

## Klasifikasi Sentimen Ulasan E-Wallet menggunakan TF-IDF dan Random Forest dengan Penyeimbangan Data SMOTE

M. Rioardian Syaputra<sup>1,\*</sup>, Muhammad Arifin<sup>1</sup>, Diana Laily Fithri<sup>1</sup>

<sup>1</sup> Program Studi Sistem Informasi, Universitas Muria Kudus, Indonesia

\* Correspondence: 202153113@std.umk.ac.id

Copyright: © 2025 by the authors

Received: 22 Juni 2025 | Revised: 30 Juni 2025 | Accepted: 26 Juli 2025 | Published: 15 Agustus 2025

### Abstrak

Peningkatan penggunaan *e-wallet* di Indonesia mendorong perlunya pemahaman terhadap persepsi pengguna secara otomatis dan efisien. Salah satu sumber data potensial adalah ulasan pengguna di *Google Play Store*. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna terhadap tiga *e-wallet* populer, yaitu GoPay, OVO, dan DANA, guna mendukung pengembangan layanan. Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis *machine learning*. Sebanyak 30.000 ulasan dikumpulkan menggunakan *google-play-scraper* (masing-masing 10.000 per aplikasi). Data diproses melalui tahapan preprocessing (pelabelan, *stopword removal*, *tokenization*, dan *stemming*), ekstraksi fitur dengan TF-IDF, penyeimbangan data menggunakan SMOTE, dan klasifikasi dengan algoritma Random Forest. Hasil temuan kami menunjukkan bahwa kombinasi Random Forest dan SMOTE meningkatkan kinerja model secara signifikan. Akurasi model Random Forest mencapai 90% (GoPay), 90% (OVO), dan 87% (DANA). Nilai *precision*, *recall*, dan *f1-score weighted* masing-masing sebesar 90%, 89%, dan 89% untuk GoPay; 90%, 90%, dan 90% untuk OVO; serta 88%, 88%, dan 88% untuk DANA. Visualisasi *WordCloud* menguatkan hasil dengan menunjukkan kata-kata dominan pada tiap label sentimen, seperti “bagus”, “bantu”, dan “hilang”. Kombinasi metode TF-IDF, SMOTE, dan Random Forest terbukti efektif dan stabil dalam klasifikasi sentimen pada ketiga aplikasi *e-wallet* tersebut.

**Kata kunci:** analisis sentimen; e-wallet; random forest; smote; tf-idf

### Abstract

The increasing use of *e-wallets* in Indonesia highlights the need to understand user perceptions automatically and efficiently. One valuable data source is user reviews from the *Google Play Store*. This study aims to classify sentiment toward three major *e-wallets*, such as GoPay, OVO, and DANA to support service improvement. A quantitative approach is used with a machine learning-based classification method. A total of 30,000 reviews (10,000 per application) were collected using the *google-play-scraper* library. The data were processed through several stages: preprocessing (labeling, *stopword removal*, *tokenization*, and *stemming*), feature extraction using TF-IDF, data balancing with SMOTE, and classification with the Random Forest algorithm. Our findings show that the combination of Random Forest and SMOTE significantly improves model performance. Accuracy reached 90% (GoPay), 90% (OVO), and 87% (DANA). Precision, recall, and weighted *f1-score* were 90%, 89%, and 89% for GoPay; 90%, 90%, and 90% for OVO; and 88%, 88%, and 88% for DANA. *WordCloud* visualizations further support the findings by highlighting dominant words in each sentiment, such as “good,” “help,” and “lost.” Overall, the integration of TF-IDF, SMOTE, and Random Forest is proven effective and reliable for sentiment classification across the three *e-wallet* platforms.

**Keywords:** sentiment analysis; e-wallet; random forest; smote; tf-idf



## PENDAHULUAN

Kemajuan teknologi informasi telah mengubah berbagai aspek kehidupan, salah satunya sistem pembayaran (Khairi et al., 2023). Di Indonesia, dompet digital (*e-wallet*) seperti GoPay, OVO, dan DANA berkembang pesat, terutama sejak pandemi COVID-19 (Nadhilah et al., 2021) dan didukung oleh pertumbuhan *start-up* di bidang ini (Widianto et al., 2024). *E-wallet* menawarkan transaksi yang mudah, aman, dan efisien melalui *smartphone* (Asfo et al., 2024). Pemerintah juga mendorong penggunaannya demi mewujudkan *cashless society* (Damayanti et al., 2021). Namun, dibalik kemudahan tersebut, pengguna sering menghadapi masalah seperti keterlambatan transaksi dan *bug* sistem. Masalah ini biasanya terungkap dari ulasan pengguna di *platform* seperti *Google Play Store*, yang menyimpan tanggapan berharga terkait performa aplikasi (Nurian & Sari, 2023). Karena volume dan keragaman ulasan sangat tinggi, analisis manual menjadi tidak efektif dan rawan bias. Oleh karena itu, diperlukan pendekatan analisis sentimen berbasis *machine learning* (Pamungkas & Cahyono, 2024).

Analisis sentimen merupakan proses klasifikasi teks ulasan menjadi sentimen positif, negatif, atau netral yang membantu pengembang mendeteksi masalah secara cepat (Setiawan & Suryono, 2024). Proses ini didukung oleh teknik *machine learning*, yaitu pendekatan komputasional yang memungkinkan sistem untuk belajar dari data dan membuat prediksi tanpa pemrograman eksplisit (Dachi & Sitompul, 2023). Dalam konteks ini, *machine learning* digunakan untuk membangun model klasifikasi berdasarkan *rating* bintang aplikasi, yang berfungsi sebagai label sentimen (Prasetyo & Fahrurrozi, 2023). Salah satu algoritma yang digunakan adalah Random Forest, yaitu metode *ensemble learning* yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting* (Haya & Ramme, 2024). Namun, tantangan umum dalam klasifikasi sentimen adalah ketidakseimbangan distribusi data, di mana ulasan positif sering kali mendominasi dan menyebabkan model bias terhadap kelas mayoritas (Berliana & Yusuf, 2025). Untuk mengatasi hal ini, digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE), yang bekerja dengan membuat data sintetis dari kelas minoritas guna menyeimbangkan jumlah data antar kelas (Istiqamah & Rijal, 2024). Kombinasi teknik ini memungkinkan proses klasifikasi berjalan secara lebih adil, akurat, dan dapat menangkap variasi sentimen dengan lebih baik.

Proses klasifikasi menggunakan algoritma Random Forest dikarenakan efektif dalam mengatasi data berdimensi tinggi dan lebih tahan terhadap *overfitting* dibandingkan SVM (Fitri et al., 2024). Proses ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) karena efisien dalam menangkap kata-kata penting dalam ulasan singkat dan informal, serta lebih ringan dibandingkan metode seperti Word2Vec atau BERT (Rahmadani et al., 2024). Penelitian sebelumnya telah menerapkan ketiga metode tersebut, seperti yang dilakukan oleh (Nawang & Aries, 2023) telah menerapkan kombinasi ketiga metode tersebut (TF-IDF, Random Forest, dan SMOTE) dalam analisis sentimen terhadap tiga aplikasi *e-wallet*: GoPay, ShopeePay, dan LinkAja, dengan hasil akurasi yang cukup tinggi, yakni 85,43% untuk GoPay, 85,50% untuk ShopeePay, dan 89,02% untuk LinkAja. Namun, penelitian tersebut belum menguji efektivitas metode yang sama pada aplikasi *e-wallet* lain yang juga populer, seperti GoPay, OVO dan DANA, serta belum mengulas lebih dalam dampak dari distribusi kelas yang tidak seimbang secara komprehensif.

Penelitian ini bertujuan untuk menerapkan metode yang sama pada GoPay, OVO, dan DANA, serta mengevaluasi dampak penyeimbangan data menggunakan SMOTE terhadap kinerja model. Temuan dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem klasifikasi sentimen otomatis, yang dapat digunakan untuk meningkatkan kualitas layanan dan pengalaman pengguna pada *platform e-wallet* di Indonesia.

## METODE

Penelitian ini merupakan penelitian kuantitatif eksperimental yang bertujuan membangun model analisis sentimen menggunakan data ulasan pengguna aplikasi. Terdapat empat tahapan utama, yaitu *preprocessing* data, ekstraksi fitur TF-IDF, *balancing data* menggunakan SMOTE dan klasifikasi menggunakan Random Forest. Analisa dibantu menggunakan bahasa pemrograman python dan *Google Colab*. *Dataset* yang digunakan dalam penelitian ini diperoleh melalui teknik *web scraping* dari platform Google Play Store menggunakan *library google-play-scraper*. Data dikumpulkan dari tiga aplikasi *e-wallet* populer di Indonesia, yaitu GoPay, OVO, dan DANA, masing-masing sebanyak 10.000 ulasan terbaru, sehingga total seluruh *dataset* berjumlah 30.000 ulasan. Ulasan yang diperoleh mencerminkan berbagai persepsi pengguna terhadap aplikasi, mulai dari komentar positif, netral dan negatif.

Tahap berikutnya adalah *preprocessing*, yang dimulai dengan pelabelan sentimen berdasarkan skor ulasan. Skor  $> 3$  dikategorikan sebagai sentimen positif, skor = 3 sebagai netral, serta skor  $< 3$  sebagai negatif. Dilanjutkan penghapusan *stopword* menggunakan *library sastrawi* untuk penghapusan kata umum. Proses *tokenization* dilakukan menggunakan fungsi *split* di python untuk memecah teks dan *stemming* menggunakan *sastrawi stemmer* untuk mengubah kata menjadi bentuk dasarnya. Proses dilanjutkan ke tahap ekstraksi fitur.

Ekstraksi fitur dilakukan menggunakan transformasi TF-IDF, yang bertujuan untuk memberikan bobot pada kata-kata berdasarkan tingkat kepentingannya dalam dokumen. Proses ini dilakukan menggunakan *library scikit-learn* di python, dengan parameter *min\_df=5* dan *max\_df=0.9* untuk menghilangkan kata-kata yang terlalu jarang atau terlalu sering muncul, serta *ngram\_range=(1,2)* untuk mempertimbangkan kombinasi kata dalam bentuk *unigram* dan *bigram* guna menangkap konteks kalimat secara lebih akurat (Yutika et al., 2021). Setelah ekstraksi fitur, dilakukan penyeimbangan data untuk mengatasi ketimpangan antar kelas sentimen.

Langkah selanjutnya adalah penyeimbangan data menggunakan metode SMOTE, dengan cara menyamakan jumlah semua kelas agar setara dengan kelas mayoritas (Sulistiyono et al., 2021). Teknik ini diterapkan pada seluruh kelas sentimen (positif, netral dan negatif), SMOTE dipilih karena lebih stabil dibandingkan metode lain seperti *random oversampling* yang rentan *overfitting*, serta ADASYN yang berpotensi menimbulkan *noise*. Implementasi dilakukan dengan *scikit-learn*.

Setelah seluruh tahapan *preprocessing*, ekstraksi fitur, dan penyeimbangan data selesai dilakukan, model dikembangkan menggunakan algoritma Random Forest untuk mengklasifikasikan sentimen ulasan pengguna ke dalam tiga kategori (positif, netral, dan negatif). Algoritma Random Forest dijalankan menggunakan konfigurasi *default*, dengan *n\_estimators=100* untuk membangun 100 pohon keputusan, *max\_depth=None* yang memungkinkan pohon tumbuh tanpa batas kedalaman, serta *criterion='gini'* sebagai fungsi pemisahan untuk mengukur impuritas setiap *node*.

Evaluasi performa model dilakukan dengan menggunakan metrik klasifikasi seperti akurasi, presisi, *recall*, dan *f1-score* untuk masing-masing kelas. Selain itu, digunakan juga pendekatan *macro average* dan *weighted average* guna memberikan gambaran menyeluruh terhadap performa model dalam konteks klasifikasi *multiclass* dengan distribusi data yang tidak seimbang, serta visualisasi *WordCloud* untuk mengidentifikasi kata-kata dominan yang sering muncul pada masing-masing kategori.

Hasil evaluasi ini menjadi dasar untuk menilai seberapa efektif model dalam memahami sentimen pengguna terhadap layanan *e-wallet*. Dengan dukungan teknik *balancing* SMOTE, evaluasi menunjukkan bahwa model tidak hanya fokus pada kelas mayoritas, tetapi juga mampu mengklasifikasikan kelas minoritas secara lebih adil. Pendekatan ini diharapkan dapat

dimanfaatkan dalam pengembangan sistem analisis sentimen otomatis yang lebih akurat dan responsif terhadap dinamika opini pengguna di *platform* digital.

## HASIL DAN PEMBAHASAN

### Hasil

Setelah tahap pengumpulan data, proses berikutnya adalah pelabelan sentimen berdasarkan nilai *rating* pada ulasan. Skor 1 dan 2 dikategorikan sebagai sentimen negatif, skor 3 sebagai netral, sedangkan skor 4 dan 5 sebagai positif. Untuk menguji keakuratan metode pelabelan ini, dilakukan validasi terhadap 500 sampel data menggunakan pendekatan berbasis kata kunci (*keyword-based*). Hasilnya menunjukkan bahwa hanya 35,20% data yang memiliki kecocokan antara label berdasarkan *keyword* dan *rating*, yang mengindikasikan bahwa pendekatan berbasis *keyword* kurang efektif dalam menangkap nuansa emosional dari suatu ulasan.

Berdasarkan temuan tersebut, pelabelan berdasarkan *rating* tetap digunakan karena dianggap lebih konsisten dan mampu merepresentasikan sentimen secara umum. Setelah proses pelabelan, dilakukan tahap *preprocessing* yang mencakup pengubahan seluruh huruf menjadi huruf kecil (*lowercasing*), pembersihan teks dari simbol dan emoji, penghapusan kata-kata umum (*stopword*), pemisahan teks menjadi unit kata (*tokenization*), dan pengembalian kata ke bentuk dasarnya (*stemming*). Tahapan ini bertujuan untuk menghasilkan teks yang bersih, terstruktur, dan siap digunakan sebagai input pada algoritma klasifikasi.

**Tabel 1.** Perbandingan data sebelum dan sesudah *preprocessing*

| Aplikasi | Sebelum <i>Preprocessing</i>                          |              | Setelah <i>Preprocessing</i>                    |          |
|----------|---|--------------|---|----------|
|          | <i>content</i>  | <i>score</i> | <i>clean review</i>                             | sentimen |
| DANA     | sangat luar biasa dan sangat membantu trimah kasih 🙌👍 | 5            | sangat luar biasa dan sangat bantu terima kasih | Positif  |
| GoPay    | menu gopay pinjam eror gk bisa dibuka                 | 3            | menu gopay pinjam eror buka                     | Netral   |
| OVO      | Biaya transfer ke bank lain potongan nya sampe 2ribu. | 2            | biaya transfer bank potong sampai ribu          | Negatif  |

Berdasarkan hasil *preprocessing data* pada tabel 1 menunjukkan bahwa proses *preprocessing* berhasil mengurangi kata-kata yang tidak penting dan menjaga inti informasi tetap utuh. Kalimat menjadi lebih padat dan bermakna, yang mendukung efektivitas algoritma dalam memahami sentimen yang terkandung di dalamnya. Untuk memperkuat temuan dari contoh tersebut, dilakukan analisis statistik guna mengukur perubahan data secara keseluruhan setelah *preprocessing*.

**Tabel 2.** Statistik hasil *preprocessing* teks ulasan

| Aplikasi | Panjang Sebelum (kata) | Panjang Setelah (kata) | Kata Unik Sebelum | Kata Unik Setelah | Emoji Dihapus |
|----------|------------------------|------------------------|-------------------|-------------------|---------------|
| DANA     | 9.16                   | 6.60                   | 10683             | 5012              | 530           |
| GoPay    | 9.75                   | 7.21                   | 11996             | 5548              | 736           |
| OVO      | 14.13                  | 10.28                  | 15266             | 6557              | 524           |

Hasil pada tabel 2 menunjukkan bahwa tahap *preprocessing* berhasil menyederhanakan data dengan mengurangi panjang rata-rata teks dan jumlah kata unik secara signifikan, serta menghilangkan *noise* seperti emoji. Setelah itu, dilakukan ekstraksi fitur menggunakan TF-IDF untuk mengubah teks menjadi bentuk numerik. TF-IDF memberikan bobot lebih besar

pada kata yang sering muncul di satu dokumen namun jarang di dokumen lain, sehingga dianggap lebih relevan.

**Tabel 3.** Jumlah fitur tf-idf

| Aplikasi | Jumlah Fitur TF-IDF |
|----------|---------------------|
| DANA     | 7964                |
| GoPay    | 9045                |
| OVO      | 12887               |

Hasil pada Tabel 3 menunjukkan bahwa jumlah fitur TF-IDF yang dihasilkan untuk masing-masing aplikasi. Terlihat bahwa aplikasi OVO menghasilkan jumlah fitur terbanyak, yaitu 12.887 kata unik setelah dilakukan proses ekstraksi, diikuti oleh GoPay sebanyak 9.045 fitur, dan DANA sebanyak 7.964 fitur. Perbedaan jumlah fitur ini mencerminkan variasi kosakata dan jumlah total kata yang digunakan oleh pengguna dalam ulasan masing-masing aplikasi. Semakin banyak fitur yang dihasilkan, semakin kompleks representasi data yang diperoleh dari ulasan pengguna.

Guna menginterpretasikan hasil ekstraksi fitur, disajikan sepuluh kata dengan bobot TF-IDF tertinggi pada masing-masing label sentimen untuk ketiga aplikasi. Pada label negatif, kata-kata seperti “hilang”, “saldo”, “potong”, “masuk”, dan “akun” mendominasi, yang mencerminkan keluhan pengguna terhadap gangguan layanan atau transaksi. Pada label netral, ditemukan kata seperti “masuk”, “bagus”, “baik”, dan “update”, yang umumnya bersifat deskriptif atau tidak mengandung emosi yang kuat. Sementara itu, label positif didominasi oleh kata-kata seperti “bagus”, “mantap”, “bantu”, “mudah”, dan “good”, yang menunjukkan kepuasan pengguna terhadap layanan. Pola ini konsisten di seluruh aplikasi (DANA, GoPay, dan OVO), menunjukkan bahwa TF-IDF tidak hanya efektif dalam mengekstraksi fitur numerik, tetapi juga mampu menggambarkan karakteristik sentimen berdasarkan kata-kata dominan. Karena parameter *min\_df* dan *max\_df* telah digunakan untuk menyaring kata-kata yang terlalu jarang atau terlalu sering muncul, proses normalisasi tambahan tidak diperlukan. Secara keseluruhan, hasil ini menunjukkan bahwa proses *preprocessing* dan ekstraksi fitur melalui TF-IDF telah menghasilkan representasi teks yang bersih, bermakna, dan siap untuk dipakai dalam proses klasifikasi sentimen.

**Tabel 4.** Hasil smote

| Aplikasi | Sebelum SMOTE |        | Sesudah SMOTE |        |
|----------|---------------|--------|---------------|--------|
|          | Sentimen      | Jumlah | Sentimen      | Jumlah |
| DANA     | Positif       | 7014   | Positif       | 7014   |
|          | Netral        | 438    | Netral        | 7014   |
|          | Negatif       | 2548   | Negatif       | 7014   |
| GoPay    | Positif       | 6432   | Positif       | 6432   |
|          | Netral        | 412    | Netral        | 6432   |
|          | Negatif       | 3156   | Negatif       | 6432   |
| OVO      | Positif       | 3235   | Positif       | 6260   |
|          | Netral        | 505    | Netral        | 6260   |
|          | Negatif       | 6260   | Negatif       | 6260   |

Selanjutnya, hasil penerapan SMOTE ditunjukkan pada tabel 4, yang memperlihatkan distribusi kelas menjadi seimbang antara label positif dan negatif. Penambahan data sintetis berhasil meningkatkan proporsi kelas minoritas, sehingga model memiliki kesempatan yang

lebih adil dalam mempelajari karakteristik masing-masing kelas. Hal ini diharapkan dapat meningkatkan akurasi dan stabilitas model dalam klasifikasi sentimen.

Hasil proses SMOTE yang ditunjukkan pada Tabel 4 merupakan distribusi data yang awalnya sangat timpang antara kelas menjadi seragam dengan data sintesis yang dibuat oleh SMOTE, ini menunjukkan bahwa SMOTE mengatasi masalah bias data dalam memproses label mayoritas. Kemudian dilakukan *data splitting*, yaitu membagi data hasil SMOTE dengan dua proporsi 80% untuk data latih (*training data*) dan 20% untuk data uji (*testing data*). Pada tahapan evaluasi model, memakai metrik *confusion matrix* untuk menilai performa Random Forest dalam memprediksi tiga label: positif, netral, dan negatif. Analisis dilakukan berdasarkan hasil *macro average* dan *weighted average* untuk memberikan gambaran menyeluruh terhadap performa model pada ketiga label secara proporsional. Hal ini bertujuan untuk mengidentifikasi potensi permasalahan model dalam menangani klasifikasi antar label.

**Tabel 5.** Evaluasi klasifikasi model

| Aplikasi | Akurasi |             | Precision |             | Recall |             | F1-Score |             |
|----------|---------|-------------|-----------|-------------|--------|-------------|----------|-------------|
|          | SMOTE   | Tanpa SMOTE | SMOTE     | Tanpa SMOTE | SMOTE  | Tanpa SMOTE | SMOTE    | Tanpa SMOTE |
| Dana     | 87%     | 84%         | 88%       | 80%         | 88%    | 82%         | 88%      | 81%         |
| GoPay    | 90%     | 86%         | 88%       | 84%         | 89%    | 86%         | 89%      | 85%         |
| OVO      | 89%     | 82%         | 90%       | 81%         | 90%    | 84%         | 90%      | 82%         |

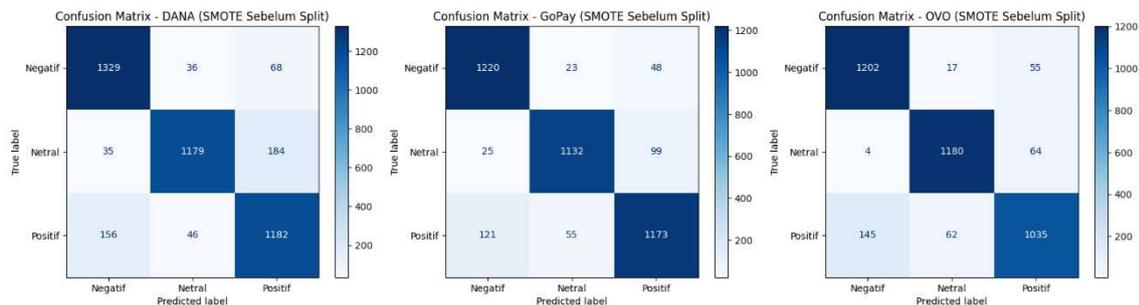
Berdasarkan hasil evaluasi klasifikasi model pada Tabel 5, SMOTE secara konsisten meningkatkan performa model pada ketiga aplikasi. Pada aplikasi GoPay, akurasi meningkat dari 86% menjadi 90% dan *f1-score* dari 85% menjadi 89%. Aplikasi Dana menunjukkan peningkatan akurasi dari 84% menjadi 87% dan *f1-score* dari 81% menjadi 88%. Sementara pada aplikasi OVO, akurasi naik dari 82% menjadi 89% dan *f1-score* dari 82% menjadi 90%. Hasil ini menunjukkan bahwa SMOTE berhasil mengurangi bias terhadap kelas mayoritas serta meningkatkan kemampuan model dalam mengenali pola dari kelas minoritas.

**Tabel 6.** Evaluasi *macro average* dan *weighted average*

| Aplikasi | Precision     |                  | Recall        |                  | F1-Score      |                  |
|----------|---------------|------------------|---------------|------------------|---------------|------------------|
|          | Macro Average | Weighted Average | Macro Average | Weighted Average | Macro Average | Weighted Average |
| Dana     | 87%           | 88%              | 86%           | 88%              | 86%           | 88%              |
| GoPay    | 90%           | 88%              | 90%           | 89%              | 90%           | 89%              |
| OVO      | 89%           | 90%              | 90%           | 90%              | 89%           | 90%              |

Selanjutnya, pada hasil yang ditunjukkan pada tabel 6, bahwa nilai *macro average* dan *weighted average* untuk *precision*, *recall*, dan *f1-score*. Pada aplikasi Dana, nilai *macro average* untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 87%, 86%, dan 86%, sementara nilai *weighted average* masing-masing sebesar 88%. Hal ini menunjukkan bahwa model mampu memberikan performa yang seimbang antar kelas, sekaligus mempertahankan akurasi tinggi pada distribusi kelas yang dominan. Untuk aplikasi GoPay, nilai *macro average* mencapai 90% pada ketiga metrik, sedangkan *weighted average* berkisar antara 88% hingga 89%. Aplikasi OVO juga menunjukkan performa serupa dengan *macro average* sebesar 89% hingga 90% dan *weighted average* sebesar 90% pada semua metrik. Nilai yang tinggi dan konsisten ini menunjukkan bahwa model tetap akurat dan stabil dalam menangani distribusi kelas yang tidak seimbang, terutama setelah diterapkannya SMOTE. Setelah memperoleh hasil evaluasi metrik

umum seperti *precision*, *recall*, dan *f1-score* pada tabel 6, analisis selanjutnya dilakukan secara lebih rinci melalui *confusion matrix*.



Gambar 1. Confusion matrix

Pada gambar 1 menunjukkan *confusion matrix* untuk DANA, GoPay, dan OVO, yang memperlihatkan bahwa sebagian besar prediksi model sesuai dengan label sebenarnya, ditandai oleh nilai diagonal yang tinggi. Khususnya pada GoPay, model mampu mengklasifikasikan label netral dan positif dengan baik. Meskipun terdapat sedikit kesalahan klasifikasi, hasil ini mencerminkan kemampuan model yang solid setelah penerapan SMOTE sebelum data dibagi.



Gambar 2. WordCloud dana



Gambar 3. WordCloud gopay



Gambar 4. WordCloud ovo

Hasil visualisasi *WordCloud* pada gambar 2, 3, dan 4 pada setiap sentimen menunjukkan bahwa kata-kata seperti “masuk”, “hilang”, “potong”, dan “transaksi” sering muncul dalam ulasan negatif, sedangkan pada sentimen netral ditemukan kata-kata seperti “saldo”, “akun”, dan “update” yang mengindikasikan adanya ketidakpastian atau harapan terhadap perbaikan layanan. Hal ini menunjukkan perlunya peningkatan pada sistem *login*, transparansi mutasi saldo, dan keandalan proses transaksi. Sementara itu, kata-kata seperti “bagus”, “mantap”, dan “bantu” mendominasi sentimen positif, namun peningkatan pengalaman pengguna secara menyeluruh tetap penting agar kepuasan pengguna lebih merata dan berkelanjutan.

## Pembahasan

Hasil penelitian menunjukkan bahwa model Random Forest mampu memberikan performa klasifikasi yang tinggi pada analisis sentimen ulasan *e-wallet*. GoPay mencatat akurasi tertinggi sebesar 90%, disusul oleh OVO sebesar 89%, dan DANA sebesar 87%. Tingginya akurasi ini mencerminkan bahwa kombinasi metode *preprocessing*, ekstraksi fitur menggunakan TF-IDF, serta penyeimbangan data dengan SMOTE berhasil meningkatkan performa model secara signifikan. Tahapan *preprocessing* memiliki peran fundamental dalam meningkatkan kualitas data. Proses ini mencakup pelabelan berdasarkan *rating* ulasan, *stopword removal*, *tokenization*, hingga *stemming*. Setiap tahapan berfungsi menyederhanakan teks dan mengurangi *noise*, sehingga data menjadi lebih bersih dan representatif untuk diolah lebih lanjut.

Proses *preprocessing* memungkinkan ekstraksi fitur dengan TF-IDF menjadi lebih akurat dalam mengidentifikasi kata-kata penting yang mencerminkan sentimen. Kata-kata seperti “bagus”, “bantu”, dan “hilang” muncul sebagai kata dominan yang relevan dengan masing-masing label sentimen, yang menunjukkan bahwa representasi fitur mampu membedakan karakteristik setiap kelas dengan baik. Aplikasi OVO menghasilkan jumlah fitur TF-IDF terbanyak, yang menandakan bahwa keragaman kosakata dalam ulasannya tinggi. Namun demikian, Random Forest tetap mampu mengklasifikasikan ulasan dengan baik berkat kemampuannya menangani data berdimensi tinggi dan kompleks.

Penerapan SMOTE juga menjadi faktor kunci yang memengaruhi keberhasilan model. Teknik ini diterapkan sebelum pembagian data latih dan uji, sehingga hanya data latih yang mengalami proses *oversampling*. Strategi ini penting untuk menghindari *data leakage* yang dapat menimbulkan hasil evaluasi yang menyesatkan. SMOTE bekerja dengan menghasilkan data sintesis untuk kelas minoritas, sehingga proporsi antar kelas menjadi seimbang. Hasil evaluasi sebelum dan sesudah penerapan SMOTE menunjukkan adanya peningkatan nilai *precision*, *recall*, dan *f1-score* pada ketiga label sentimen. Ini membuktikan bahwa SMOTE tidak hanya mengatasi permasalahan ketidakseimbangan kelas, tetapi juga membantu model belajar pola dari semua kelas secara proporsional tanpa menyebabkan *overfitting*.

Hasil temuan yang dilakukan oleh (Nawang & Aries, 2023) menunjukkan bahwa kombinasi TF-IDF, Random Forest, dan SMOTE dapat menghasilkan akurasi tinggi pada beberapa aplikasi *e-wallet*. Namun, evaluasi terhadap SMOTE belum dilakukan secara mendalam dan tidak mencakup aplikasi seperti OVO dan DANA. Sementara itu, penelitian oleh (Larasati et al., 2022) yang hanya menggunakan Random Forest tanpa SMOTE pada aplikasi DANA menghasilkan akurasi 84%, dengan *recall* kelas negatif yang rendah 32%, menunjukkan kurang optimalnya performa dalam menangani ketidakseimbangan data. Selain itu, penelitian oleh (Kaeren & Andrianingsih, 2025) menganalisis aplikasi LinkAja menggunakan Naive Bayes dan Random Forest tanpa *balancing data*, dan menemukan bahwa mayoritas ulasan bersentimen negatif. Meskipun Random Forest menunjukkan performa lebih baik dengan akurasi 82% dibandingkan Naive Bayes 79%, evaluasi belum menyertakan teknik penyeimbangan data seperti SMOTE maupun eksplorasi pengaruhnya terhadap distribusi kelas.

Penelitian ini menawarkan pendekatan yang lebih menyeluruh, dengan strategi *preprocessing* yang dirancang khusus untuk teks informal berbahasa Indonesia, serta penerapan SMOTE sebelum *data splitting*. Hasilnya menunjukkan peningkatan akurasi dan stabilitas model secara konsisten pada tiga aplikasi *e-wallet*. Kebaruan penelitian ini terletak pada integrasi *preprocessing* yang sistematis, *balancing data* berbasis SMOTE, serta evaluasi menyeluruh menggunakan metrik seperti *confusion matrix*, *macro average*, dan *weighted average*, sehingga dapat menjadi acuan dalam pengembangan sistem analisis sentimen otomatis di aplikasi digital lainnya.

## SIMPULAN

Penelitian ini membuktikan bahwa kombinasi metode TF-IDF, algoritma Random Forest, dan teknik SMOTE mampu mengklasifikasikan sentimen pengguna *e-wallet* secara efektif. Model yang dihasilkan menunjukkan akurasi tinggi dan metrik evaluasi yang seimbang pada ketiga kelas sentimen. SMOTE terbukti meningkatkan performa model dengan mengatasi ketidakseimbangan data. Visualisasi *WordCloud* turut memperkuat hasil dengan menampilkan kata-kata dominan di tiap label. Namun, pendekatan berbasis TF-IDF dan model pohon keputusan masih memiliki keterbatasan dalam memahami konteks semantik. Disarankan untuk penelitian berikutnya dapat mencoba kombinasi metode berbasis *deep learning* seperti *Long Short Term Memory* (LSTM) atau *Bidirectional Encoder Representations from Transformers* (BERT) untuk hasil yang lebih mendalam dan akurat.

## REFERENSI

- Asfo, N. S., Setiawan, H. A., Pessireron, A. G., Saleh, M., & Ladjin, N. (2024). Analisis Determinan Penggunaan E Wallet Analysis Of Determinants Of E Wallet Use. *Jurnal Kolaboratif Sains*, 7(5), 1874–1884.
- Berliana, H., & Yusuf, R. (2025). Analisis Sentimen Terhadap Penggunaan Donasi Korban Penyiraman Air Keras Pada Media Sosial X.Com Menggunakan Metode Bert. *Journal of Science and Social Research*, 4307(2), 1134–1142.
- Dachi, J. M. A. S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma Xgboost Dan Algoritma Random Forest Ensemble Learning Pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, 2(2), 87–103. <https://doi.org/10.55606/jurrimipa.v2i2.1470>
- Damayanti, P., Purwitasari, D., & Suciati, N. (2021). Eliminasi Data Non-Topic Menggunakan Pemodelan Topik Untuk Peringkasan Otomatis Data Tweet Dengan Konteks Covid-19. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 8(1), 199–208. <https://doi.org/10.25126/jtiik.202184324>
- Fitri, D. D., Agustian, S., Pizaini, P., & Sanjaya, S. (2024). Klasifikasi Sentimen Pada Dataset Terbatas Menggunakan Random Forest Dan Word2Vec. *Journal of Computer System and Informatics (JoSYC)*, 6(1), 214–222. <https://doi.org/10.47065/josyc.v6i1.6246>
- Haya, A. N., & Ramme, M. Y. (2024). Penerapan Algoritma Stacking Ensemble Machine Learning Berbasis Pohon Untuk Prediksi Penyakit Diabetes. *Prosiding Seminar Nasional Sains Data*, 4(1), 954–961. <https://doi.org/10.33005/senada.v4i1.388>
- Istiqamah, N., & Rijal, M. (2024). Klasifikasi Ulasan Konsumen Menggunakan Random Forest Dan SMOTE. *Journal of System and Computer Engineering (JSCE)*, 5(1), 66–77. <https://doi.org/10.61628/jsce.v5i1.1061>
- Kaeren, & Andrianingsih. (2025). Analisis Sentimen Aplikasi Linkaja Di Google Play Store Menggunakan Algoritma Naïve Bayes Dan Random Forest. *Jurnal Riset dan Aplikasi Mahasiswa Informatika*, 6(2), 438–447. <https://doi.org/10.30998/jrami.v6i02.13821>
- Khairi, M., Rianto, B., Chrismondari, Y., Jalil, M., Juita, H., & Sudeska, E. (2023). Peran Teknologi Dalam Transformasi Ekonomi Dan Bisnis Di Era Digital. *JMEB Jurnal Manajemen Ekonomi & Bisnis*, 1(01), 11–22. <https://doi.org/10.59561/jmeh.v1i01.89>
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana Dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(9), 4305–4313.
- Nadhilah, P., Jatikusumo, R. I., & Permana, E. (2021). Efektifitas Penggunaan E-Wallet Dikalangan Mahasiswa Dalam Proses Menentukan Keputusan Pembelian. *JEMMA (Jurnal of Economic, Management, and Accounting)*, 4(9), 128–138. <https://doi.org/10.35914/jemma.v4i2.725>
- Nawang, A. H., & Aries, D. I. (2023). Analisis Sentimen Pada Aplikasi Dompot Digital

- Menggunakan Algoritma Random Forest. *Journal of Emerging Information Systems and Business Intelligence*, 4(3), 186–192.
- Nurian, A., & Sari, B. N. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3s1), 829–835. <https://doi.org/10.23960/jitet.v11i3s1.3348>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review Chatgpt Di Play Store Menggunakan Support Vector Machine Dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Prasetyo, M. R., & Fahrurozi, A. (2023). Analisa Sentimen Pada Ulasan Google Untuk Hotel Gran Mahakam Jakarta Menggunakan Pendekatan Machine Learning. *Jurnal Ilmiah Informatika Komputer*, 28(3), 203–217. <https://doi.org/10.35760/ik.2023.v28i3.9761>
- Rahmadani, R., Rahim, A., & Rudiman, R. (2024). Analisis Sentimen Ulasan “Ojol The Game” Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3), 145–156. <https://doi.org/10.23960/jitet.v12i3.4988>
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara Menggunakan Algoritma Support Vector Machine Dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183–192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Sulistiyono, M., Pristyanto, Y., Adi, S., & Gumelar, G. (2021). Implementasi algoritma synthetic minority over-sampling technique untuk menangani ketidakseimbangan kelas pada dataset klasifikasi. *Sistemasi: Jurnal Sistem Informasi*, 10(2), 445-459. <https://doi.org/10.32520/stmsi.v10i2.1303>
- Widianto, I. S., Ramadhan, Y. R., Ramadhan, Y. R., Komara, M. A., & Komara, M. A. (2024). Analisis Sentimen E-Wallet Gopay, Shopeepay, Dan Ovo Menggunakan Algoritma Naive Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1), 4155–4163. <https://doi.org/10.23960/jitet.v12i3s1.5277>
- Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek Pada Review Female Daily Menggunakan TF-IDF Dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422-430. <https://doi.org/10.30865/mib.v5i2.2845>