Journal of Computer Science and Technology
Volume 4 ; Nomor 2 ; Mei 2026 ; Page 97-109
DOI : https://doi.org/10.59435/jocstec.v4i2.759
https://jurnal.padangtekno.com/index.php/jocstec
E-ISSN : 2985-4318
P-ISSN : 2985-5772

Peningkatan Akurasi Klasifikasi Sentimen Pengguna Dompet Digital Menggunakan Stacking Ensemble Machine Learning

Ilmawati[1*], Nadya Alinda Rahmi[2], Elvira Sawitri[3]

[1,2,3]Jurusan Sistem Informasi, Universitas Putra Indonesia YPTK
[1]ilmawati@upiyptk.ac.id, [2]nadyaalindaa@upiyptk.ac.id [3]elvira_sawitri@upiyptk.ac.id

Abstrak

Penggunaan dompet digital yang terus meningkat menghasilkan banyak ulasan pengguna yang dapat dimanfaatkan untuk mengevaluasi kualitas layanan. Penelitian ini bertujuan meningkatkan akurasi klasifikasi sentimen pengguna dompet digital menggunakan metode Stacking Ensemble Machine Learning yang mengombinasikan Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost dengan Logistic Regression sebagai *meta-learner*. Data ulasan diproses melalui tahapan *text preprocessing* meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, *stemming*, dan pembobotan fitur menggunakan TF-IDF. Penyeimbangan data dilakukan dengan SMOTE, sedangkan evaluasi model menggunakan 5-Fold Cross-Validation. Hasil penelitian menunjukkan bahwa model Stacking Ensemble memperoleh akurasi rata-rata 80,55%, lebih tinggi dibandingkan algoritma dasar. Evaluasi menggunakan Confusion Matrix, Classification Report, dan ROC Curve juga menunjukkan peningkatan nilai precision, recall, F1-score, dan kemampuan diskriminasi model. Hasil ini menunjukkan bahwa pendekatan Stacking Ensemble Machine Learning efektif untuk meningkatkan akurasi klasifikasi sentimen pengguna dompet digital serta mendukung evaluasi kualitas layanan berbasis opini pengguna.

Kata Kunci : Dompet Digital, Stacking Ensemble, Machine Learning, TF-IDF, SMOTE

Abstract

The rapid growth of digital wallet usage has generated a large volume of user reviews that can be utilized to evaluate service quality. This study aims to improve the accuracy of digital wallet user sentiment classification using a Stacking Ensemble Machine Learning approach that combines Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), and AdaBoost with Logistic Regression as the meta-learner. User reviews were processed through text preprocessing stages, including case folding, text cleaning, tokenization, stopword removal, stemming, and TF-IDF feature weighting. Synthetic Minority Over-sampling Technique (SMOTE) was employed to address class imbalance, while model performance was evaluated using 5-Fold Cross-Validation. The experimental results show that the proposed Stacking Ensemble model achieved an average accuracy of 80.55%, outperforming the individual base learners. Furthermore, evaluations based on the Confusion Matrix, Classification Report, and Receiver Operating Characteristic (ROC) Curve demonstrated improvements in precision, recall, F1-score, and the model's discriminative capability. These findings indicate that the proposed Stacking Ensemble Machine Learning approach is effective in improving the accuracy of digital wallet user sentiment classification and can serve as a reliable tool for supporting service quality evaluation based on user opinions.

Keyword : Digital Wallet, Stacking Ensemble, Machine Learning, TF-IDF, SMOTE.

E-ISSN : 2985-4318
P-ISSN : 2985-5772

PENDAHULUAN

Seiring dengan pesatnya perkembangan teknologi finansial (financial technology), penggunaan dompet digital semakin meningkat sebagai salah satu metode pembayaran yang praktis, cepat, dan aman. Berbagai layanan dompet digital seperti pembayaran tagihan, transfer dana, pembelian produk, hingga transaksi di berbagai platform digital telah menjadi bagian dari aktivitas masyarakat sehari-hari. Peningkatan jumlah pengguna tersebut juga diikuti dengan meningkatnya jumlah ulasan yang diberikan pengguna pada berbagai platform digital, seperti Google Play Store dan App Store. Ulasan tersebut mengandung informasi penting mengenai kualitas layanan, kepuasan pengguna, serta berbagai permasalahan yang dialami selama menggunakan aplikasi. Oleh karena itu, analisis sentimen terhadap ulasan pengguna dompet digital menjadi salah satu pendekatan yang penting untuk membantu penyedia layanan memahami kebutuhan pengguna dan meningkatkan kualitas layanan [1].

Analisis sentimen berbasis machine learning telah berkembang menjadi salah satu metode yang banyak digunakan dalam mengekstraksi opini pengguna dari data teks karena mampu mengenali pola-pola kompleks yang sulit diidentifikasi menggunakan metode konvensional. Berbagai penelitian sebelumnya telah menerapkan algoritma machine learning untuk melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi digital. Penelitian lain membandingkan beberapa algoritma, seperti Naïve Bayes, Random Forest, dan Support Vector Machine, dengan pendekatan K-Fold Cross Validation, yang menunjukkan bahwa setiap algoritma memiliki keunggulan dan keterbatasan pada karakteristik data yang berbeda [2]. Selain itu, penelitian lain menunjukkan bahwa penggunaan algoritma boosting maupun metode ensemble learning mampu meningkatkan performa klasifikasi dibandingkan algoritma tunggal, terutama pada data teks yang memiliki karakteristik kompleks dan jumlah fitur yang besar [3].

Meskipun demikian, penelitian-penelitian terdahulu masih memiliki beberapa keterbatasan. Sebagian besar penelitian masih menggunakan pendekatan algoritma tunggal atau hanya melakukan perbandingan performa beberapa algoritma tanpa memanfaatkan teknik ensemble learning yang mampu mengombinasikan keunggulan dari berbagai model. Selain itu, beberapa penelitian belum mengatasi permasalahan ketidakseimbangan kelas (class imbalance) yang sering ditemukan pada data ulasan pengguna, sehingga berpotensi menurunkan kinerja model klasifikasi [4]. Di sisi lain, variasi karakteristik dataset, seperti jumlah data, distribusi kelas, serta proses text preprocessing, juga belum dijelaskan secara komprehensif sehingga memengaruhi kemampuan generalisasi model [5]. Beberapa penelitian juga belum menerapkan metode validasi yang konsisten, sehingga hasil akurasi yang diperoleh sulit dibandingkan secara objektif antarpenelitian.

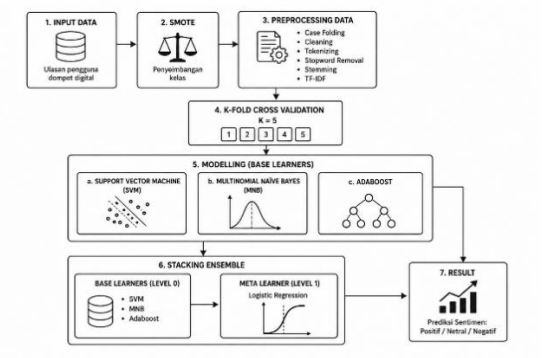
Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan Stacking Ensemble Machine Learning untuk meningkatkan akurasi klasifikasi sentimen pengguna dompet digital. Tahapan penelitian dimulai dengan proses text preprocessing yang meliputi case folding, cleaning, tokenization, stopword removal, stemming, dan pembobotan fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF) untuk menghasilkan representasi data yang optimal. Selanjutnya, diterapkan metode Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas sehingga distribusi data menjadi lebih seimbang. Dataset kemudian divalidasi menggunakan K-Fold Cross Validation agar seluruh data memiliki kesempatan yang sama sebagai data pelatihan dan pengujian, sehingga hasil evaluasi menjadi lebih andal. Model yang diusulkan menggunakan tiga algoritma sebagai base learner, yaitu Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost, sedangkan Logistic Regression digunakan sebagai meta learner untuk mengombinasikan hasil prediksi dari masing-masing algoritma dasar sehingga menghasilkan model klasifikasi yang lebih akurat dan robust.

Penelitian ini memiliki beberapa keunggulan dibandingkan penelitian sebelumnya. Pertama, penelitian ini menerapkan teknik stacking ensemble yang mampu menggabungkan kelebihan beberapa algoritma machine learning, sehingga menghasilkan performa klasifikasi yang lebih baik dibandingkan penggunaan algoritma tunggal. Kedua, penelitian ini mengintegrasikan proses

SMOTE, TF-IDF, dan K-Fold Cross Validation dalam satu kerangka kerja sehingga mampu mengatasi permasalahan ketidakseimbangan data, meningkatkan kualitas representasi fitur, serta menghasilkan evaluasi model yang lebih stabil. Ketiga, kombinasi algoritma SVM, Multinomial Naïve Bayes, dan AdaBoost memungkinkan model mempelajari karakteristik data dari berbagai sudut pandang, sedangkan Logistic Regression sebagai meta learner mengoptimalkan hasil prediksi akhir melalui proses penggabungan keluaran dari seluruh base learner. Dengan pendekatan tersebut, penelitian ini diharapkan mampu menghasilkan model klasifikasi sentimen pengguna dompet digital yang memiliki tingkat akurasi lebih tinggi, lebih stabil, dan lebih efektif dalam mendukung evaluasi kualitas layanan berdasarkan opini pengguna.

METODOLOGI PENELITIAN

Penelitian ini memiliki signifikansi yang kuat dalam bidang machine learning, text mining, dan analisis sentimen karena mengintegrasikan beberapa teknik modern untuk menghasilkan model klasifikasi sentimen yang lebih akurat, robust, dan andal dalam menganalisis opini pengguna dompet digital. Kerangka penelitian yang diusulkan dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

Berikut merupakan kontribusi utama dari penelitian ini.

A. Peningkatan Kinerja Klasifikasi Sentimen Melalui Stacking Ensemble

Penelitian ini mengombinasikan tiga algoritma *machine learning* sebagai base learner, yaitu Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost, yang masing-masing memiliki karakteristik dan keunggulan berbeda dalam melakukan klasifikasi data teks.

1. Support Vector Machine (SVM) memiliki kemampuan yang baik dalam menangani data berdimensi tinggi dan membangun *hyperplane* optimal sehingga efektif untuk mengklasifikasikan data teks dengan jumlah fitur yang besar.
2. Multinomial Naïve Bayes (MNB) merupakan algoritma yang efisien untuk klasifikasi dokumen teks karena bekerja berdasarkan distribusi probabilitas kata, sehingga sangat sesuai digunakan pada analisis sentimen berbasis TF-IDF.

Kombinasi ketiga algoritma melalui metode stacking ensemble memungkinkan setiap model saling melengkapi keunggulan dan kelemahannya sehingga menghasilkan prediksi sentimen yang lebih akurat, stabil, dan tidak bergantung pada satu algoritma saja.

B. Optimalisasi Prediksi Menggunakan Logistic Regression sebagai Meta Learner

Prediksi yang dihasilkan oleh masing-masing *base learner* selanjutnya diproses menggunakan Logistic Regression sebagai meta learner. Model ini berfungsi mempelajari pola terbaik dari keluaran setiap algoritma dasar dan menghasilkan keputusan klasifikasi akhir yang lebih optimal.

Keunggulan pendekatan ini meliputi:

1. Meningkatkan kemampuan generalisasi model karena kesalahan prediksi dari satu algoritma dapat dikompensasi oleh algoritma lainnya.

2. Logistic Regression mampu memberikan pembobotan terhadap probabilitas keluaran dari masing-masing *base learner*, sehingga menghasilkan keputusan akhir yang lebih akurat.

C. Penyeimbangan Data Menggunakan SMOTE

Permasalahan ketidakseimbangan kelas (*class imbalance*) sering dijumpai pada data ulasan pengguna dompet digital, terutama ketika jumlah ulasan positif jauh lebih banyak dibandingkan ulasan negatif atau netral

Untuk mengatasi kondisi tersebut, penelitian ini menerapkan Synthetic Minority Oversampling Technique (SMOTE) [6], sehingga:

1. Distribusi data antar kelas menjadi lebih seimbang.
2. Model mampu mengenali karakteristik kelas minoritas dengan lebih baik.
3. Risiko bias terhadap kelas mayoritas dapat diminimalkan.
4. Nilai *precision*, *recall*, dan *F1-score* meningkat pada seluruh kelas.

D. Validasi Model Menggunakan K-Fold Cross Validation

Untuk memperoleh hasil evaluasi yang lebih objektif dan mengurangi risiko overfitting, penelitian ini menggunakan K-Fold Cross Validation dengan K = 5.

Pendekatan ini memberikan beberapa keuntungan, yaitu:

1. Seluruh data memperoleh kesempatan menjadi data latih maupun data uji.
2. Evaluasi model menjadi lebih stabil.
3. Hasil akurasi tidak bergantung pada satu pembagian dataset tertentu.
4. Kemampuan generalisasi model dapat diukur secara lebih baik.

Dengan mengombinasikan TF-IDF, SMOTE, Stacking Ensemble, dan Logistic Regression, penelitian ini diharapkan mampu menghasilkan model klasifikasi sentimen pengguna dompet digital yang memiliki tingkat akurasi, stabilitas, dan kemampuan generalisasi yang lebih baik dibandingkan pendekatan konvensional, sehingga dapat mendukung data-driven decision making dalam pengembangan layanan keuangan digital.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dan pembahasan mengenai klasifikasi sentimen pengguna dompet digital menggunakan metode *Stacking Ensemble Machine Learning*. Pembahasan diawali dengan proses pengumpulan data ulasan pengguna dompet digital dari platform digital, kemudian dilanjutkan dengan proses pelabelan sentimen, *text preprocessing*, analisis distribusi data, serta validasi model menggunakan metode 5-Fold Cross-Validation. Selanjutnya, dilakukan pengujian terhadap masing-masing algoritma dasar (*base learner*), yaitu Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost. Pada tahap akhir, kinerja model Stacking Ensemble dibandingkan dengan kinerja masing-masing algoritma dasar untuk mengetahui efektivitas pendekatan ensemble dalam meningkatkan akurasi klasifikasi sentimen pengguna dompet digital.

Hasil pengujian dievaluasi menggunakan beberapa metrik performa, yaitu Accuracy, Precision, Recall, dan F1-Score. Selain itu, analisis juga dilakukan terhadap Confusion Matrix untuk mengidentifikasi kemampuan model dalam mengklasifikasikan sentimen positif, negatif, dan netral secara lebih rinci. Pembahasan pada bagian ini bertujuan untuk menunjukkan sejauh mana metode *Stacking Ensemble* mampu menghasilkan performa yang lebih baik dibandingkan penggunaan algoritma tunggal dalam proses klasifikasi sentimen.

A. Pengumpulan Data Ulasan Pengguna Dompet Digital

Dataset yang digunakan dalam penelitian ini terdiri atas ulasan pengguna dompet digital yang dikumpulkan dari berbagai platform digital yang menyediakan ulasan publik, seperti Google Play Store dan App Store. Ulasan tersebut merepresentasikan pengalaman, kepuasan, keluhan, serta

harapan pengguna terhadap layanan dompet digital yang digunakan dalam aktivitas transaksi sehari-hari [7]

Ulasan yang dikumpulkan membahas berbagai aspek yang berkaitan dengan kualitas layanan dompet digital, antara lain kemudahan penggunaan aplikasi (*usability*), kecepatan transaksi, keamanan akun dan data pengguna, kemudahan proses verifikasi, kualitas layanan pelanggan (*customer service*), stabilitas aplikasi, fitur promosi dan cashback, proses pembayaran melalui QRIS, transfer dana, top-up saldo, serta pengalaman pengguna secara keseluruhan. Aspek-aspek tersebut dipilih karena merupakan faktor utama yang memengaruhi tingkat kepuasan pengguna terhadap layanan dompet digital.

Pada tahap awal, seluruh data ulasan melalui proses penyaringan (*data screening*) untuk menghilangkan data yang tidak relevan, seperti ulasan yang tidak berkaitan dengan penggunaan aplikasi, data duplikat, ulasan kosong, ulasan yang hanya berisi simbol atau karakter khusus, serta teks yang tidak memberikan informasi yang bermakna. Setiap ulasan yang memenuhi kriteria dianggap sebagai satu observasi, sedangkan kelas sentimen (positif, negatif, atau netral) digunakan sebagai variabel target dalam proses klasifikasi.

Setelah proses penyaringan selesai, diperoleh 1.800 ulasan pengguna dompet digital yang digunakan sebagai dataset penelitian. Dataset tersebut kemudian dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) melalui mekanisme 5-Fold Cross-Validation, sehingga seluruh data dapat dimanfaatkan secara optimal dalam proses pembentukan dan evaluasi model.

Penggunaan ulasan yang dihasilkan langsung oleh pengguna memberikan representasi empiris mengenai persepsi masyarakat terhadap kualitas layanan dompet digital. Namun demikian, ulasan pada platform digital umumnya ditulis menggunakan bahasa yang bersifat informal, mengandung singkatan, kesalahan penulisan (*typo*), bahasa gaul, campuran bahasa Indonesia dan bahasa Inggris, penggunaan emoji, serta istilah-istilah yang spesifik pada layanan dompet digital. Oleh karena itu, diperlukan tahapan text preprocessing yang sistematis, meliputi *case folding*, *cleaning*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*, serta proses pelabelan sentimen sebelum data digunakan pada tahap klasifikasi menggunakan metode Stacking Ensemble Machine Learning.

B. Pelabelan Sentimen

Setelah proses pengumpulan dan penyaringan data selesai dilakukan, setiap ulasan pengguna dompet digital diberikan label sentimen berdasarkan isi dan makna yang terkandung dalam teks ulasan. Pada penelitian ini digunakan dua kelas sentimen, yaitu kelas 0 yang merepresentasikan sentimen negatif dan kelas 1 yang merepresentasikan sentimen positif. Contoh data ulasan beserta label sentimennya disajikan pada Tabel 1.

Ulasan yang berisi keluhan, kritik, kekecewaan, atau pengalaman yang kurang memuaskan dikategorikan sebagai sentimen negatif. Beberapa contoh ulasan negatif meliputi keluhan mengenai aplikasi yang sering mengalami *crash*, proses transaksi yang gagal, lambatnya proses verifikasi akun, gangguan sistem saat melakukan pembayaran, kesulitan melakukan *top-up* saldo, kendala pada proses transfer dana, masalah keamanan akun, serta respons layanan pelanggan (*customer service*) yang kurang memuaskan. [8]

Sebaliknya, ulasan yang mengandung ungkapan kepuasan, apresiasi, rekomendasi, atau pengalaman positif dikategorikan sebagai sentimen positif. Ulasan positif umumnya berisi pernyataan mengenai kemudahan penggunaan aplikasi, kecepatan transaksi, keamanan sistem, kelengkapan fitur pembayaran digital, kemudahan penggunaan QRIS, banyaknya promo dan cashback, stabilitas aplikasi, serta kualitas layanan pelanggan yang baik.

Tahap pelabelan sentimen merupakan salah satu proses yang sangat penting karena kualitas model klasifikasi sangat dipengaruhi oleh ketepatan dan konsistensi pemberian label pada setiap data. Ulasan yang mengandung opini campuran, yaitu memuat unsur positif dan negatif secara bersamaan, memerlukan analisis yang lebih cermat. Oleh karena itu, proses pelabelan dilakukan

berdasarkan sentimen yang paling dominan dalam setiap ulasan agar menghasilkan label yang konsisten dan representatif.
Data yang telah diberi label selanjutnya digunakan sebagai ground truth dalam proses pelatihan (*training*) dan evaluasi model klasifikasi menggunakan algoritma Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), AdaBoost, serta model Stacking Ensemble Machine Learning yang diusulkan pada penelitian ini.

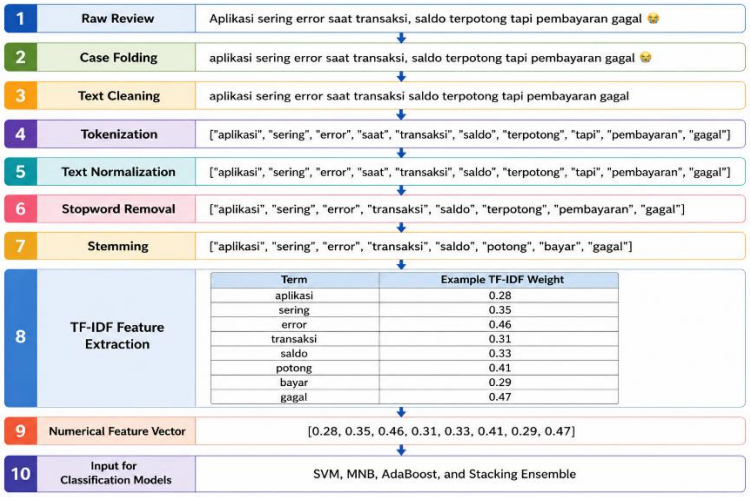
Tabel 1. Contoh Data Ulasan Pengguna Dompet Digital dan Label Sentimen

No	Ulasan Pengguna Dompet Digital	Sentimen
1	Aplikasi sering error saat melakukan pembayaran, transaksi gagal terus.	Negative
2	Sudah berkali-kali login tetapi selalu gagal verifikasi OTP.	Negative
3	Saldo saya terpotong, tetapi transaksi tidak berhasil. Sangat mengecewakan.	Negative
4	Customer service sangat lambat merespons keluhan pengguna.	Negative
5	Aplikasi sering keluar sendiri (*force close*) ketika digunakan.	Negative
6	Cashback yang dijanjikan tidak pernah masuk ke akun saya.	Negative
7	Proses transfer sangat cepat dan biaya administrasinya murah.	Positive
8	Aplikasi mudah digunakan dan tampilannya sangat nyaman.	Positive
9	Pembayaran menggunakan QRIS sangat praktis dan berhasil setiap saat.	Positive
10	Banyak promo menarik sehingga transaksi menjadi lebih hemat.	Positive
11	Proses top-up saldo sangat cepat dan tidak pernah mengalami kendala.	Positive
12	Fitur keamanan sangat baik karena menggunakan verifikasi biometrik.	Positive
13	Customer service membantu dengan cepat saat saya mengalami masalah.	Positive
14	Sangat puas menggunakan aplikasi ini untuk pembayaran sehari-hari.	Positive

C. Text Data Preprocessing

Tahap *text preprocessing* dilakukan untuk mengurangi *noise*, menyeragamkan isi ulasan, serta mempersiapkan data teks agar dapat dikonversi menjadi fitur numerik yang siap digunakan pada proses klasifikasi. Ulasan pengguna dompet digital yang diperoleh dari Google Play Store maupun App Store umumnya mengandung tanda baca, angka, simbol, URL, emoji, singkatan, bahasa gaul, karakter yang berulang, serta penggunaan huruf kapital yang tidak konsisten. Elemen-elemen tersebut dapat menurunkan kualitas representasi fitur apabila tidak diproses terlebih dahulu.
Proses *text preprocessing* pada penelitian ini terdiri atas enam tahapan, yaitu case folding, text cleaning, tokenization, normalization, stopword removal, dan stemming. Tahap case folding mengubah seluruh karakter menjadi huruf kecil (*lowercase*) sehingga kata yang memiliki perbedaan kapitalisasi diperlakukan sebagai kata yang sama. Selanjutnya, text cleaning dilakukan untuk menghapus URL, angka, tanda baca, simbol, emoji, karakter khusus, serta elemen lain yang tidak memberikan informasi penting terhadap proses klasifikasi[9]
Tahap tokenization memisahkan setiap kalimat ulasan menjadi kumpulan kata (*tokens*) sehingga setiap kata dapat diproses secara individual. Setelah itu, dilakukan normalization untuk mengubah kata-kata tidak baku, singkatan, bahasa gaul, dan variasi penulisan menjadi bentuk baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) atau kamus normalisasi yang digunakan. Sebagai contoh, kata *"gk"*, *"ga"*, dan *"nggak"* diubah menjadi *"tidak"*, sedangkan kata *"bgt"* diubah menjadi *"banget"*.
Tahap berikutnya adalah stopword removal, yaitu menghapus kata-kata umum yang sering muncul tetapi memiliki kontribusi yang rendah dalam membedakan kelas sentimen, seperti *yang*, *dan*, *di*, *ke*, *dari*, serta kata-kata umum lainnya. Selanjutnya, dilakukan proses stemming untuk mengubah setiap kata menjadi bentuk dasarnya sehingga variasi kata akibat imbuhan dapat direduksi. Sebagai contoh, kata *"membayar"*, *"dibayar"*, dan *"pembayaran"* akan diubah menjadi kata dasar *"bayar"*.
memperoleh bobot yang lebih tinggi sehingga lebih berkontribusi dalam membedakan kelas sentimen.

Hasil transformasi TF-IDF menghasilkan matriks fitur berdimensi tinggi (*high-dimensional sparse matrix*) yang merepresentasikan karakteristik tekstual dari ulasan pengguna dompet digital. Matriks fitur tersebut selanjutnya digunakan sebagai data masukan (*input features*) pada proses pelatihan dan pengujian algoritma Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), AdaBoost, serta model Stacking Ensemble Machine Learning yang diusulkan dalam penelitian ini.



Gambar 3. Alur *Text Preprocessing* Ulasan Pengguna Dompet Digital

D. Distribusi Data

Sebelum proses pelatihan model dilakukan, distribusi kelas sentimen dianalisis untuk mengetahui apakah terdapat ketidakseimbangan (*class imbalance*) pada dataset. Analisis distribusi data bertujuan untuk memastikan bahwa jumlah data pada masing-masing kelas relatif seimbang sehingga model klasifikasi tidak cenderung berpihak pada salah satu kelas. Distribusi data sentimen pengguna dompet digital disajikan pada Tabel 2.

Tabel 2. Distribusi Data Sentimen Pengguna Dompet Digital

Kelas Sentimen	Jumlah Ulasan	Persentase
Negatif (Class 0)	910	50,56%
Positif (Class 1)	890	49,44%
Total	1.800	100,00%

Berdasarkan Tabel 1, kelas Negatif (Class 0) terdiri atas 910 ulasan atau 50,56% dari keseluruhan dataset, sedangkan kelas Positif (Class 1) terdiri atas 890 ulasan atau 49,44%. Perbedaan jumlah data pada kedua kelas hanya 20 ulasan atau sekitar 1,12%, sehingga distribusi data dapat dikatakan relatif seimbang (*balanced dataset*). Kondisi ini menunjukkan bahwa dataset yang digunakan tidak mengalami permasalahan *class imbalance* yang signifikan.

Meskipun demikian, distribusi data pada setiap subset pelatihan (*training set*) dapat mengalami sedikit perubahan ketika proses validasi dilakukan menggunakan metode 5-Fold Cross-Validation. Oleh karena itu, penelitian ini menerapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) pada setiap data pelatihan untuk menjaga keseimbangan jumlah data antar kelas selama proses pembentukan model.

Penerapan SMOTE dilakukan hanya pada data pelatihan setelah setiap *fold* dibagi menjadi data pelatihan dan data validasi. Data validasi tidak mengalami proses *oversampling* sehingga tetap mempertahankan distribusi data aslinya. Pendekatan ini bertujuan untuk mencegah terjadinya data leakage, yaitu masuknya data sintetis ke dalam proses evaluasi model yang dapat menyebabkan hasil pengujian menjadi bias [10]

Dengan demikian, penggunaan SMOTE pada penelitian ini dimaksudkan sebagai strategi untuk menjaga keseimbangan data selama proses pelatihan model, bukan karena dataset awal memiliki

ketidakseimbangan kelas yang tinggi. Untuk mengetahui kontribusi SMOTE secara lebih komprehensif terhadap peningkatan performa klasifikasi, penelitian selanjutnya dapat melakukan perbandingan antara model yang dilatih menggunakan SMOTE dan model tanpa SMOTE.

E. Hasil K-Fold Cross-Validation

Kinerja model klasifikasi dievaluasi menggunakan metode 5-Fold Cross-Validation. Seluruh dataset yang terdiri atas 1.800 ulasan pengguna dompet digital dibagi menjadi lima subset dengan ukuran yang relatif sama. Pada setiap iterasi, empat subset digunakan sebagai data pelatihan (*training data*), sedangkan satu subset lainnya digunakan sebagai data validasi (*validation data*). Proses ini diulang sebanyak lima kali sehingga setiap subset memperoleh kesempatan yang sama untuk menjadi data validasi[11]

Pada setiap *fold*, proses ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF) dibangun menggunakan data pelatihan. Selanjutnya, apabila digunakan pada penelitian ini, teknik Synthetic Minority Over-sampling Technique (SMOTE) diterapkan hanya pada data pelatihan untuk menjaga keseimbangan kelas, sedangkan data validasi tetap mempertahankan distribusi aslinya. Pendekatan tersebut dilakukan untuk mencegah terjadinya *data leakage* sehingga performa model dievaluasi menggunakan data yang belum pernah dilihat selama proses pelatihan.

Hasil pengujian menggunakan algoritma dasar (*base learner*), yaitu Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost, ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian Algoritma Dasar Menggunakan 5-Fold Cross-Validation

Algoritma	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Rata-rata
SVM	80,00%	80,00%	79,44%	80,83%	78,61%	79,77%
Multinomial Naïve Bayes	79,72%	79,17%	78,61%	80,28%	78,61%	79,27%
AdaBoost	71,39%	71,39%	69,72%	70,00%	71,39%	70,77%

Berdasarkan Tabel 3, algoritma Support Vector Machine (SVM) memperoleh nilai akurasi rata-rata tertinggi sebesar 79,77%. Nilai akurasi pada setiap *fold* berada pada rentang 78,61% hingga 80,83%, dengan performa terbaik diperoleh pada Fold 4 dan performa terendah pada Fold 5. Perbedaan akurasi yang relatif kecil menunjukkan bahwa SVM memiliki tingkat kestabilan yang baik selama proses validasi. Hasil tersebut menunjukkan bahwa SVM mampu membentuk *decision boundary* yang optimal pada ruang fitur TF-IDF yang bersifat berdimensi tinggi (*high-dimensional*) dan jarang (*sparse*), sehingga efektif dalam membedakan sentimen positif dan negatif pada ulasan pengguna dompet digital.

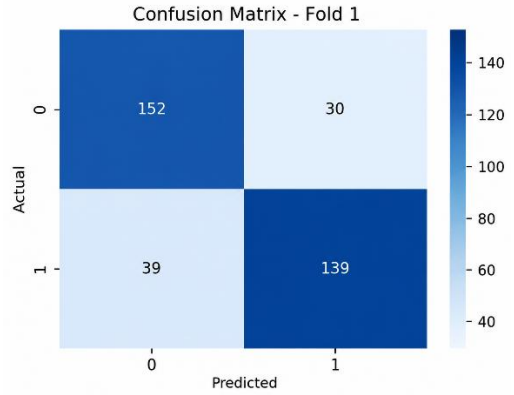
Algoritma Multinomial Naïve Bayes (MNB) memperoleh akurasi rata-rata sebesar 79,27%, hanya berbeda 0,50% dibandingkan dengan SVM. Nilai akurasi yang relatif stabil pada seluruh *fold* menunjukkan bahwa distribusi probabilitas kata yang digunakan oleh MNB cukup efektif dalam mengenali pola sentimen pada ulasan pengguna. Meskipun demikian, asumsi independensi antar fitur yang dimiliki MNB menyebabkan algoritma ini kurang optimal dalam menangkap hubungan antarkata, terutama pada ulasan yang mengandung negasi, kalimat majemuk, atau kombinasi kata yang memiliki makna kontekstual.

Sementara itu, algoritma AdaBoost menghasilkan akurasi rata-rata sebesar 70,77%, yang merupakan nilai terendah dibandingkan dua algoritma lainnya. Performa AdaBoost berada pada rentang 69,72% hingga 71,39%, atau sekitar 9% lebih rendah dibandingkan SVM. Kondisi ini menunjukkan bahwa konfigurasi AdaBoost kurang sesuai untuk representasi fitur TF-IDF yang memiliki dimensi tinggi dan bersifat *sparse*. Pada data ulasan pengguna dompet digital, masih banyak ditemukan kalimat yang mengandung bahasa informal, singkatan, kesalahan penulisan (*typo*), serta opini yang bersifat ambigu. Karakteristik tersebut menyebabkan AdaBoost cenderung memberikan bobot yang lebih besar terhadap sampel yang sulit diklasifikasikan sehingga berpotensi menurunkan kemampuan generalisasi model [12]

F. Pengujian Model Stacking Ensemble

Setelah dilakukan evaluasi terhadap masing-masing algoritma dasar (*base learner*), yaitu Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost, ketiga algoritma tersebut dikombinasikan menggunakan pendekatan Stacking Ensemble Machine Learning dengan Logistic Regression sebagai meta-learner. Pada arsitektur ini, hasil prediksi yang dihasilkan oleh ketiga *base learner* digunakan sebagai meta-features yang selanjutnya dipelajari oleh Logistic Regression untuk menghasilkan prediksi akhir terhadap sentimen pengguna dompet digital. Berbeda dengan metode *majority voting*, pendekatan stacking memungkinkan meta-learner memberikan bobot yang berbeda pada setiap algoritma dasar sesuai dengan tingkat kontribusinya dalam menghasilkan prediksi yang akurat.

Untuk mengevaluasi performa model, digunakan Confusion Matrix yang menggambarkan jumlah prediksi benar maupun salah pada masing-masing kelas sentimen. Hasil *Confusion Matrix* ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix Model Stacking Ensemble pada Klasifikasi Sentimen Pengguna Dompet Digital

Berdasarkan Gambar 4, model Stacking Ensemble mampu mengklasifikasikan sebagian besar ulasan pengguna dompet digital dengan benar. Nilai pada diagonal utama menunjukkan jumlah ulasan yang berhasil diklasifikasikan sesuai dengan kelas sebenarnya, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi (*misclassification*). Sebagian besar ulasan dengan sentimen Negatif (Class 0) berhasil dikenali sebagai negatif, begitu pula sebagian besar ulasan Positif (Class 1) berhasil diprediksi sebagai positif. Namun demikian, masih terdapat beberapa ulasan negatif yang diprediksi sebagai positif (*false positive*) serta beberapa ulasan positif yang diprediksi sebagai negatif (*false negative*). Kesalahan tersebut umumnya disebabkan oleh adanya kalimat yang mengandung bahasa informal, singkatan, opini yang ambigu, negasi, maupun kombinasi sentimen positif dan negatif dalam satu ulasan.

Untuk memperoleh evaluasi yang lebih rinci pada setiap kelas sentimen, dilakukan analisis menggunakan metrik Precision, Recall, dan F1-Score. Hasil evaluasi tersebut disajikan pada Gambar 5.

```
--- Fold 1 ---
Accuracy on Test Set: 0.8083
Classification Report on Test Set:
              precision    recall  f1-score   support

           0       0.80      0.84      0.82       182
           1       0.82      0.78      0.80       178

    accuracy                           0.81       360
   macro avg       0.81      0.81      0.81       360
weighted avg       0.81      0.81      0.81       360
```

Gambar 5. Classification Report Model Stacking Ensemble

Berdasarkan Gambar 5, model Stacking Ensemble memperoleh nilai Accuracy sebesar 80,83% pada *fold* yang diuji. Untuk kelas Negatif (Class 0), model menghasilkan nilai Precision sebesar

0,80, Recall sebesar 0,84, dan F1-Score sebesar 0,82. Nilai precision menunjukkan bahwa 80% dari seluruh ulasan yang diprediksi sebagai negatif memang merupakan ulasan negatif, sedangkan nilai recall menunjukkan bahwa model berhasil mengenali 84% dari seluruh ulasan negatif yang sebenarnya. Nilai recall yang lebih tinggi dibandingkan precision menunjukkan bahwa model cukup sensitif dalam mendeteksi sentimen negatif, meskipun masih terdapat beberapa ulasan positif yang salah diklasifikasikan sebagai negatif.

G. Perbandingan Model Stacking Ensemble dengan Model Individu

Perbandingan performa antara model Stacking Ensemble Machine Learning dan masing-masing algoritma dasar (*base learner*) ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Performa Model Stacking Ensemble dan Algoritma Dasar

Model	Rata-rata Akurasi
Support Vector Machine (SVM)	79,77%
Multinomial Naïve Bayes (MNB)	79,27%
AdaBoost	70,77%
Stacking Ensemble	80,55%

Berdasarkan Tabel 4, model Stacking Ensemble Machine Learning memperoleh nilai akurasi rata-rata tertinggi sebesar 80,55%. Nilai tersebut lebih tinggi dibandingkan Support Vector Machine (SVM) sebesar 79,77%, Multinomial Naïve Bayes (MNB) sebesar 79,27%, dan AdaBoost sebesar 70,77%. Secara berturut-turut, model stacking mampu meningkatkan akurasi sebesar 0,78% dibandingkan SVM, 1,28% dibandingkan MNB, dan 9,78% dibandingkan AdaBoost.

Meskipun peningkatan akurasi terhadap SVM dan Multinomial Naïve Bayes relatif tidak terlalu besar, hasil tersebut menunjukkan bahwa pendekatan Stacking Ensemble mampu menghasilkan performa klasifikasi yang lebih baik dibandingkan penggunaan algoritma tunggal. Hal ini menunjukkan bahwa kombinasi beberapa algoritma pembelajaran dapat memanfaatkan keunggulan masing-masing model sehingga menghasilkan prediksi yang lebih akurat pada klasifikasi sentimen pengguna dompet digital[13]

Peningkatan performa tersebut tidak hanya ditunjukkan melalui nilai akurasi, tetapi juga didukung oleh hasil evaluasi menggunakan Confusion Matrix, Classification Report, dan ROC Curve. Model Stacking Ensemble memperoleh nilai ROC-AUC sebesar 0,8795, yang menunjukkan kemampuan diskriminasi yang baik dalam membedakan ulasan dengan sentimen positif dan negatif. Selain itu, nilai Precision, Recall, dan F1-Score pada kedua kelas sentimen relatif seimbang, sehingga model mampu mempertahankan kualitas prediksi tanpa terlalu berpihak pada salah satu kelas.

Keunggulan utama pendekatan Stacking Ensemble terletak pada kemampuannya menggabungkan karakteristik pembelajaran dari beberapa algoritma yang berbeda. Support Vector Machine (SVM) memiliki kemampuan yang baik dalam membentuk *decision boundary* pada data berdimensi tinggi, Multinomial Naïve Bayes (MNB) efektif dalam memodelkan distribusi probabilitas kata pada data teks, sedangkan AdaBoost mampu meningkatkan perhatian terhadap data yang sulit diklasifikasikan melalui proses *boosting*. Seluruh informasi prediksi tersebut kemudian dipelajari kembali oleh Logistic Regression sebagai meta-learner, sehingga keputusan akhir yang dihasilkan menjadi lebih akurat dan stabil dibandingkan apabila hanya menggunakan satu algoritma.

Secara keseluruhan, hasil penelitian menunjukkan bahwa metode Stacking Ensemble Machine Hasil ini menunjukkan bahwa pendekatan *ensemble learning* dapat menjadi alternatif yang efektif untuk meningkatkan akurasi analisis sentimen serta mendukung pengambilan keputusan berbasis data bagi penyedia layanan dompet digital. Informasi yang diperoleh dari hasil klasifikasi sentimen dapat dimanfaatkan untuk mengevaluasi kualitas layanan, mengidentifikasi permasalahan yang sering dikeluhkan pengguna, meningkatkan pengalaman pengguna (*user experience*), serta menyusun strategi pengembangan fitur dan layanan yang lebih sesuai dengan kebutuhan pelanggan[14]

E-ISSN : 2985-4318
P-ISSN : 2985-5772

KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan Stacking Ensemble Machine Learning mampu memberikan performa klasifikasi sentimen pengguna dompet digital yang lebih baik dibandingkan penggunaan satu algoritma *machine learning* secara individual. Dengan mengombinasikan Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), dan AdaBoost sebagai *base learner*, serta Logistic Regression sebagai *meta-learner*, model yang diusulkan berhasil mengintegrasikan karakteristik pembelajaran dari beberapa algoritma sehingga menghasilkan performa klasifikasi terbaik di antara seluruh model yang dievaluasi. Meskipun peningkatan akurasi dibandingkan algoritma terbaik, yaitu SVM, relatif tidak terlalu besar, hasil tersebut menunjukkan bahwa penggabungan pola prediksi yang saling melengkapi mampu meningkatkan konsistensi klasifikasi dan mengurangi ketergantungan terhadap keterbatasan satu algoritma tertentu.[15]

Evaluasi pada tingkat kelas (*class-level evaluation*) juga menunjukkan bahwa model yang diusulkan mampu mempertahankan keseimbangan performa dalam mengklasifikasikan sentimen positif maupun negatif. Temuan ini menjadi penting karena ulasan pengguna dompet digital umumnya memuat berbagai aspek layanan, seperti kecepatan transaksi, keamanan akun, kemudahan penggunaan aplikasi, proses top-up saldo, transfer dana, pembayaran menggunakan QRIS, program cashback, stabilitas aplikasi, serta kualitas layanan pelanggan (customer service). Dengan demikian, manfaat penelitian ini tidak hanya terletak pada peningkatan akurasi klasifikasi, tetapi juga pada kemampuannya mengubah data ulasan yang tidak terstruktur menjadi informasi yang bernilai bagi penyedia layanan dompet digital. Informasi tersebut dapat dimanfaatkan untuk mengevaluasi kualitas layanan, mengidentifikasi faktor-faktor yang paling sering menimbulkan kepuasan maupun keluhan pengguna, serta mendukung pengambilan keputusan berbasis data dalam pengembangan layanan digital.[16]

Kontribusi metodologis penelitian ini terletak pada penerapan Stacking Ensemble Machine Learning yang mengintegrasikan beberapa algoritma dengan karakteristik pembelajaran yang berbeda, penggunaan Logistic Regression sebagai *meta-learner*, penerapan SMOTE pada setiap data pelatihan, serta evaluasi model menggunakan 5-Fold Cross-Validation dan berbagai metrik evaluasi, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC-AUC. Penerapan SMOTE hanya pada data pelatihan di setiap *fold* membantu meminimalkan risiko data leakage, sehingga hasil evaluasi yang diperoleh lebih merepresentasikan kemampuan generalisasi model terhadap data baru. Selain itu, dokumentasi yang sistematis mengenai tahapan text preprocessing, ekstraksi fitur menggunakan TF-IDF, penyeimbangan data, skema validasi, arsitektur model, serta metrik evaluasi meningkatkan tingkat reproduktibilitas penelitian sehingga lebih mudah direplikasi maupun dikembangkan pada penelitian selanjutnya.[17]

Penelitian berikutnya juga dapat melakukan ablation study untuk menganalisis kontribusi masing-masing *base learner* terhadap performa model Stacking Ensemble. Analisis tersebut akan menunjukkan apakah seluruh algoritma yang digunakan memberikan kontribusi yang signifikan atau terdapat algoritma tertentu yang memiliki pengaruh dominan terhadap peningkatan akurasi. Selain itu, analisis terhadap ulasan yang salah diklasifikasikan (*misclassified reviews*) perlu dilakukan untuk mengidentifikasi pola bahasa yang masih sulit dikenali oleh model, seperti penggunaan negasi, bahasa gaul, singkatan, *typo*, opini campuran, kalimat sarkastik, maupun istilah-istilah khusus yang sering digunakan dalam layanan dompet digital. Hasil analisis tersebut dapat menjadi dasar dalam mengembangkan metode representasi fitur yang lebih baik, misalnya dengan memanfaatkan model berbasis word embedding, FastText, BERT, atau IndoBERT, sehingga akurasi klasifikasi sentimen pada penelitian mendatang dapat ditingkatkan lebih lanjut.

DAFTAR PUSTAKA

[1] H. Ferdiansyah, "Pengembangan Pariwisata Halal di Indonesia Melalui Konsep Smart Tourism," Journal of Sustainable Tourism Research, vol. 4, no. 1, pp. 30–34, 2020, doi: 10.24198/tornare.v2i1.25831.

[2] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[3] N. A. M. Roslan, N. M. Diah, Z. Ibrahim, Y. Munarko, and A. E. Minarno, "Automatic plant recognition using convolutional neural network on Malaysian medicinal herbs: the value of data augmentation," International Journal of Advances in Intelligent Informatics, vol. 9, no. 1, pp. 136–147, Mar. 2023, doi: 10.26555/ijain.v9i1.1076.

[4] P. Lahagun, B. Devkota, S. Giri, and P. Budha, "Machine Learning-Based Social Media Review Analysis for Recommending Tourist Spots," Journal of Engineering and Sciences, vol. 3, no. 1, pp. 45–52, Jun. 2024, doi: 10.3126/jes2.v3i1.66234.

[5] Q. Sui and S. K. Ghosh, "Active Learning for Stacking and AdaBoost-Related Models," Stats (Basel)., vol. 7, no. 1, pp. 110–137, Mar. 2024, doi: 10.3390/stats7010008.

[6] H. Zhu and A. Zhu, "Application Research of the XGBoost-SVM Combination Model in Quantitative Investment Strategy," in International Conference on Systems and Informatics, Kunming: Institute of Electrical and Electronics Engineers Inc., Jan. 2022, pp. 1–7. doi: 10.1109/ICSAI57119.2022.10005355.

[7] J. M. Ahn, J. Kim, and K. Kim, "Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting," Toxins (Basel)., vol. 15, no. 10, pp. 1–15, Oct. 2023, doi: 10.3390/toxins15100608.

[8] Agusviyanda, Hamdani, M. K. Anam, Agustin, and M. I. Wibowo, "Stacking Ensemble Machine Learning Model for Early Detection of Chronic Kidney Disease in Indonesia," Rabit : Jurnal Teknologi dan Sistem Informasi Univrab, vol. 10, no. 2, pp. 1244–1252, Jul. 2025, doi: 10.36341/rabit.v10i2.6501.

[9] M. M. R. Mubarak, Y. H. Chrisnanto, and P. N. Sabrina, "Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce," Enrichment: Journal of Multidisciplinary Research and Development, vol. 1, no. 8, pp. 463–477, Nov. 2023, doi: 10.55324/enrichment.v1i8.69.

[10] H. Rhomadhona and J. Permadi, "Klasifikasi Berita Kriminal Menggunakan NaÃ¯ve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation," Jurnal Sains dan Informatika, vol. 5, no. 2, pp. 108–117, 2019, doi: 10.34128/jsi.v5i2.177.

[11] R. J. Dhanal and V. R. Ghorpade, "Aspect-based sentiment-analysis using topic modelling and machine-learning," International Journal of Electrical and Computer Engineering, vol. 14, no. 6, pp. 6689–6698, Dec. 2024, doi: 10.11591/ijece.v14i6.pp6689-6698.

[12] E. Ikhwan Saputra, M. Khairul Anam, H. Yenni, Hamdani, and A. Zamsuri, "Optimalisasi Algoritma Support Vector Machine pada Aspect-Based Sentiment Analysis Menggunakan GridSearchCV," Zonasi (Jurnal Sistem Informasi), vol. 7, no. 1, pp. 271–279, 2025, doi: 10.31849/zn.v7i1.17800.

[13] M. K. Anam, M. I. Mahendra, W. Agustin, Rahmaddeni, and Nurjayadi, "Framework for Analyzing Netizen Opinions on BPJS Using Sentiment Analysis and Social Network Analysis (SNA)," Intensif, vol. 6, no. 1, pp. 2549–6824, 2022, doi: 10.29407/intensif.v6i1.15870.

[14] B. N. Pikir, M. K. Anam, H. Asnal, Rahmaddeni, and T. A. Fitri, "Sentiment Analysis of Technology Utilization by Pekanbaru City Government Based on Community Interaction in

Social Media," JAIA – Journal Of Artificial Intelligence And Applications, vol. 2, no. 1, pp. 32–40, 2021.

[15] R. S. Putra, W. Agustin, M. K. Anam, L. Lusiana, and S. Yaakub, "The Application of Naïve Bayes Classifier Based Feature Selection on Analysis of Online Learning Sentiment in Online Media," Jurnal Transformatika, vol. 20, no. 1, pp. 44–56, Jul. 2022, doi: 10.26623/transformatika.v20i1.5144.

[16] Junadhi, Agustin, M. Rifqi, and M. K. Anam, "Sentiment Analysis of Online Lectures using K-Nearest Neighbors based on Feature Selection," Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), vol. 11, no. 3, pp. 216–225, Dec. 2022, doi: 10.23887/janapati.v11i3.51531.

[17] A. N. Ulfah, M. K. Anam, N. Y. S. Munti, S. Yaakub, and M. B. Firdaus, "Sentiment Analysis of the Convict Assimilation Program on Handling Covid-19," JUITA: Jurnal Informatika, vol. 10, no. 2, pp. 209–216, 2022, doi: 10.30595/juita.v10i2.12308.