasi, data kosong, serta ulasan yang tidak relevan jumlah data yang digunakan dalam penelitian menjadi 4.980 ulasan.

Berdasarkan hasil pengelompokan, dataset dipisahkan berdasarkan aplikasi untuk mendukung analisis yang lebih spesifik pada masing-masing platform. Distribusi jumlah ulasan untuk masing-masing aplikasi ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Data

Aplikasi	Jumlah Data
DANA	2497
OVO	2490
Total	4987

Berdasarkan Tabel 1, jumlah data antara aplikasi DANA dan OVO relatif seimbang dengan selisih hanya tujuh ulasan. Keseimbangan distribusi data ini penting untuk memastikan bahwa perbandingan performa model antar aplikasi tidak dipengaruhi oleh perbedaan jumlah data yang signifikan, sehingga hasil analisis sentimen dan evaluasi model diharapkan lebih objektif dan representatif.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

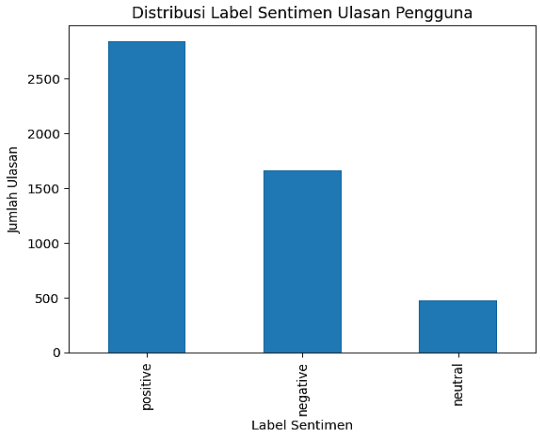
3.2 Pelabelan Sentimen

Pelabelan sentimen pada penelitian ini dilakukan menggunakan pendekatan *lexicon-based*, yaitu dengan memanfaatkan kamus kata positif dan negatif berbahasa Indonesia. Pendekatan ini dipilih karena tidak memerlukan data berlabel sebelumnya dan telah terbukti efektif digunakan dalam penelitian analisis sentimen pada ulasan aplikasi berbahasa Indonesia [9]. Penentuan label dilakukan berdasarkan akumulasi skor kata pada setiap ulasan. Jika skor akhir bernilai positif, maka ulasan dikategorikan sebagai sentimen positif. Jika skor bernilai negatif, maka dikategorikan sebagai sentimen negatif. Sedangkan jika skor bernilai nol, maka dikategorikan sebagai sentimen netral. Distribusi hasil pelabelan sentimen ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Sentimen

Sentimen	Jumlah Data
Positif	2842
Negatif	1659
Netral	479
Total	4980

Berdasarkan Tabel 2, distribusi sentimen menunjukkan ketidakseimbangan yang cukup signifikan antar kelas. Kelas positif mendominasi dengan 2.842 ulasan (57%), diikuti kelas negatif sebanyak 1.659 ulasan (33%), sedangkan kelas netral hanya berjumlah 479 ulasan (10%). Ketidakseimbangan ini berpotensi mempengaruhi kemampuan model dalam mengenali kelas minoritas, khususnya sentimen netral. Distribusi sentimen tersebut divisualisasikan pada Gambar 2 untuk memperjelas proporsi setiap kelas.



Gambar 2. Distribusi Sentimen

Ketidakseimbangan distribusi kelas sentimen menjadi salah satu tantangan utama dalam penelitian ini. Untuk mengatasi hal tersebut, dilakukan eksperimen menggunakan teknik *oversampling* SMOTE (*Synthetic Minority Over-sampling Technique*) guna meningkatkan representasi kelas minoritas, khususnya sentimen netral. Hasil eksperimen SMOTE pada masing-masing algoritma dibahas secara rinci pada sub-bab 3.7, 3.8, dan 3.9. Berdasarkan hasil evaluasi tersebut, penerapan SMOTE menunjukkan dampak yang berbeda-beda pada setiap algoritma, memberikan perubahan signifikan pada Naïve Bayes, namun tidak memberikan peningkatan performa yang berarti pada SVM dan XGBoost. Oleh karena itu, keputusan penggunaan SMOTE pada model final didasarkan pada hasil evaluasi empiris terhadap masing-masing algoritma, bukan diterapkan secara seragam pada seluruh model.

3.3 Pelabelan Aspek

Pelabelan aspek dilakukan menggunakan metode topic modeling dengan algoritma *Latent Dirichlet Allocation* (LDA). Metode ini digunakan untuk mengelompokkan kata-kata dalam ulasan berdasarkan pola kemunculan kata yang sering muncul bersamaan, sehingga menghasilkan beberapa topik utama yang merepresentasikan aspek layanan yang dibahas pengguna. Sebelum proses pemodelan, data ulasan terlebih dahulu melalui tahap *preprocessing* meliputi pembersihan teks, penghapusan *stopwords*, dan tokenisasi. Setelah itu, teks diubah ke dalam bentuk representasi *Bag of Words* untuk digunakan dalam proses LDA. Berdasarkan hasil pemodelan, diperoleh empat aspek utama dengan distribusi jumlah ulasan yang ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Aspek

Aspek	Jumlah
Transaksi	2460
UI/UX & Akses Akun	690

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Aspek	Jumlah
Kenerja Sistem	671
Fitur & Biaya	1159

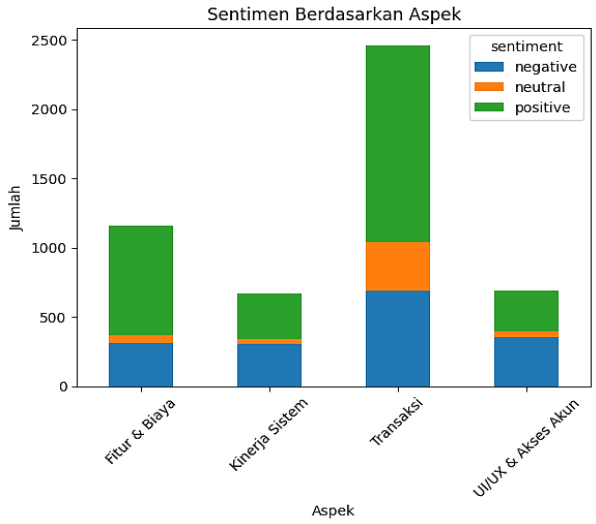
Berdasarkan Tabel 3, aspek Transaksi mendominasi dengan 2.460 ulasan (49%), diikuti aspek Fitur & Biaya sebanyak 1.159 ulasan (23%), UI/UX & Akses Akun sebanyak 690 ulasan (14%), dan Kinerja Sistem sebanyak 671 ulasan (13%). Dominasi aspek Transaksi mengindikasikan bahwa pengalaman pengguna terkait proses transaksi seperti pengiriman saldo, top up, dan kegagalan transaksi merupakan perhatian utama pengguna aplikasi DANA dan OVO.

Untuk memperoleh pemahaman yang lebih mendalam, dilakukan analisis sentimen per aspek guna mengidentifikasi kecenderungan sentimen pada masing-masing aspek layanan. Hasil analisis cross aspek dan sentimen ditunjukkan pada Tabel 4.

Tabel 4. Distribusi Sentimen per Aspek

Aspek	Negatif	Netral	Positif	Total
Transaksi	692 (28%)	345 (14%)	1.423 (58%)	2.460
UI/UX & Akses Akun	352 (51%)	44 (6%)	294 (43%)	690
Kenerja Sistem	302 (45%)	38 (6%)	331 (49%)	671
Fitur & Biaya	313 (27%)	52 (4%)	794 (69%)	1.159

Distribusi sentimen per aspek divisualisasikan pada Gambar 3 untuk mempermudah interpretasi hasil analisis secara visual.



Gambar 3. Sentimen Berdasarkan Aspek

Berdasarkan Tabel 4 dan Gambar 3, setiap aspek menunjukkan karakteristik distribusi sentimen yang berbeda-beda. Aspek Transaksi dan Fitur & Biaya didominasi oleh sentimen positif dengan masing-masing 58% dan 69%, mengindikasikan bahwa pengguna secara umum merasa puas dengan fitur transaksi dan layanan biaya yang ditawarkan kedua aplikasi. Meskipun demikian, aspek Transaksi tetap mencatat jumlah ulasan negatif tertinggi secara absolut sebanyak 692 ulasan, yang berkaitan dengan permasalahan kegagalan transaksi dan keterlambatan pemrosesan.

Sebaliknya, aspek UI/UX & Akses Akun menjadi satu-satunya aspek yang didominasi sentimen negatif dengan proporsi 51%, jauh melampaui sentimen positifnya yang hanya 43%. Temuan ini mengindikasikan bahwa pengguna menghadapi permasalahan yang cukup signifikan terkait antarmuka aplikasi, proses login, dan verifikasi akun. Aspek Kinerja Sistem juga menunjukkan distribusi yang hampir berimbang antara sentimen negatif (45%) dan positif (49%), yang mengindikasikan bahwa stabilitas dan kecepatan sistem masih menjadi perhatian sebagian pengguna. Temuan ini memberikan gambaran yang lebih spesifik mengenai aspek layanan yang perlu mendapat prioritas perbaikan oleh pengembang aplikasi DANA dan OVO, khususnya pada aspek UI/UX & Akses Akun dan Kinerja Sistem.

3.4 Hasil Pra-pemrosesan

Tahap pra-pemrosesan dilakukan melalui *case folding*, *cleaning*, tokenisasi, serta penghapusan *stopwords*. Selain itu, dilakukan juga normalisasi kata untuk mengatasi penggunaan bahasa tidak baku yang sering ditemukan dalam ulasan pengguna. Contoh hasil perubahan teks setelah proses pra-pemrosesan ditunjukkan pada Tabel 5 berikut.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Tabel 5. Contoh Hasil Pra-pemrosesan

Teks Asli	Hasil Preprocessing
tiba tiba hilang premium ketika mau upgrade ti...	hilang premium upgrade ktp no akun aman data a...
Aplikasi penipuan yang sangat jahat tidak bisa...	penipuan jahat menjaga data pengguna ktp diamb..
dipaksa update untuk isi jenis pekerjaan dan i...	dipaksa update isi jenis pekerjaan income isi ...
pelayanannya kurang tidak jelas tidak ada sele...	pelayanannya selesai pemindahan akun gagal pro...
Gmn si kk balikin dong uang saya ilang tb? 120...	gmn si kk balikin uang ilang tb tolong dikemba...

Berdasarkan Tabel 5, proses pra-pemrosesan berhasil mereduksi *noise* pada teks ulasan dengan mempertahankan kata-kata yang memiliki makna relevan terhadap sentimen dan aspek layanan. Kata-kata tidak baku seperti "gmn" dan "tb" dinormalisasi, sementara kata yang tidak memiliki kontribusi semantik dihapus melalui proses penghapusan *stopwords*. Hasil pra-pemrosesan ini selanjutnya digunakan sebagai input pada tahap representasi fitur menggunakan TF-IDF. Untuk memberikan gambaran mengenai kata-kata yang paling sering muncul dalam keseluruhan data ulasan setelah pra-pemrosesan, dilakukan visualisasi menggunakan *WordCloud* pada Gambar 4.

Gambar 4. WordCloud Seluruh Data

Berdasarkan Gambar 4, kata-kata seperti "*saldo*", "*transaksi*", dan "*aplikasi*" muncul dengan frekuensi paling tinggi dalam keseluruhan ulasan. Temuan ini konsisten dengan hasil analisis aspek pada sub-bab 3.3, di mana aspek Transaksi merupakan aspek dengan jumlah ulasan terbanyak (2.460 ulasan atau 49% dari total data). Selain itu, kemunculan kata "*akun*" dan "*verifikasi*" yang cukup dominan juga selaras dengan temuan bahwa aspek UI/UX & Akses Akun memiliki proporsi sentimen negatif tertinggi (51%), mengindikasikan bahwa permasalahan akses akun menjadi salah satu keluhan utama pengguna. Dengan demikian, visualisasi *WordCloud* ini memperkuat temuan analisis aspek dan memberikan gambaran awal mengenai topik-topik dominan yang dibahas pengguna sebelum tahap klasifikasi dilakukan.

3.5 Representasi TF-IDF

Teks ulasan pengguna yang telah melalui tahap pra-pemrosesan selanjutnya diubah ke dalam bentuk representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini dipilih karena mampu memberikan bobot lebih tinggi pada kata yang relevan dan informatif dalam suatu dokumen, namun jarang muncul di dokumen lain, sehingga dapat meningkatkan kualitas fitur dalam proses klasifikasi dibandingkan metode sederhana seperti *Bag of Words* yang hanya menghitung frekuensi kemunculan kata tanpa mempertimbangkan distribusinya di seluruh dokumen.

Pada penelitian ini, TF-IDF diimplementasikan menggunakan parameter *max_features* sebesar 5.000, yang berarti hanya 5.000 kata dengan bobot TF-IDF tertinggi yang dipertahankan sebagai fitur. Selain itu, digunakan *ngram_range* (1,2) yang mencakup kombinasi *unigram* (satu kata) dan *bigram* (dua kata berurutan), sehingga model dapat menangkap konteks makna yang lebih luas dari sekadar kata tunggal. Sebagai contoh, frasa seperti "*tidak bisa*" atau "*gagal transaksi*" dapat direpresentasikan sebagai satu fitur bigram yang lebih informatif dibandingkan kata "*tidak*" dan "*bisa*" secara terpisah. Hasil representasi TF-IDF menghasilkan matriks fitur dengan dimensi 4.980 × 5.000, yang berarti setiap ulasan direpresentasikan dalam bentuk vektor dengan 5.000 dimensi fitur. Matriks fitur ini selanjutnya digunakan sebagai input pada tahap pembagian data dan pemodelan klasifikasi.

3.6 Pembagian Data (Train-Test Split)

Data hasil representasi TF-IDF selanjutnya dibagi menjadi data latih (*training*) dan data uji (*testing*) menggunakan rasio 80:20. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur performa model pada data yang belum pernah dilihat sebelumnya. Rasio 80:20 dipilih karena merupakan proporsi yang umum digunakan dalam penelitian klasifikasi teks, di mana porsi data latih yang lebih besar memungkinkan model untuk mempelajari pola yang lebih beragam, sementara porsi data uji yang memadai tetap menjamin validitas evaluasi [12], [13]. Hasil pembagian data ditunjukkan pada Tabel 6.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Tabel 6. Pembagian Data

Jenis Data	Jumlah Data	Persentase
Data Latih	3984	80%
Data Uji	996	20%
Total	4980	100%

Berdasarkan Tabel 6, dari total 4.980 ulasan yang tersedia, sebanyak 3.984 ulasan digunakan sebagai data latih dan 996 ulasan digunakan sebagai data uji. Pembagian ini memastikan bahwa model dilatih pada data yang cukup representatif, sekaligus dievaluasi pada data yang sepenuhnya independen dari proses pelatihan. Data latih dan data uji yang telah terbentuk ini selanjutnya digunakan sebagai input pada tahap pemodelan menggunakan algoritma Naïve Bayes, SVM, dan XGBoost.

3.7 Penerapan Algoritma Naïve Bayes

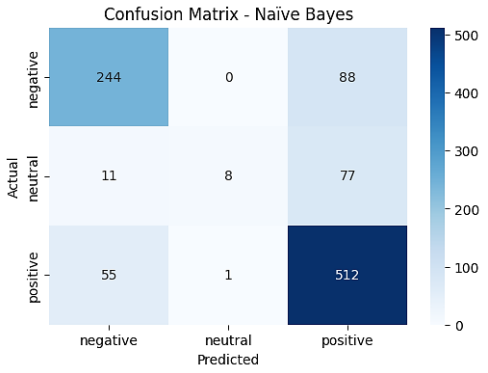
Algoritma Naïve Bayes diterapkan sebagai model baseline dalam penelitian ini untuk mengklasifikasikan sentimen ulasan pengguna aplikasi DANA dan OVO. Metode ini dipilih karena memiliki proses komputasi yang cepat dan mampu bekerja secara efektif pada data teks berdimensi tinggi seperti hasil representasi TF-IDF. Pemodelan dilakukan pada data latih sebanyak 3.984 ulasan dan dievaluasi pada data uji sebanyak 996 ulasan. Hasil evaluasi model Naïve Bayes ditunjukkan pada Tabel 7.

Tabel 7. Hasil Evaluasi Naïve Bayes

Kelas	Precision	Recall	F1-Score	Support
Negative	0.79	0.73	0.76	332
Neutral	0.89	0.08	0.15	96
Positive	0.76	0.90	0.82	568
Accuracy			0.77	996

Berdasarkan Tabel 7, model Naïve Bayes memperoleh akurasi sebesar 77%. Model menunjukkan performa yang cukup baik pada kelas positif dengan nilai recall sebesar 0.90, yang berarti model mampu mengenali 90% ulasan positif dengan benar. Namun demikian, terdapat anomali yang signifikan pada kelas netral, di mana nilai precision mencapai 0.89 namun recall hanya sebesar 0.08. Kondisi ini mengindikasikan bahwa model sangat jarang memprediksi suatu ulasan sebagai kelas netral, namun ketika model memberikan prediksi netral, prediksi tersebut hampir selalu benar. Fenomena ini terjadi karena jumlah data kelas netral yang sangat sedikit (479 ulasan atau 10% dari total data), sehingga model cenderung mengklasifikasikan ulasan netral ke dalam kelas mayoritas, yaitu positif atau negatif. Distribusi hasil prediksi model secara lebih rinci ditunjukkan pada Gambar 5.

Berdasarkan Gambar 5, confusion matrix mengonfirmasi bahwa sebagian besar kesalahan klasifikasi terjadi pada kelas netral, yang mayoritas diprediksi sebagai kelas positif atau negatif. Hal ini memperkuat bahwa ketidakseimbangan distribusi data merupakan faktor utama yang mempengaruhi performa model Naïve Bayes, khususnya pada kelas minoritas. Untuk mengatasi permasalahan tersebut, dilakukan eksperimen tambahan dengan menerapkan teknik *oversampling* SMOTE (*Synthetic Minority Over-sampling Technique*) pada data latih sebelum proses pemodelan. Hasil evaluasi model Naïve Bayes dengan SMOTE ditunjukkan pada Tabel 8.



Gambar 5. Confusion Matrix - Naïve Bayes

Tabel 8. Hasil Evaluasi Naïve Bayes + SMOTE

Kelas	Precision	Recall	F1-Score	Support
Negative	0.60	0.89	0.72	332
Neutral	0.41	0.29	0.34	96
Positive	0.92	0.70	0.79	568

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Kelas	Precision	Recall	F1-Score	Support
Accuracy			0.72	996

Berdasarkan Tabel 8, penerapan SMOTE pada model Naïve Bayes menghasilkan akurasi sebesar 72%, yang mengalami penurunan sebesar 5% dibandingkan model tanpa SMOTE. Meskipun demikian, terdapat peningkatan yang signifikan pada kemampuan model dalam mengenali kelas netral, di mana nilai recall kelas netral meningkat drastis dari 0.08 menjadi 0.29. Peningkatan ini menunjukkan bahwa SMOTE berhasil memperbaiki representasi kelas minoritas dalam proses pelatihan model, sehingga model lebih mampu mengenali pola ulasan netral yang sebelumnya hampir seluruhnya salah diklasifikasikan. Namun demikian, peningkatan recall kelas netral ini diikuti dengan penurunan precision pada kelas negatif dari 0.79 menjadi 0.60, yang mengindikasikan adanya *trade-off* antara peningkatan sensitivitas terhadap kelas minoritas dengan peningkatan kesalahan klasifikasi pada kelas lainnya.

Secara keseluruhan, model Naïve Bayes tanpa SMOTE menghasilkan akurasi yang lebih tinggi namun kurang mampu mengenali kelas netral, sedangkan model dengan SMOTE menunjukkan keseimbangan klasifikasi yang lebih baik pada kelas minoritas dengan konsekuensi penurunan akurasi secara keseluruhan. Perbedaan dampak SMOTE ini menjadi dasar evaluasi lebih lanjut pada algoritma SVM dan XGBoost untuk melihat konsistensi pengaruh SMOTE pada model yang berbeda, sebagaimana dibahas pada sub-bab 3.8 dan 3.9.

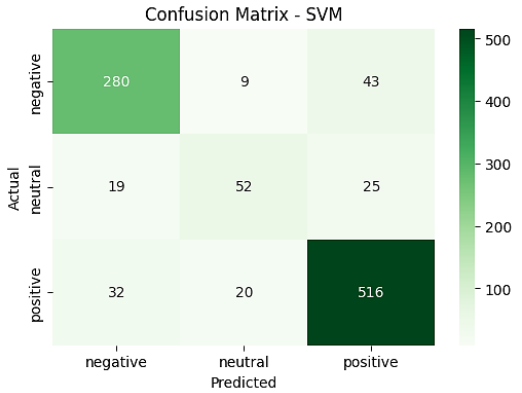
3.8 Penerapan Algoritma Support Vector Machine (SVM)

Model *Support Vector Machine* (SVM) diterapkan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi DANA dan OVO. SVM dipilih karena kemampuannya dalam menemukan hyperplane optimal yang memisahkan antar kelas pada ruang fitur berdimensi tinggi, sehingga sangat sesuai untuk data teks hasil representasi TF-IDF yang memiliki 5.000 dimensi fitur. Pemodelan dilakukan pada data latih sebanyak 3.984 ulasan dan dievaluasi pada data uji sebanyak 996 ulasan. Hasil evaluasi model SVM ditunjukkan pada Tabel 9.

Tabel 9. Hasil Evaluasi SVM

Kelas	Precision	Recall	F1-Score	Support
Negative	0.85	0.84	0.84	332
Neutral	0.64	0.54	0.59	96
Positive	0.88	0.91	0.90	568
Accuracy			0.85	996

Berdasarkan Tabel 9, model SVM memperoleh akurasi sebesar 85%, yang merupakan peningkatan signifikan sebesar 8% dibandingkan model Naïve Bayes sebelumnya. Peningkatan ini terjadi pada seluruh kelas sentimen, termasuk kelas netral yang sebelumnya menjadi kelemahan utama Naïve Bayes. Nilai recall kelas netral meningkat drastis dari 0.08 pada Naïve Bayes menjadi 0.54 pada SVM, yang menunjukkan bahwa SVM jauh lebih mampu mengenali pola ulasan netral meskipun jumlah datanya terbatas. Kemampuan ini didukung oleh mekanisme *margin maximization* pada SVM yang memungkinkan model untuk menemukan batas keputusan yang lebih optimal antar kelas, bahkan pada kondisi data yang tidak seimbang. Distribusi hasil prediksi model secara lebih rinci ditunjukkan pada Gambar 6.



Gambar 6. Confusion Matrix - SVM

Berdasarkan Gambar 6, confusion matrix menunjukkan bahwa kesalahan klasifikasi pada model SVM relatif lebih sedikit dibandingkan model Naïve Bayes pada seluruh kelas sentimen. Model ini mampu mengurangi kesalahan prediksi pada kelas netral secara signifikan, meskipun masih terdapat beberapa ulasan netral yang salah diklasifikasikan ke dalam kelas positif atau negatif. Untuk mengevaluasi konsistensi dampak SMOTE pada SVM, dilakukan eksperimen tambahan dengan menerapkan SMOTE pada data latih sebelum proses pemodelan. Hasil evaluasi model SVM dengan SMOTE ditunjukkan pada Tabel 10.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

Tabel 10. Hasil Evaluasi SVM + SMOTE

Kelas	Precision	Recall	F1-Score	Support
Negative	0.83	0.86	0.84	332
Neutral	0.57	0.60	0.59	96
Positive	0.91	0.88	0.89	568
Accuracy			0.85	996

Berdasarkan Tabel 10, penerapan SMOTE pada model SVM menghasilkan akurasi yang sama sebesar 85%, yang berarti SMOTE tidak memberikan dampak negatif terhadap akurasi keseluruhan. Terdapat peningkatan pada recall kelas netral dari 0.54 menjadi 0.60, yang mengindikasikan bahwa SMOTE sedikit meningkatkan kemampuan model dalam mengenali ulasan netral. Namun demikian, peningkatan ini tidak sebesar dampak SMOTE pada Naïve Bayes, di mana recall kelas netral meningkat dari 0.08 menjadi 0.29. Hal ini menunjukkan bahwa SVM secara inheren sudah lebih mampu menangani ketidakseimbangan data dibandingkan Naïve Bayes, sehingga kontribusi SMOTE terhadap performa SVM relatif lebih kecil. Berdasarkan hasil evaluasi tersebut, model SVM tanpa SMOTE dipilih sebagai model final karena menghasilkan akurasi dan keseimbangan klasifikasi yang sudah optimal tanpa memerlukan proses *oversampling* tambahan. Evaluasi dampak SMOTE selanjutnya dilakukan pada algoritma XGBoost untuk melengkapi analisis komparatif, sebagaimana dibahas pada sub-bab 3.9.

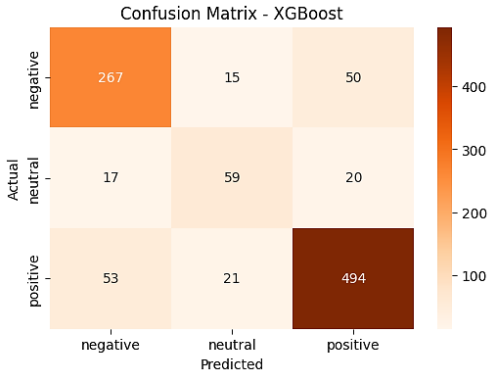
3.9 Penerapan Algoritma XGBoost

Algoritma *Extreme Gradient Boosting* (XGBoost) diterapkan sebagai model *ensemble* berbasis *boosting* untuk mengklasifikasikan sentimen ulasan pengguna aplikasi DANA dan OVO. XGBoost dipilih karena kemampuannya dalam menangani pola data yang kompleks melalui kombinasi beberapa model lemah (*weak learners*) secara bertahap, di mana setiap model berikutnya berfokus pada kesalahan prediksi model sebelumnya sehingga menghasilkan model akhir yang lebih akurat. Pemodelan dilakukan pada data latih sebanyak 3.984 ulasan dan dievaluasi pada data uji sebanyak 996 ulasan. Hasil evaluasi model XGBoost ditunjukkan pada Tabel 11.

Tabel 11. Hasil Evaluasi XGBoost

Kelas	Precision	Recall	F1-Score	Support
Negative	0.79	0.80	0.80	332
Neutral	0.62	0.61	0.62	96
Positive	0.88	0.87	0.87	568
Accuracy			0.82	996

Berdasarkan Tabel 11, model XGBoost memperoleh akurasi sebesar 82%, yang berada di antara performa Naïve Bayes (77%) dan SVM (85%). Dibandingkan Naïve Bayes, XGBoost menunjukkan peningkatan yang signifikan pada kelas netral, di mana recall meningkat dari 0.08 menjadi 0.61. Hal ini menunjukkan bahwa mekanisme *boosting* pada XGBoost mampu memperbaiki kesalahan klasifikasi pada kelas minoritas secara bertahap melalui proses iterasi. Meskipun demikian, performa XGBoost masih berada di bawah SVM, yang dapat dijelaskan secara teoritis bahwa data teks berdimensi tinggi hasil representasi TF-IDF cenderung lebih *linearly separable*, sehingga SVM dengan kemampuan *hyperplane* optimalnya lebih unggul pada karakteristik data seperti ini. Distribusi hasil prediksi model secara lebih rinci ditunjukkan pada Gambar 7.



Gambar 7. Confusion Matrix - XGBoost

Berdasarkan Gambar 7, confusion matrix menunjukkan bahwa model XGBoost mampu mengurangi kesalahan klasifikasi pada kelas netral dibandingkan Naïve Bayes, serta mempertahankan performa yang baik pada kelas positif dan negatif. Dibandingkan SVM, kesalahan klasifikasi XGBoost pada kelas netral sedikit lebih tinggi, yang konsisten dengan perbedaan nilai recall kelas netral antara kedua model (0.61 vs 0.54 pada SVM

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

tanpa SMOTE). Untuk mengevaluasi konsistensi dampak SMOTE pada XGBoost, dilakukan eksperimen tambahan dengan menerapkan SMOTE pada data latih sebelum proses pemodelan. Hasil evaluasi model XGBoost dengan SMOTE ditunjukkan pada Tabel 12.

Tabel 12. Hasil Evaluasi XGBoost + SMOTE

Kelas	Precision	Recall	F1-Score	Support
Negative	0.77	0.82	0.79	332
Neutral	0.51	0.68	0.58	96
Positive	0.90	0.82	0.86	568
Accuracy			0.81	996

Berdasarkan Tabel 12, penerapan SMOTE pada model XGBoost menghasilkan akurasi sebesar 81%, yang mengalami penurunan tipis sebesar 1% dibandingkan model tanpa SMOTE. Meskipun demikian, terdapat peningkatan yang cukup signifikan pada recall kelas netral dari 0.61 menjadi 0.68, yang menunjukkan bahwa SMOTE berhasil meningkatkan kemampuan model dalam mengenali ulasan netral. Jika dibandingkan dengan dampak SMOTE pada ketiga algoritma, terlihat pola yang konsisten SMOTE memberikan dampak paling besar pada Naïve Bayes (recall netral 0.08 → 0.29), dampak sedang pada XGBoost (0.61 → 0.68), dan dampak paling kecil pada SVM (0.54 → 0.60). Pola ini mengindikasikan bahwa algoritma yang secara inheren lebih mampu menangani ketidakseimbangan data akan mendapat manfaat yang lebih kecil dari penerapan SMOTE. Berdasarkan hasil evaluasi tersebut, model XGBoost tanpa SMOTE dipilih sebagai model final karena penurunan akurasi akibat SMOTE tidak sebanding dengan peningkatan recall kelas netral yang dihasilkan. Hasil evaluasi ketiga algoritma beserta eksperimen SMOTE selanjutnya dibandingkan secara komprehensif pada sub-bab 3.10.

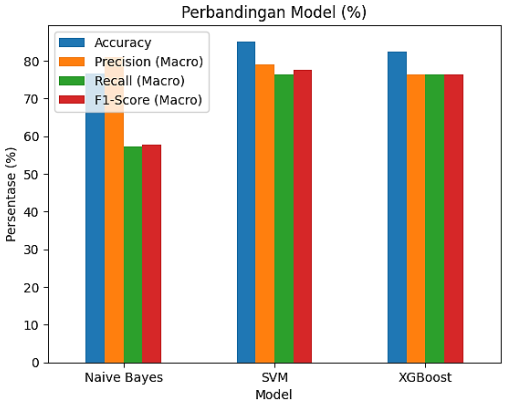
3.10 Perbandingan Kinerja Model

Tahap ini membandingkan performa keenam model klasifikasi yang telah dibangun, yaitu Naïve Bayes, Naïve Bayes + SMOTE, SVM, SVM + SMOTE, XGBoost, dan XGBoost + SMOTE pada data ulasan pengguna DANA dan OVO. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *dan F1-score* (macro average) untuk melihat keseimbangan performa antar kelas secara menyeluruh. Hasil perbandingan kinerja seluruh model ditunjukkan pada Tabel 13.

Tabel 13. Perbandingan Kinerja Model

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	76.71%	81.08%	57.32%	57.83%
Naïve Bayes + SMOTE	72.00%	64.00%	63.00%	62.00%
SVM	85.14%	79.05%	76.45%	77.60%
SVM + SMOTE	85.00%	77.00%	78.00%	77.00%
XGBoost	82.33%	76.31%	76.28%	76.29%
XGBoost + SMOTE	81.00%	73.00%	77.00%	75.00%

Berdasarkan Tabel 13 dan visualisasi pada Gambar 8, setiap model menunjukkan karakteristik performa yang berbeda-beda. Naïve Bayes menghasilkan akurasi sebesar 76.71%, namun memiliki nilai recall makro yang sangat rendah (57.32%), yang mencerminkan ketidakmampuan model dalam mengenali kelas netral secara konsisten. Penerapan SMOTE pada Naïve Bayes menurunkan akurasi menjadi 72.00%, namun meningkatkan recall makro menjadi 63.00%, yang menunjukkan adanya *trade-off* antara akurasi keseluruhan dan keseimbangan klasifikasi antar kelas. Untuk memperjelas perbandingan performa antar model, digunakan visualisasi pada Gambar 8.



Gambar 8. Perbandingan Model NB SVM XGBoost (%)

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

SVM menghasilkan performa terbaik dengan akurasi 85.14% dan F1-score makro 77.60%, yang menunjukkan keseimbangan klasifikasi paling konsisten di antara seluruh model yang diuji. Penerapan SMOTE pada SVM menghasilkan akurasi yang hampir identik (85.00%) dengan peningkatan recall makro tipis dari 76.45% menjadi 78.00%, yang mengindikasikan bahwa SVM secara inheren sudah mampu menangani ketidakseimbangan data dengan baik tanpa memerlukan *oversampling* tambahan. XGBoost menempati posisi kedua dengan akurasi 82.33% dan F1-score makro 76.29%, menunjukkan performa yang kompetitif dan distribusi prediksi yang relatif seimbang antar kelas. Penerapan SMOTE pada XGBoost sedikit menurunkan akurasi menjadi 81.00%, namun meningkatkan recall makro dari 76.28% menjadi 77.00%, yang menunjukkan pola serupa dengan SVM meskipun dengan magnitude yang berbeda.

Secara keseluruhan, terdapat pola yang konsisten dari hasil eksperimen SMOTE pada ketiga algoritma dampak SMOTE paling signifikan terjadi pada Naïve Bayes, diikuti XGBoost, dan paling kecil pada SVM. Pola ini mengonfirmasi bahwa algoritma yang secara inheren lebih robust terhadap ketidakseimbangan data akan memperoleh manfaat yang lebih kecil dari penerapan SMOTE. Temuan ini juga menjawab pertanyaan mengenai konsistensi dampak SMOTE lintas algoritma, yang menjadi salah satu kontribusi utama penelitian ini.

Jika dibandingkan dengan penelitian sejenis, hasil SVM pada penelitian ini (85.14%) sedikit di bawah hasil yang dilaporkan pada penelitian ulasan aplikasi OVO yang mencapai akurasi sekitar 89% [3]. Perbedaan ini dapat dijelaskan oleh perbedaan karakteristik dataset, di mana penelitian ini menggunakan data gabungan dari dua aplikasi dengan distribusi sentimen yang lebih beragam dan periode pengumpulan yang lebih baru. Sementara itu, penelitian pada aplikasi DANA yang menggunakan XGBoost dan SVM dengan SMOTE melaporkan akurasi hingga 93% [5], namun penelitian tersebut hanya berfokus pada satu aplikasi sehingga distribusi datanya lebih homogen dibandingkan dataset gabungan yang digunakan dalam penelitian ini. Hasil perbandingan per aplikasi secara lebih rinci dibahas pada sub-bab 3.11.

3.11 Analisis Perbandingan Model Berdasarkan Aplikasi

Tahap ini membandingkan performa model klasifikasi SVM pada tiga skenario dataset, yaitu data ulasan aplikasi DANA, OVO, dan gabungan keduanya. Tujuan analisis ini adalah untuk memahami bagaimana perbedaan karakteristik distribusi sentimen pada masing-masing aplikasi berdampak terhadap performa model klasifikasi. Sebelum membahas hasil evaluasi model, perlu dipahami terlebih dahulu perbedaan distribusi sentimen antara kedua aplikasi. Dataset DANA memiliki distribusi sentimen yang tidak seimbang namun cenderung searah, dengan dominasi sentimen positif sebesar 67% (1.674 ulasan), diikuti sentimen negatif 23% (567 ulasan), dan netral 10% (249 ulasan). Sebaliknya, dataset OVO menunjukkan distribusi yang jauh lebih berimbang antara sentimen positif (47%, 1.168 ulasan) dan negatif (44%, 1.092 ulasan), dengan netral sebesar 9% (230 ulasan). Perbedaan karakteristik distribusi inilah yang menjadi faktor utama perbedaan performa model pada kedua aplikasi. Hasil evaluasi model SVM pada dataset DANA ditunjukkan pada Tabel 14.

Tabel 14. Hasil Evaluasi SVM - DANA

Kelas	Precision	Recall	F1-Score	Support
Negative	0.81	0.83	0.82	113
Neutral	0.77	0.66	0.71	50
Positive	0.91	0.92	0.92	335
Accuracy			0.88	498

Berdasarkan Tabel 14, model SVM pada dataset DANA menghasilkan akurasi sebesar 88%, dengan performa yang konsisten pada seluruh kelas sentimen. Kelas positif menunjukkan performa terbaik dengan F1-score 0.92, yang sejalan dengan dominasi sentimen positif pada dataset ini. Kelas netral mencapai recall 0.66, yang menunjukkan bahwa model cukup mampu mengenali ulasan netral meskipun jumlahnya terbatas. Tingginya proporsi sentimen positif pada dataset DANA membuat pola klasifikasi lebih mudah dikenali oleh model, karena model dapat mempelajari fitur-fitur yang khas dari kelas mayoritas secara lebih optimal. Hasil evaluasi model SVM pada dataset OVO ditunjukkan pada Tabel 15.

Tabel 15. Hasil Evaluasi SVM - OVO

Kelas	Precision	Recall	F1-Score	Support
Negative	0.83	0.85	0.84	218
Neutral	0.72	0.39	0.51	46
Positive	0.83	0.88	0.85	234
Accuracy			0.82	498

Berdasarkan Tabel 15, model SVM pada dataset OVO menghasilkan akurasi sebesar 82%, lebih rendah dibandingkan dataset DANA. Penurunan performa ini paling terlihat pada kelas netral, di mana recall turun drastis dari 0.66 pada DANA menjadi hanya 0.39 pada OVO. Kondisi ini dapat dijelaskan oleh distribusi OVO yang hampir berimbang antara sentimen positif (47%) dan negatif (44%), sehingga batas keputusan antar kelas menjadi lebih sulit ditentukan oleh model. Keberadaan ulasan negatif dalam jumlah besar yang berdekatan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

secara fitur dengan ulasan netral menyebabkan model lebih sering salah mengklasifikasikan ulasan netral sebagai negatif. Meskipun demikian, performa pada kelas positif dan negatif tetap cukup baik dengan F1-score masing-masing 0.85 dan 0.84, yang menunjukkan bahwa model masih mampu membedakan kedua kelas mayoritas dengan konsisten. Perbandingan performa SVM pada ketiga skenario dataset ditunjukkan pada Tabel 16.

Tabel 16. Perbandingan Performa SVM pada Dataset DANA, OVO, dan Gabungan

Dataset	Accuracy	Precision	Recall	F1-Score
DANA	87.55%	82.97%	80.47%	81.58%
OVO	82.12%	79.10%	70.68%	73.24%
Gabungan	85.14%	79.04%	76.44%	77.60%

Berdasarkan Tabel 16, terdapat perbedaan performa yang signifikan antara ketiga skenario dataset. Dataset DANA menghasilkan akurasi tertinggi sebesar 87.55% dengan F1-score makro 81.58%, sedangkan dataset OVO menghasilkan akurasi 82.12% dengan F1-score makro 73.24%. Perbedaan F1-score makro sebesar 8.34% antara DANA dan OVO mencerminkan bahwa distribusi sentimen yang lebih berimbang pada OVO menghadirkan tantangan klasifikasi yang jauh lebih tinggi dibandingkan DANA. Dataset gabungan menghasilkan performa di antara keduanya dengan akurasi 85.14% dan F1-score makro 77.60%, yang menunjukkan bahwa penggabungan data dari dua aplikasi dengan karakteristik berbeda menghasilkan kompleksitas yang lebih tinggi namun tetap dapat ditangani oleh model dengan baik.

Secara keseluruhan, hasil analisis per aplikasi ini mengonfirmasi bahwa performa model klasifikasi tidak hanya ditentukan oleh algoritma yang digunakan, tetapi juga sangat dipengaruhi oleh karakteristik distribusi sentimen pada dataset. Temuan ini memiliki implikasi praktis bagi pengembang aplikasi DANA dan OVO tingginya proporsi sentimen negatif pada OVO (44%) mengindikasikan adanya permasalahan layanan yang lebih mendesak untuk ditangani dibandingkan DANA (23%), khususnya pada aspek UI/UX & Akses Akun yang telah teridentifikasi sebagai aspek dengan sentimen negatif dominan pada sub-bab 3.3.

4. KESIMPULAN

Penelitian ini membandingkan performa algoritma Naïve Bayes, Support Vector Machine (SVM), dan XGBoost dalam klasifikasi sentimen ulasan pengguna aplikasi dompet digital DANA dan OVO, sekaligus mengevaluasi efektivitas SMOTE dalam menangani ketidakseimbangan data pada masing-masing algoritma. Hasil pengujian menunjukkan bahwa SVM merupakan algoritma terbaik dengan akurasi 85.14% dan F1-score makro 77.60% pada data gabungan, diikuti XGBoost dengan akurasi 82.33% dan Naïve Bayes dengan akurasi 76.71%. Evaluasi SMOTE pada ketiga algoritma mengungkapkan pola yang konsisten dampak SMOTE paling signifikan terjadi pada Naïve Bayes dengan peningkatan recall kelas netral dari 0.08 menjadi 0.29, diikuti XGBoost (0.61 → 0.68), dan paling kecil pada SVM (0.54 → 0.60), yang mengindikasikan bahwa algoritma yang secara inheren lebih robust terhadap ketidakseimbangan data akan memperoleh manfaat lebih kecil dari penerapan SMOTE. Analisis per aplikasi menunjukkan bahwa model pada dataset DANA menghasilkan akurasi lebih tinggi (87.55%) dibandingkan OVO (82.12%), yang disebabkan oleh perbedaan karakteristik distribusi sentimen DANA didominasi sentimen positif (67%) sementara OVO memiliki distribusi yang lebih berimbang antara positif (47%) dan negatif (44%), sehingga menghadirkan tantangan klasifikasi yang lebih tinggi. Analisis aspek menggunakan LDA mengidentifikasi empat topik utama layanan, yaitu Transaksi, Fitur & Biaya, Kinerja Sistem, dan UI/UX & Akses Akun, di mana aspek UI/UX & Akses Akun menjadi satu-satunya aspek dengan sentimen negatif dominan (51%), mengindikasikan bahwa permasalahan antarmuka aplikasi dan proses verifikasi akun merupakan prioritas utama yang perlu mendapat perhatian dari pengembang DANA dan OVO. Meskipun demikian, penelitian ini memiliki keterbatasan pada proses pelabelan otomatis berbasis leksikon yang berpotensi menghasilkan label yang kurang sesuai dengan makna sebenarnya, serta kerentanan terhadap bahasa tidak baku, singkatan, dan sarkasme yang umum ditemukan dalam ulasan pengguna, sehingga penelitian selanjutnya dapat mempertimbangkan penggunaan model berbasis transformer seperti IndoBERT yang lebih mampu memahami konteks bahasa Indonesia informal secara lebih akurat.

REFERENCES

[1] D. Toresa, S. Rico Francsisco Sitorus, I. Muzdalifah, F. Wiza, and R. Syelly, "Analisis Sentimen Terhadap Ulasan Penggunaan Dompet Digital Dana Mengunakan Metode Klasifikasi Support Vector Machine," *Technologica*, vol. 3, no. 2, pp. 64–74, Jul. 2024, doi: 10.55043/technologica.v3i2.163.

[2] M. Ramdan Adi Surya, M. Martanto, and U. Hayati, "Analisis Sentimen Ulasan Pengguna Ovo Menggunakan Algoritma Naive Bayes Pada Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2780–2786, May 2024, doi: 10.36040/jati.v8i3.8739.

[3] A. Lowell, A. Lowell, K. Candra, E. Indra, and S. Informasi, "Perbandingan Metode Support Vector Machine (SVM) Dan Naive Bayes Pada Analisis Sentimen Ulasan Aplikasi OVO," *Jurnal Media Informatika [JUMIN]*, vol. 6, no. 2, pp. 896–905, 2025, doi: 10.55338/jumin.v6i2.5134.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 136-149
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.10002

[4] P. Anggraini and W. Winarsih, "Komparasi Naïve Bayes, Vector Machine, Dan Random Forest Dalam Analisis Sentimen Aplikasi Shopee Di Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4451–4457, May 2025, doi: 10.36040/jati.v9i3.13675.

[5] M. J. Setiawan and V. R. S. Nastiti, "DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 3, pp. 337–345, Nov. 2024, doi: 10.32736/sisfokom.v13i3.2239.

[6] S. A. Helmayanti, F. Hamami, and R. Y. Fa'rifah, "Penerapan Algoritma Tf-Idf Dan Naïve Bayes Untuk Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Flip Pada Google Play Store," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1822–1834, Sep. 2023, doi: 10.35870/jimik.v4i3.415.

[7] H. Dafa Ibrahim, N. Heryana, and A. Primajaya, "Analisis Sentimen Pengguna E-Wallet Ovo Dan Dana Pada Twitter Dengan Metode Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 9858–9864, Sep. 2024, doi: 10.36040/jati.v8i5.10738.

[8] I. S. Widianto, Y. R. Ramadhan, Y. R. Ramadhan, M. A. Komara, and M. A. Komara, "Analisis Sentimen E-Wallet Gopay, Shopeepay, Dan Ovo Menggunakan Algoritma Naive Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3s1.5277.

[9] H. Firda, P. Putra, N. R. Oktadini, P. E. Sevtiyuni, and A. Meiriza, "Comparison of Rating-based and Inset Lexicon-based Labeling in Sentiment Analysis using SVM (Case Study: GoBiz Application Reviews on Google Play Store)," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 2, p. 516, Mar. 2025, doi: 10.32520/stmsi.v14i2.4795.

[10] R. Rahmat, A. Rahim, and A. Arbansyah, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Alibaba Di Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3050–3057, Mar. 2025, doi: 10.36040/jati.v9i2.13269.

[11] N. Fibriyanti Arminda, N. Sulistiyowati, and T. Nur Padilah, "Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1817–1822, Nov. 2023, doi: 10.36040/jati.v7i3.7012.

[12] N. Nurzaman, N. Suarna, and W. Prihartono, "Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 967–974, Mar. 2024, doi: 10.36040/jati.v8i1.8708.

[13] D. S. Putri and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine," *Jurnal Ilmiah Informatika*, vol. 11, no. 1, pp. 32–40, 2023, Accessed: Jun. 03, 2026. [Online]. Available: https://ejournal.upbatam.ac.id/index.php/jif/article/download/6611/3098

[14] O. Oktafia and R. S. A. Nugroho, "Comparison Of Support Vector Machine(SVM), Xgboost And Random Forest For Sentiment Analysis Of Bumble App User Comments," *Proxies : Jurnal Informatika*, vol. 6, no. 1, pp. 32–46, Aug. 2024, doi: 10.24167/proxies.v6i1.12453.

[15] Z. Arif and I. Irwansyah, "Analisis Sentimen Ulasan Pengguna Pada Aplikasi Wondr By BNI Menggunakan Metode Klasifikasi Algoritma Naïve Bayes," *Journal of Information System Research (JOSH)*, vol. 6, no. 3, pp. 1591–1597, Apr. 2025, doi: 10.47065/josh.v6i3.7100.

[16] M. Musdalifah and E. L. Hadisaputro, "Analisis Kepuasan Pengguna Menggunakan Technology Acceptance Model Pada Aplikasi Dana," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 1, pp. 72–78, Dec. 2022, doi: 10.47065/josyc.v4i1.2493.

[17] E. O. I. Pratiwi and W. Yustanti, "Analisis Sentimen Kualitas Layanan Teknologi Pembayaran Elektronik pada Twitter (Studi Kasus Ovo dan Dana)," *JEISBI*, vol. 02, no. 3, p. 2021, 2021, Accessed: Jun. 03, 2026. [Online]. Available: https://ejournal.unesa.ac.id/index.php/JEISBI/article/download/41597/35804

[18] A. J. Tobing and A. Febriandirza, "Analisis Sentimen Aplikasi Mobile Bangking Bca Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Naive Bayes," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 998–1005, Jul. 2024, doi: 10.47065/josh.v5i4.5485.

[19] G. Riansyah Ramadhan and C. Agus Sugianto, "Analisis Sentimen Ulasan Aplikasi Dana Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 9849–9857, Sep. 2024, doi: 10.36040/jati.v8i5.10732.

[20] G. Rafianto, H. -, and A. Ma'sum, "Analisis Sentimen Berbasis Topik Ulasan Pengguna Aplikasi Roblox Menggunakan Integrasi LDA Dan SVM," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 14, no. 2, Apr. 2026, doi: 10.23960/jitet.v14i2.9336.

[21] Y. Kustiyahningsih and Y. Permana, "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) Untuk Menganalisis Sentimen Berdasarkan Aspek Dalam Ulasan Aplikasi EdLink," *Teknika*, vol. 13, no. 1, pp. 127–136, Mar. 2024, doi: 10.34148/teknika.v13i1.746.