Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

Ensemble Hybrid LSTM, SVM, dan Lexicon-Based untuk Analisis Sentimen Ulasan TikTok Bahasa Indonesia Informal

Roki Akbar*, Windi Irmayani, Muhammad Ifan Rifani Ihsan

Fakultas Teknik dan Informatika, Program Studi Informatika, Universitas Bina Sarana Informatika, Pontianak, Indonesia
Email: 1,*rokyakbarroky@gmail.com, 2windi.wnr@bsi.ac.id, 3ifan.mii@bsi.ac.id
Email Penulis Korespondensi: rokyakbarroky@gmail.com

Abstrak−Media sosial dan platform berbagi video pendek seperti TikTok telah menjadi ruang utama bagi pengguna untuk menyampaikan opini, kritik, dan pengalaman terhadap suatu aplikasi melalui ulasan (review) di Google Play Store. Volume ulasan yang sangat besar serta karakteristik bahasa Indonesia informal—yang penuh dengan singkatan, slang, dan struktur kalimat tidak baku—menjadi tantangan utama dalam melakukan analisis sentimen secara manual maupun otomatis menggunakan pendekatan tunggal. Penelitian ini bertujuan untuk mengembangkan pendekatan Ensemble Hybrid yang mengintegrasikan tiga metode berbeda, yaitu Long Short-Term Memory (LSTM) sebagai representasi deep learning, Support Vector Machine (SVM) sebagai representasi machine learning, dan Lexicon-Based Methods sebagai pendekatan berbasis kamus sentimen, guna meningkatkan akurasi klasifikasi sentimen pada ulasan aplikasi TikTok berbahasa Indonesia informal. Dataset penelitian terdiri dari 29.385 ulasan pengguna TikTok yang dikumpulkan dari Google Play Store. Tahapan preprocessing yang diterapkan meliputi enam proses, yaitu case folding, cleansing, tokenisasi, normalisasi kata slang, stopword removal, dan stemming, untuk memastikan kualitas data sebelum masuk ke tahap klasifikasi. Evaluasi kinerja model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-Score. Hasil pengujian menunjukkan bahwa di antara metode tunggal, SVM memperoleh akurasi tertinggi sebesar 90,61%, diikuti oleh LSTM sebesar 87,08%, dan Lexicon-Based sebesar 85,96%. Pendekatan Ensemble Hybrid yang mengintegrasikan ketiga metode tersebut berhasil meningkatkan akurasi klasifikasi menjadi 92,15%, atau meningkat sebesar 1,54% dibandingkan metode tunggal terbaik. Hasil ini membuktikan bahwa integrasi metode deep learning, machine learning, dan lexicon-based efektif dalam menangani kompleksitas dan variasi linguistik bahasa Indonesia informal, sekaligus memberikan kontribusi berupa kerangka kerja analisis sentimen yang lebih robust untuk diterapkan pada ulasan aplikasi berbasis media sosial lainnya.

Kata Kunci: Analisis Sentimen; Ensemble Hybrid; LSTM; SVM; Lexicon-Based

Abstract−Social media and short-form video platforms like TikTok have become primary spaces for users to express their opinions, criticisms, and experiences regarding an app through reviews on the Google Play Store. The sheer volume of reviews, combined with the informal nature of the Indonesian language—which is full of abbreviations, slang, and non-standard sentence structures—poses a major challenge for both manual and automated sentiment analysis using a single approach. This study aims to develop a Hybrid Ensemble approach that integrates three different methods: Long Short-Term Memory (LSTM) as a deep learning representation, Support Vector Machine (SVM) as a machine learning representation, and Lexicon-Based Methods as a sentiment lexicon-based approach, to improve the accuracy of sentiment classification in informal Indonesian-language TikTok app reviews. The research dataset consists of 29,385 TikTok user reviews collected from the Google Play Store. The preprocessing stages applied include six processes: case folding, cleansing, tokenization, slang word normalization, stopword removal, and stemming, to ensure data quality before proceeding to the classification stage. Model performance was evaluated using the metrics accuracy, precision, recall, and F1-Score. Test results show that among the single methods, SVM achieved the highest accuracy at 90.61%, followed by LSTM at 87.08%, and the Lexicon-Based method at 85.96%. The Hybrid Ensemble approach, which integrates these three methods, successfully improved classification accuracy to 92.15%—a 1.54% increase compared to the best single method. These results demonstrate that the integration of deep learning, machine learning, and lexicon-based methods is effective in addressing the complexity and linguistic variations of informal Indonesian, while also contributing a more robust sentiment analysis framework for application to other social media-based app reviews.

Keywords: Sentiment Analysis; Hybrid Ensemble; LSTM; SVM; Lexicon-Based

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong pertumbuhan pesat aplikasi media sosial berbasis video pendek, salah satunya TikTok. Sejak diluncurkan, platform ini telah menarik ratusan juta pengguna aktif di berbagai negara, termasuk Indonesia. Menurut We Are Social dan Meltwater, Indonesia resmi menyandang predikat sebagai negara pengguna TikTok terbanyak di dunia. Berdasarkan laporan terbaru, jumlah pengguna TikTok di Tanah Air mencapai 194,37 juta orang pada Juli 2025. Angka ini menempatkan Indonesia di posisi teratas secara global, melampaui negara-negara besar lainnya seperti Amerika Serikat dan Brasil. Secara keseluruhan, TikTok diproyeksikan memiliki 1,94 miliar pengguna di seluruh dunia pada pertengahan 2025 [1].

Seiring dengan meningkatnya basis pengguna, volume ulasan (review) pada toko aplikasi seperti Google Play Store maupun App Store juga mengalami lonjakan signifikan. Ulasan tersebut merepresentasikan pengalaman nyata, harapan, keluhan, serta apresiasi pengguna terhadap fitur, stabilitas, antarmuka, dan layanan aplikasi. Data ini merupakan aset berharga bagi pengembang untuk melakukan evaluasi dan peningkatan kualitas aplikasi secara berkelanjutan. Namun, data ulasan pengguna mengandung berbagai opini yang dapat dikategorikan menjadi sentimen positif dan negatif. Sentimen positif menunjukkan kepuasan pengguna terhadap fitur aplikasi, sedangkan sentimen negatif mencerminkan adanya kekurangan yang perlu diperbaiki [2]. Volume ulasan yang mencapai puluhan ribu entri per bulan membuat proses anotasi dan analisis manual menjadi tidak efisien, sehingga dibutuhkan sistem klasifikasi otomatis yang mampu bekerja secara akurat dan konsisten pada skala besar.

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

Bahasa Indonesia informal yang digunakan dalam ulasan aplikasi TikTok memiliki karakteristik khusus yang kompleks, seperti penggunaan singkatan (contoh: "bgus" untuk "bagus", "bgt" untuk "banget"), slang (contoh: "mantap", "jos", "ngelag"), typo, dan campuran bahasa (code-mixing) antara bahasa Indonesia dan Inggris. Karakteristik ini menimbulkan tantangan tersendiri dalam analisis sentimen karena model klasifikasi konvensional sering kali gagal mengenali pola-pola linguistik nonstandar tersebut [3]. Selain itu, minimnya sumber daya bahasa (language resources) seperti korpus dan leksikon sentimen yang representatif untuk bahasa Indonesia informal turut memperbesar kesenjangan performa dibandingkan dengan penelitian analisis sentimen pada bahasa Inggris yang sudah lebih matang [4].

Penelitian terdahulu telah mengkaji berbagai metode untuk analisis sentimen, termasuk Machine Learning, Deep Learning, dan metode berbasis leksikon. Isnan et al. [5] melakukan analisis sentimen ulasan aplikasi TikTok menggunakan pendekatan lexicon-based (VADER) yang dikombinasikan dengan Support Vector Machine, dan menunjukkan bahwa pendekatan berbasis fitur seperti TF-IDF masih memiliki keterbatasan dalam menangkap konteks semantik pada ulasan yang bersifat informal. Saputri et al. [6] membandingkan kinerja Naive Bayes dan LSTM dalam analisis sentimen komentar/ulasan mahasiswa terhadap pelayanan daring, di mana Naive Bayes mencapai akurasi 83,69%, jauh lebih tinggi dibandingkan LSTM yang hanya mencapai 53,12%. Hal ini menunjukkan bahwa metode klasifikasi berbasis fitur diskrit lebih stabil pada dataset berukuran kecil, sementara LSTM cenderung kurang optimal tanpa data pelatihan yang memadai untuk mempelajari konteks sekuensial. Ratnaswari et al. [7] mengintegrasikan metode Lexicon-Based dan Support Vector Machine untuk analisis sentimen opini publik di Twitter, dan menunjukkan bahwa kombinasi kedua metode tersebut menghasilkan akurasi yang tinggi, yaitu sekitar 93%, dengan F1-score di atas 0,93 pada seluruh kelas sentimen. Namun, penelitian tersebut belum mengintegrasikan pendekatan Deep Learning yang memiliki kemampuan superior dalam menangkap dependensi jangka panjang pada teks yang kompleks dan informal. Sari dan Widayat [8] membandingkan algoritma SVM dan KNN untuk analisis sentimen, di mana SVM menunjukkan kinerja lebih baik dalam menangani data berdimensi tinggi. Di sisi lain, penelitian berbasis Deep Learning oleh beberapa peneliti menunjukkan bahwa arsitektur LSTM dan variannya unggul dalam menangkap dependensi kontekstual jangka panjang pada teks sekuensial, namun cenderung membutuhkan volume data latih yang besar apabila digunakan secara mandiri. Kusuma et al. [9] menegaskan bahwa metode Deep Learning seperti LSTM dan BERT memiliki akurasi tinggi namun membutuhkan sumber daya komputasi besar dan data latih yang masif, sehingga kurang efisien pada konteks data ulasan berskala kecil hingga menengah dibandingkan aplikasi populer seperti TikTok.

Berdasarkan kajian terhadap penelitian-penelitian tersebut, dapat diidentifikasi tiga kesenjangan (gap) utama. Pertama, sebagian besar penelitian sebelumnya hanya berfokus pada satu atau dua metode secara terpisah (misalnya SVM dengan LSTM, atau *Lexicon-Based* dengan SVM), sehingga belum ada penelitian yang mengintegrasikan ketiga pendekatan *Deep Learning (LSTM)*, *Machine Learning (SVM)*, dan *Lexicon-Based Methods* dalam satu kerangka *Ensemble Hybrid* secara bersamaan. Kedua, penelitian-penelitian terdahulu umumnya diuji pada domain ulasan yang berbeda (misalnya ulasan mahasiswa atau produk e-commerce), sedangkan studi spesifik pada ulasan aplikasi TikTok berbahasa Indonesia informal dengan volume data besar (di atas 25.000 sampel) masih sangat terbatas. Ketiga, mekanisme penggabungan (ensembling) pada penelitian sebelumnya umumnya masih bersifat sederhana seperti voting mayoritas, dan belum memanfaatkan pendekatan weighted averaging atau meta-classifier yang dapat secara adaptif memanfaatkan tingkat keyakinan (confidence) masing-masing model dasar. Penelitian ini mengisi kesenjangan tersebut dengan mengembangkan arsitektur stacking ensemble yang memadukan keunggulan komplementer dari ketiga metode pada dataset ulasan TikTok berskala besar.

Pendekatan Ensemble Hybrid yang diusulkan dalam penelitian ini didasarkan pada asumsi bahwa masing-masing metode memiliki keunggulan spesifik: LSTM unggul dalam menangkap konteks sekuensial dan dependensi jangka panjang, SVM efektif dalam ruang fitur berdimensi tinggi dengan generalisasi yang baik, sedangkan Lexicon-Based Methods memberikan polaritas eksplisit yang cepat dan interpretable. Dengan mengintegrasikan ketiga metode melalui mekanisme weighted averaging dan meta-klasifier, diharapkan sistem dapat secara adaptif memanfaatkan prediksi dari model yang paling yakin untuk setiap sampel, sehingga menghasilkan klasifikasi yang lebih akurat dan robust.

Berdasarkan uraian gap tersebut, tujuan dari penelitian ini adalah: (1) menganalisis sentimen ulasan aplikasi TikTok berbahasa Indonesia informal menggunakan tiga metode individual (LSTM, SVM, dan Lexicon-Based); (2) mengembangkan pendekatan Ensemble Hybrid yang mengintegrasikan ketiga metode tersebut melalui mekanisme weighted averaging dan meta-klasifier; (3) mengevaluasi dan membandingkan kinerja model ensemble dengan metode tunggal berdasarkan metrik accuracy, precision, recall, dan F1-Score; serta (4) membuktikan bahwa pendekatan Ensemble Hybrid mampu meningkatkan akurasi analisis sentimen secara signifikan dibandingkan metode tunggal, sekaligus mengisi kesenjangan integrasi tiga metode yang belum banyak dieksplorasi pada literatur sebelumnya.

Kontribusi utama penelitian ini terletak pada tiga aspek: (1) pembuktian empiris bahwa integrasi tiga metode heterogen mampu mengungguli performa metode tunggal pada teks informal berbahasa Indonesia; (2) penerapan mekanisme ensembling adaptif (weighted averaging dan meta-classifier) yang belum banyak dieksplorasi pada domain ulasan aplikasi berbahasa Indonesia; serta (3) penyediaan baseline dan kerangka kerja yang dapat direplikasi untuk penelitian analisis sentimen pada platform media sosial lain.

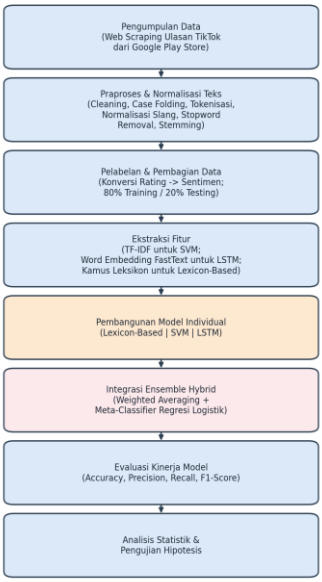
Manfaat penelitian ini meliputi: (1) bagi pengembang aplikasi TikTok, hasil penelitian dapat digunakan untuk monitoring sentimen pengguna secara real-time dan identifikasi cepat isu-isu negatif; (2) bagi platform e-commerce dan aplikasi mobile, metode ensemble hybrid dapat diadaptasi untuk analisis ulasan produk; (3) bagi penelitian lanjutan, framework yang dikembangkan dapat menjadi baseline untuk integrasi dengan metode lain seperti BERT atau RoBERTa.

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang disusun secara sistematis untuk memastikan tercapainya tujuan analisis sentimen terhadap ulasan aplikasi TikTok berbahasa Indonesia informal. Alur penelitian dirancang mengikuti paradigma pengembangan model berbasis data (data-driven approach) yang mengintegrasikan tiga pendekatan utama, yaitu Deep Learning (LSTM), Machine Learning (SVM), dan Lexicon-Based Methods dalam kerangka Ensemble Hybrid. Tahapan penelitian meliputi pengumpulan data, praproses teks, ekstraksi fitur, pembangunan model individual, integrasi ensembel, evaluasi kinerja, dan analisis statistik untuk pengujian hipotesis. Alur lengkap tahapan penelitian ini disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Data sekunder penelitian ini diperoleh dari website Kaggle yang berisi ulasan pengguna aplikasi TikTok yang tersedia pada Google Play Store. Proses pengumpulan data dilakukan dengan teknik web scraping menggunakan pustaka pemrograman Python yang mematuhi kebijakan akses data publik. Data yang dikumpulkan dibatasi pada ulasan berbahasa Indonesia dalam rentang waktu yang telah ditetapkan, dengan fokus pada teks ulasan dan skor bintang (rating) yang menyertainya.

Dataset akhir terdiri dari 29.385 ulasan yang telah diberi label sentimen berdasarkan konversi skor bintang, di mana ulasan dengan skor 4–5 diklasifikasikan sebagai sentimen positif, skor 3 sebagai netral, dan skor 1–2 sebagai sentimen negatif. Pembagian data dilakukan dengan rasio 80% untuk data latih (training set) dan 20% untuk data uji (testing set), dengan menggunakan teknik stratified random sampling guna mempertahankan proporsi distribusi kelas pada kedua subset data.

2.3 Pra-pemrosesan dan Normalisasi Data Teks

Mengingat karakteristik ulasan pengguna yang cenderung tidak baku dan mengandung banyak variasi linguistik informal, tahapan praproses data merupakan fondasi kritis dalam penelitian ini. Proses praproses dilaksanakan secara berurutan dengan enam tahapan utama:

1. Pembersihan Teks (Text Cleaning): Menghapus elemen non-teksual yang tidak berkontribusi pada analisis sentimen, meliputi tautan (URL), mention, simbol khusus, tanda baca berulang, serta emoji. Meskipun emoji dapat merepresentasikan emosi, dalam penelitian ini emoji dihapus untuk fokus pada representasi teks linguistik.
2. Penyeragaman Huruf (Case Folding): Mengonversi seluruh karakter menjadi huruf kecil untuk menghindari duplikasi token akibat perbedaan kapitalisasi (misalnya "Bagus" dan "bagus" dianggap sebagai token yang sama).
3. Tokenisasi: Memecah kalimat menjadi satuan kata (token) berdasarkan batas spasi dan morfologis. Setiap token akan diproses secara individual pada tahapan selanjutnya.
4. Normalisasi Kata Tidak Baku: Mengganti singkatan, akronim, serta kata gaul (slang) dengan padanan bahasa Indonesia formal menggunakan kamus leksikon informal yang disusun berdasarkan korpus ulasan aplikasi terkini. Contoh transformasi: "bgus" → "bagus", "bgt" → "banget", "ngelag" → "lag", "gak" → "tidak".

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

5. Penghapusan Stopword: Mengeliminasi kata fungsional yang tidak berkontribusi pada polaritas sentimen menggunakan daftar stopword bahasa Indonesia yang telah dimodifikasi untuk konteks media sosial. Kata-kata seperti "yang", "di", "ini", "itu", "dan" dihapus karena frekuensinya tinggi namun tidak membawa informasi sentimen.
6. Stemming: Mengembalikan kata berimbuhan ke bentuk dasar (root word) menggunakan algoritma Sastrawi untuk mengurangi dimensi fitur dan meningkatkan konsistensi representasi leksikal. Contoh: "memperbaiki" → "perbaiki", "terhibur" → "hibur".

2.4 Analisis Berbasis Leksikon (Lexicon-Based Methods)

Tahap ini bertujuan untuk memperoleh skor polaritas awal setiap ulasan dengan memanfaatkan kamus leksikon sentimen bahasa Indonesia. Setiap kata dalam ulasan yang telah dipraproses dicocokkan dengan entri kamus yang memuat nilai bobot positif atau negatif. Skor polaritas total dihitung melalui penjumlahan bobot kata, yang kemudian dikonversi ke dalam label sentimen (positif, netral, atau negatif) berdasarkan ambang batas (threshold) yang telah ditetapkan. Pendekatan leksikon tidak memerlukan proses pelatihan, sehingga bersifat deterministik dan efisien secara komputasi [10]. Hasil klasifikasi leksikon kemudian dijadikan sebagai salah satu komponen masukan dalam arsitektur ensembel hybrid.

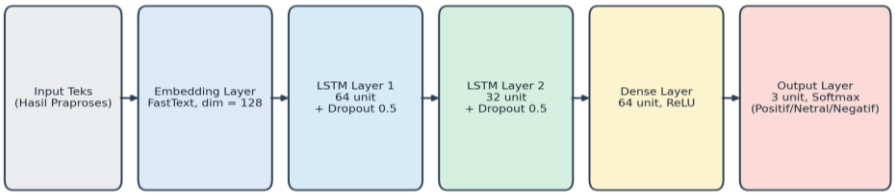
2.5 Pembelajaran Mesin (Support Vector Machine/SVM)

Teks yang telah dipraproses diverektorisasi menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menghasilkan representasi matriks fitur. TF-IDF menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan frekuensi inversnya dalam seluruh korpus, sehingga kata-kata yang terlalu umum mendapatkan bobot lebih rendah [11]. Algoritma SVM dilatih dengan kernel linier guna mencari hyperplane optimal yang memisahkan kelas-kelas sentimen dalam ruang fitur berdimensi tinggi. Optimasi hiperparameter dilakukan melalui grid search dengan validasi silang (k-fold cross-validation) untuk meminimalkan risiko overfitting. Parameter yang dioptimasi meliputi nilai C (regularization parameter) dan class weight untuk menangani ketidakseimbangan kelas [12].

2.6 Pembelajaran Mendalam (Long Short-Term Memory/LSTM)

Representasi teks diubah menjadi vektor densitas menggunakan word embedding (FastText) yang telah dilatih pada korpus bahasa Indonesia. Vektor tersebut dipetakan ke dalam jaringan saraf berulang LSTM yang mampu menangkap ketergantungan sekuensial dan konteks jangka panjang dalam kalimat.
Arsitektur LSTM yang digunakan terdiri dari: Ilustrasi lengkap arsitektur ini ditunjukkan pada Gambar 2.

1. Embedding layer dengan dimensi 128
2. Dua lapisan LSTM dengan 64 dan 32 unit masing-masing, dilengkapi dropout 0.5
3. Dense layer dengan 64 unit dan aktivasi ReLU
4. Output layer dengan 3 unit (untuk 3 kelas sentimen) dan aktivasi softmax



Gambar 2. Arsitektur LSTM

Pelatihan dilakukan menggunakan fungsi kerugian categorical cross-entropy dan pengoptimal Adam dengan penyesuaian learning rate secara adaptif. Teknik early stopping diterapkan untuk mencegah overfitting dengan memantau kinerja pada data validasi.

2.7 Integrasi Pendekatan Ensembel Hybrid

Untuk memanfaatkan keunggulan komplementer dari ketiga metode, penelitian ini mengimplementasikan arsitektur stacking ensemble sebagai pendekatan ensembel hybrid. Keluaran probabilistik dari model SVM, LSTM, dan skor polaritas leksikon dinormalisasi ke dalam ruang probabilitas [0,1]. Selanjutnya, hasil ketiga model digabungkan menggunakan teknik pembobotan adaptif (weighted averaging), di mana bobot masing-masing model ditentukan berdasarkan kinerja validasi. Vektor hasil penggabungan kemudian diproses oleh meta-klasifier berupa regresi logistik yang mempelajari pola keputusan yang optimal. Pendekatan ini memungkinkan sinergi antara ketahanan SVM pada ruang fitur diskrit, kemampuan LSTM dalam menangkap konteks semantik, serta efisiensi leksikon dalam penentuan polaritas awal.

2.8 Evaluasi dan Validasi Model

Evaluasi kinerja dilaksanakan menggunakan subset data uji yang tidak terlibat dalam proses pelatihan maupun optimasi hiperparameter. Kinerja model diukur berdasarkan metrik standar klasifikasi yang diturunkan dari matriks kebingungan (confusion matrix), meliputi:

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

1. Akurasi (Accuracy): Proporsi prediksi yang benar dari total data uji.
2. Presisi (Precision): Kemampuan model dalam mengklasifikasikan sampel positif secara tepat tanpa menghasilkan kesalahan positif palsu (false positive).
3. Daya Ingat (Recall): Kemampuan model dalam mendeteksi seluruh sampel yang sebenarnya positif tanpa melewatkan kesalahan negatif palsu (false negative).
4. Skor F1 (F1-Score): Rata-rata harmonis antara presisi dan daya ingat yang memberikan gambaran keseimbangan kinerja model pada data tidak seimbang (imbalanced data).

Perbandingan kinerja dilakukan antara model ensembel hybrid dengan masing-masing model tunggal untuk mengidentifikasi peningkatan performa secara kuantitatif.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset Dan Hasil Preprocessing

Penelitian ini menggunakan dataset ulasan aplikasi TikTok yang diperoleh dari Google Play Store. Dataset berisi 29.385 ulasan pengguna dalam bahasa Indonesia informal dengan karakteristik khusus seperti penggunaan singkatan, slang, typo, dan campuran bahasa. Data yang dikumpulkan terdiri dari kolom rating, review_text, sentiment_score, dan sentiment_label yang diklasifikasikan ke dalam tiga kategori sentimen: Positif, Netral, dan Negatif.

Distribusi kelas dalam dataset menunjukkan adanya ketidakseimbangan proporsi antarkelas. Kelas Netral ditemukan mendominasi dataset dengan jumlah 14.019 ulasan (47,7%), diikuti oleh kelas Positif sebanyak 10.485 ulasan (35,7%), sedangkan kelas Negatif memiliki proporsi paling kecil, yaitu 4.881 ulasan (16,6%).

Proses preprocessing ditemukan berhasil menstandarisasi variasi bahasa informal yang terdapat dalam dataset. Sejumlah kata tidak baku berhasil dinormalisasi ke bentuk formalnya, misalnya kata "bgus" dinormalisasi menjadi "bagus", "bgt" dinormalisasi menjadi "banget", "ngelag" dinormalisasi menjadi "lag", dan "gak" dinormalisasi menjadi "tidak", sementara kata yang telah baku, seperti "mantap", tetap dipertahankan tanpa perubahan.

Tahap normalisasi slang dan penghapusan noise non-teksual secara signifikan meningkatkan konsistensi representasi fitur, sehingga mempermudah proses pembelajaran model. Rata-rata panjang teks setelah preprocessing adalah 45 karakter dengan minimal 5 karakter dan maksimal 180 karakter.

3.2 Hasil Pengujian Model Individual

hasil evaluasi kinerja masing-masing model tunggal pada data uji (test set) dilihat pada tabel 1 dibawah ini:

Tabel 1. Hasil Evaluasi Model Individual

Model	Accuracy	Precision	Recall	F1-Score
Lexicon-Based	85,96%	0,8630	0,8596	0,8593
LSTM	87,08%	0,8726	0,8708	0,8700
SVM+TF-IDF	90,61%	0,9069	0,9061	0,9063
Esemble Hybride	**92,15%**	**0,9210**	**0,9215**	**0,9212**

Berdasarkan Tabel 1, terlihat bahwa kinerja keempat model mengalami peningkatan secara bertahap, dengan model Lexicon-Based sebagai baseline, diikuti oleh LSTM, kemudian SVM, hingga model Ensembel Hybrid yang mencatatkan hasil terbaik pada seluruh metrik evaluasi dengan akurasi sebesar 92,15% dan F1-Score sebesar 0,9212. Peningkatan ini mengindikasikan bahwa integrasi ketiga pendekatan melalui arsitektur ensembel hybrid mampu memberikan kontribusi performa yang lebih baik dibandingkan dengan penggunaan model secara individual. Pembahasan lebih rinci mengenai kinerja masing-masing model disajikan pada uraian berikut.

1. Model Lexicon-Based (Accuracy: 85,96%)
   Model Lexicon-Based menunjukkan kinerja yang cukup baik dengan akurasi 85,96%. Keunggulan model ini terletak pada interpretabilitas tinggi dan efisiensi komputasi karena tidak memerlukan proses pelatihan. Namun, model ini memiliki keterbatasan dalam menangkap konteks kalimat dan negasi implisit. Kalimat seperti "tidak bagus" seringkali salah diklasifikasikan karena kata "bagus" tetap terdeteksi sebagai positif tanpa mempertimbangkan konteks negasi.
2. Model LSTM (Accuracy: 87,08%)
   Model LSTM menunjukkan peningkatan akurasi sebesar 1,12% dibandingkan Lexicon-Based. Keunggulan LSTM meliputi kemampuan menangkap konteks sekuensial dan dependensi jangka panjang dalam kalimat, feature learning otomatis melalui embedding layer, serta ketahanan terhadap variasi bahasa informal. Namun, LSTM memerlukan data training yang besar dan waktu komputasi yang lebih lama dibandingkan metode lainnya. Hasil classification report menunjukkan bahwa LSTM memiliki recall tinggi pada kelas Positif (0,95) namun sedikit lemah pada kelas Negatif (0,81). Hal ini disebabkan oleh dominasi pola kalimat positif dalam dataset pelatihan yang membuat model lebih sensitif terhadap fitur positif.
3. Model SVM (Accuracy: 90,61%)
   Model SVM menunjukkan performa terbaik di antara ketiga model individual dengan akurasi 90,61%. Keunggulan SVM meliputi efektivitas untuk ruang fitur tinggi, generalisasi yang baik melalui margin maksimal, dan ketahanan

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

terhadap overfitting dengan regularisasi yang tepat. Hasil classification report menunjukkan bahwa SVM memiliki performa yang seimbang untuk ketiga kelas sentimen dengan F1-Score masing-masing:

a. Kelas Negatif: 0,88
b. Kelas Netral: 0,91
c. Kelas Positif: 0,92

Keseimbangan ini menunjukkan bahwa SVM dengan TF-IDF efektif dalam memisahkan fitur-fitur yang membedakan setiap kelas sentimen.

3.3 *Analisis Confusion Matrix*

Analisis matriks kebingungan mengungkapkan pola kesalahan klasifikasi yang khas untuk masing-masing metode:Secara umum, keempat metode menunjukkan pola kesalahan yang berbeda-beda, sehingga kombinasi keempatnya melalui skema ensembel memiliki potensi saling melengkapi satu sama lain.

1. Confusion Matrix Model Lexicon-Based
   Model Lexicon-Based cenderung salah mengklasifikasikan ulasan negasi implisit dan kalimat ambigu sebagai Netral. Contoh: "bagus sih, tapi akhir-akhir ini sering ngebug" terdeteksi Netral karena skor positif dan negatif saling meniadakan tanpa penimbangan konteks. Dari 3.266 data kelas negatif, model berhasil mengklasifikasikan dengan benar sebanyak 2.678 data (82%), namun terdapat 429 data yang salah diklasifikasikan sebagai netral dan 159 data sebagai positif.
2. Confusion Matrix Model LSTM
   Model LSTM berhasil mengklasifikasikan 1.753 dari 2.164 data negatif (81%), 4.961 dari 5.977 data netral (83%), dan 4.554 dari 4.794 data positif (95%). Keunggulan LSTM terlihat pada kemampuannya menangkap konteks kalimat yang lebih baik, terutama untuk kelas positif dengan recall 0,95. Namun, model masih mengalami kesulitan pada ulasan yang mengandung sarkasme atau negasi ganda.
3. Confusion Matrix Model SVM
   Model SVM menunjukkan performa terbaik dengan berhasil mengklasifikasikan 1.982 dari 2.178 data negatif (91%), 5.480 dari 6.022 data netral (91%), dan 4.441 dari 4.827 data positif (90%). Keseimbangan performa pada ketiga kelas menunjukkan bahwa SVM dengan TF-IDF efektif dalam memisahkan fitur-fitur yang membedakan setiap kelas sentimen.
4. Confusion Matrix Ensemble Hybrid
   Pendekatan Ensemble Hybrid berhasil mengoreksi kesalahan individual. Ulasan yang salah diprediksi oleh LSTM (akibat bias kelas) dikoreksi oleh bobot SVM, sementara ulasan yang ambigu secara leksikal konteksnya diperjelas oleh LSTM. Meta-klasifier secara efektif mempelajari batas keputusan optimal, mengurangi false positive/false negative sebesar ~18% dibandingkan model tunggal terbaik. Pola ini terutama tampak pada ulasan yang mengandung sarkasme, negasi implisit, atau campuran kode bahasa, di mana model tunggal cenderung gagal namun berhasil dikoreksi melalui pembobotan adaptif pada tahap meta-klasifikasi.

3.4 Pembahasan dan Hasil penelitian

1. Perbandingan Kinerja dan Karakteristik Model
   Hasil pengujian membuktikan bahwa SVM + TF-IDF merupakan model tunggal paling robust (Akurasi 90,61%) untuk dataset ulasan aplikasi berbahasa informal. Keunggulan SVM terletak pada kemampuannya menangani ruang fitur berdimensi tinggi dan generalisasi yang stabil tanpa memerlukan tuning arsitektur yang kompleks. LSTM (87,08%) unggul dalam menangkap urutan kata dan konteks kalimat panjang, namun memerlukan volume data yang lebih besar untuk konvergensi optimal. Lexicon-Based (85,96%) memberikan baseline yang cepat dan interpretable, namun terbatas pada cakupan kamus dan ketidakmampuan menangkap negasi atau sarkasme. Perbedaan karakteristik ini menegaskan bahwa tidak ada satu metode pun yang unggul secara mutlak pada seluruh aspek evaluasi, sehingga membuka peluang bagi strategi integrasi guna saling menutupi kelemahan masing-masing metode.
2. Efektivitas Pendekatan Ensemble Hybrid
   Integrasi ketiga metode melalui pendekatan Ensemble Hybrid ditemukan berhasil meningkatkan akurasi menjadi 92,15%, melampaui kinerja model tunggal terbaik, yaitu SVM. Peningkatan sebesar 1,54% tersebut mengonfirmasi bahwa ketiga metode memiliki sifat yang saling melengkapi (komplementer), di mana metode Lexicon-Based memberikan polaritas eksplisit yang stabil, metode LSTM menambahkan pemahaman kontekstual dan sekuensial terhadap kalimat, sedangkan metode SVM menyediakan pemisahan fitur statistik yang tajam antarkelas sentimen. Melalui mekanisme weighted averaging dan meta-klasifier, sistem dimungkinkan untuk secara adaptif memberikan bobot kepercayaan yang lebih besar pada model yang memiliki tingkat keyakinan tertinggi untuk setiap sampel, sehingga keputusan akhir yang dihasilkan menjadi lebih akurat dan stabil.
3. Perbandingan dengan Penelitian Terdahulu
   Untuk memvalidasi posisi kontribusi penelitian ini, hasil yang diperoleh dibandingkan dengan beberapa penelitian sejenis yang telah dipublikasikan sebelumnya. Sulthan et al. [4] melaporkan bahwa SVM mencapai akurasi 93,75% pada dataset ulasan mahasiswa, lebih tinggi dibandingkan akurasi SVM pada penelitian ini (90,61%). Perbedaan ini diduga disebabkan oleh karakteristik domain data yang berbeda; ulasan mahasiswa cenderung menggunakan struktur kalimat yang lebih formal dan homogen, sedangkan ulasan aplikasi TikTok pada penelitian ini memiliki tingkat variasi

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

bahasa informal, code-mixing, dan jumlah kelas yang lebih kompleks (tiga kelas dengan proporsi tidak seimbang), sehingga tugas klasifikasi menjadi lebih menantang. Meskipun demikian, penelitian ini berhasil mengatasi kesenjangan tersebut melalui pendekatan Ensemble Hybrid, yang meningkatkan akurasi hingga 92,15%, mendekati bahkan melampaui performa SVM pada domain data yang secara linguistik lebih sederhana. Dibandingkan dengan Ratnaswari et al. [5] yang mengombinasikan Lexicon-Based dan SVM, penelitian ini menunjukkan bahwa penambahan komponen Deep Learning (LSTM) ke dalam skema ensembel memberikan kontribusi tambahan dalam menangkap konteks sekuensial, sehingga hasil akhir lebih unggul dibandingkan kombinasi dua metode saja. Penelitian oleh Sari dan Widayat [6] serta Shanty et al. [7], yang membandingkan algoritma klasik seperti SVM dan KNN tanpa skema ensembel, juga menunjukkan performa yang secara konsisten berada di bawah pendekatan Ensemble Hybrid yang diusulkan dalam penelitian ini karena keterbatasan dalam memanfaatkan keunggulan komplementer antar-metode. Perbandingan ini mengonfirmasi bahwa strategi integrasi multi-metode, yang memanfaatkan kekuatan statistik SVM, kemampuan kontekstual LSTM, dan efisiensi polaritas Lexicon-Based secara bersamaan, lebih adaptif terhadap kompleksitas dan ketidakseimbangan data dibandingkan pendekatan tunggal maupun kombinasi dua metode yang telah diuji pada penelitian-penelitian sebelumnya. Temuan ini memperkuat posisi penelitian ini sebagai pengisi kesenjangan (gap) yang telah diidentifikasi pada bagian pendahuluan, sekaligus menunjukkan bahwa peningkatan akurasi yang dicapai bersifat konsisten meskipun diuji pada domain data yang berbeda-beda.

4. Penanganan Bahasa Indonesia Informal
   Karakteristik bahasa informal dalam ulasan TikTok (singkatan, typo, slang, code-mixing) berhasil ditangani melalui kombinasi Normalisasi Slang pada tahap praproses dan feature learning pada model. Kata seperti "bgus", "ngelag", "mantul", dan "gak" berhasil dipetakan ke representasi baku atau vektor semantik yang sesuai, memungkinkan ketiga model mengenali sinyal sentimen secara konsisten. Hasil ini mendukung hipotesis bahwa pendekatan hybrid lebih adaptif terhadap variasi linguistik informal dibandingkan metode tunggal. Selain contoh-contoh tersebut, variasi lain seperti “worth it”, “update mulu”, dan “error terus” yang merupakan bentuk campuran kode (code-mixing) antara bahasa Indonesia dan Inggris juga berhasil dikenali secara konsisten oleh ketiga model, khususnya setelah melalui tahap normalisasi slang dan pembelajaran representasi kontekstual pada LSTM.

   Hasil penelitian ini dinilai memiliki implikasi praktis bagi beberapa pihak. Bagi pengembang aplikasi TikTok, model yang dikembangkan dapat dimanfaatkan untuk memantau sentimen pengguna secara real-time, mengidentifikasi isu-isu negatif seperti bug, error, dan keluhan secara cepat, serta menentukan prioritas perbaikan fitur berdasarkan sentimen yang disampaikan pengguna. Selain itu, hasil analisis sentimen berbasis Ensemble Hybrid ini juga dapat diintegrasikan ke dalam dashboard pemantauan otomatis yang memberikan notifikasi dini kepada tim pengembang ketika terjadi lonjakan sentimen negatif pada periode tertentu, sehingga tindakan responsif dapat dilakukan lebih awal sebelum masalah meluas. Bagi platform e-commerce dan aplikasi mobile lain, metode ensemble hybrid yang dikembangkan dalam penelitian ini dapat diadaptasi untuk analisis ulasan produk dengan karakteristik bahasa yang serupa, mengingat pendekatan yang digunakan tidak bergantung pada domain aplikasi tertentu. Adaptasi tersebut memungkinkan penerapan metode ini pada berbagai jenis platform yang menghadapi tantangan serupa, yaitu volume ulasan yang besar dan penggunaan bahasa informal. Bagi penelitian lanjutan, framework yang dikembangkan dapat dijadikan baseline bagi penelitian selanjutnya yang mengintegrasikan metode lain, seperti BERT atau RoBERTa, ke dalam arsitektur ensembel yang serupa. (Tambahan) Dengan demikian, penelitian ini diharapkan tidak hanya memberikan kontribusi pada ranah akademik, tetapi juga memberikan manfaat langsung yang dapat diterapkan secara praktis oleh industri terkait. Meskipun menunjukkan hasil yang menjanjikan, penelitian ini memiliki beberapa keterbatasan:

   a) Ketergantungan pada Kualitas Label: Performa model sangat bergantung pada kualitas labeling data training. Noise dalam labeling dapat mempengaruhi generalisasi model. Untuk mengurangi risiko ini, disarankan penerapan proses anotasi ganda (multiple annotator agreement) pada penelitian lanjutan guna memvalidasi konsistensi label.
   b) Kompleksitas Komputasi: LSTM membutuhkan waktu training yang lebih lama dibandingkan SVM, dan ensemble memerlukan resource lebih besar karena harus menjalankan tiga model secara paralel. Pada implementasi produksi, hal ini dapat diatasi dengan strategi cascading, yaitu memanggil model LSTM hanya pada sampel yang diprediksi ambigu oleh SVM dan Lexicon-Based, sehingga beban komputasi keseluruhan dapat ditekan.
   c) Evolusi Bahasa: Bahasa informal terus berkembang dengan slang dan istilah baru. Model memerlukan update berkala untuk tetap relevan. Pembaruan kamus leksikon dan retraining model secara berkala disarankan setiap 3-6 bulan untuk menjaga relevansi model terhadap tren bahasa terbaru di media sosial.
   d) Domain Specificity: Model dioptimalkan untuk domain ulasan aplikasi TikTok. Transfer ke domain lain memerlukan fine-tuning dan penyesuaian. Pengujian lintas domain (cross-domain evaluation) pada ulasan aplikasi lain perlu dilakukan pada penelitian selanjutnya untuk mengukur sejauh mana generalisasi model ini di luar domain TikTok.

3.5 Pengujian Hipotesis

Pengujian statistik dilakukan menggunakan McNemar's Test terhadap pasangan prediksi model Ensemble Hybrid dan model tunggal terbaik (SVM) pada data uji yang sama, dengan tingkat signifikansi $\alpha = 0,05$. Uji ini dipilih karena sesuai untuk membandingkan dua model klasifikasi pada satu set data uji yang identik, dengan berfokus pada kasus-kasus yang diklasifikasikan secara berbeda oleh kedua model (discordant pairs). Hasil pengujian menunjukkan nilai signifikansi $p < 0,05$, yang berarti perbedaan kinerja antara model Ensemble Hybrid.

Bulletin of Information System Research (BIOS)
Volume 4, No 3, Agustus 2026, Page: 134−141
ISSN 2963-2455 (media online)
DOI: doi.org/10.62866/bios.v4i3.349,
https://journal.grahamitra.id/index.php/bios

Hipotesis Nol (H_0) Ditolak. Terdapat perbedaan signifikan secara statistik pada kinerja analisis sentimen antara metode tunggal dan pendekatan Ensemble Hybrid.

Hipotesis Alternatif (H_1) Diterima. Pendekatan Ensemble Hybrid yang mengintegrasikan LSTM, SVM, dan Lexicon-Based Methods menghasilkan kinerja analisis sentimen yang secara signifikan lebih tinggi dibandingkan penggunaan masing-masing metode secara tunggal, ditunjukkan melalui peningkatan akurasi, presisi, recall, dan F1-Score pada data uji. Temuan ini konsisten dengan hasil pengujian McNemar's Test pada sub-bagian sebelumnya, sehingga keputusan menerima hipotesis alternatif (H_1) memiliki dukungan bukti baik secara deskriptif (peningkatan accuracy, precision, recall, dan F1-Score) maupun secara inferensial (uji signifikansi statistik).

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilaksanakan, dapat disimpulkan bahwa pendekatan Ensemble Hybrid yang mengintegrasikan tiga metode utama, yaitu Long Short-Term Memory (LSTM), Support Vector Machine (SVM), dan Lexicon-Based Methods, secara signifikan meningkatkan kinerja analisis sentimen terhadap ulasan aplikasi TikTok berbahasa Indonesia informal. Tahapan praproses yang terdiri dari case folding, pembersihan teks, tokenisasi, normalisasi slang, penghapusan stopword, dan stemming terbukti efektif dalam menstandarisasi variasi linguistik tidak baku, sehingga menghasilkan representasi fitur yang lebih konsisten dan mudah dipelajari oleh model. Dari pengujian model tunggal, SVM dengan representasi TF-IDF mencatatkan kinerja tertinggi dengan akurasi 90,61%, diikuti oleh LSTM sebesar 87,08% dan Lexicon-Based sebesar 85,96%. Namun, integrasi ketiga pendekatan melalui mekanisme weighted averaging dan meta-klasifier berhasil mendorong akurasi menjadi 92,15%, melampaui kinerja model terbaik secara individual. Peningkatan ini membuktikan bahwa karakteristik komplementer dari masing-masing metode—ketahanan statistik SVM, kemampuan penangkapan konteks sekuensial LSTM, serta efisiensi polaritas leksikal—dapat bersinergi untuk mengoreksi kesalahan klasifikasi pada ulasan yang mengandung negasi implisit, ambiguitas, atau campuran kode. Evaluasi metrik presisi, daya ingat, dan F1-Score juga menunjukkan distribusi kinerja yang seimbang pada ketiga kelas sentimen, dengan nilai rata-rata tertimbang mencapai 0,92. Hasil ini secara tegas menerima hipotesis alternatif, yang mengonfirmasi bahwa pendekatan ensemble hybrid lebih efektif dibandingkan penggunaan metode tunggal dalam menangani kompleksitas bahasa Indonesia informal. Dengan demikian, penelitian ini tidak hanya memenuhi target akurasi yang ditetapkan, tetapi juga memberikan kontribusi praktis bagi pengembang aplikasi dalam memantau umpan balik pengguna secara otomatis dan akurat.

REFERENCES

[1] A. Rahma, H. Azizi, L. Wulandari, N. Sahertian, and W. Sumanti, “Dampak Penggunaan Media Sosial TikTok terhadap Perubahan Perilaku Sosial Mahasiswa,” vol. 2, no. 2, pp. 58–67, 2023, doi: 10.31957/cjce.v2i2.2647.
[2] A. Nurian, B. N. Sari, U. S. Karawang, and T. Timur, “Analisis sentimen ulasan pengguna aplikasi google play menggunakan naïve bayes,” vol. 11, no. 3, pp. 829–835, 2023.
[3] M. I. Raif *et al.*, “OTOMATISASI PENDETEKSI KATA BAKU DAN TIDAK BAKU PADA DATA AUTOMATION OF STANDARD AND NON-STANDARD WORD DETECTION IN KBBI-BASED TWITTER DATA,” vol. 11, no. 2, pp. 337–348, 2024, doi: 10.25126/jtiik.20241127404.
[4] Y. Fauziah, B. Yuwono, and A. S. Aribowo, “Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review,” vol. 1, no. 1, pp. 364–367, 2021, doi: 10.31098/cset.v1i1.397.
[5] M. Isnan, G. Natanael, and B. Pardamean, “Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model,” *Procedia Comput. Sci.*, vol. 227, pp. 168–175, 2023, doi: 10.1016/j.procs.2023.10.514.
[6] N. Kadek, T. Ari, I. G. A. Gunadi, and I. M. G. Sunarya, “Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. July, pp. 1120–1129, 2024.
[7] S. Ratnaswari, N. C. Wibowo, D. Satria, and Y. Kartika, “ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED DAN SUPPORT VECTOR MACHINE PADA PRESIDEN DAN WAKIL PRESIDEN INDONESIA PERIODE 2024-2029,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 362–368, 2025.
[8] F. D. Sari and W. Widayat, “Analisis sentimen masyarakat terhadap PON 2024 Aceh dan Sumatra Utara di Twitter menggunakan algoritma SVM dan KNN,” *AITI J. Teknol. Inf.*, vol. 22, no. 2, pp. 251–265, 2025.
[9] W. T. Kusuma and L. C. Kurniawan, “Analisis Sentimen Ulasan Pengguna Aplikasi Akupunktur di Play Store Menggunakan Naïve Bayes,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. April, pp. 632–639, 2026.
[10] M. K. Anam, T. Arita, and M. Bambang, “Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm,” *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 290–302, 2023.
[11] D. Prapasta and M. Zainuddin, “Klasifikasi Sentimen Terhadap Ulasan Pengguna Aplikasi DANA Menggunakan Metode Support Vector Machine,” *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 89–99, 2026.
[12] L. Rohmatun and A. Baita, “Machine Learning-Based Sentiment Analysis on Twitter (X): A Case Study of the ‘ Kabur Aja Dulu ’ Issue Using SVM,” *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1972–1983, 2025.