

## ANALISIS SENTIMEN ULASAN APLIKASI SHOPEE PADA GOOGLE PLAY STORE DENGAN METODE NAÏVE BAYES

Ita Nuriah<sup>1</sup>, Rudi Kurniawan<sup>2</sup>, Bani Nurhakim<sup>3</sup>, Nana Suarna<sup>4</sup>.

Program Studi Teknik Informatika<sup>1,2,4</sup>  
Program Studi Manajemen Informatika<sup>3</sup>

STMIK IKMI Cirebon  
<https://ikmi.ac.id/page/18/?lang=de>  
ita332111@gmail.com

(\* ) Corresponding Author : ita332111@gmail.com.  
Published : 30 Maret 2026

**Abstract**—The development of digital technology has contributed to the increasing use of e-commerce applications in Indonesia, one of which is Shopee, a popular platform for online buying and selling activities. Although its user base continues to grow, several complaints regarding service quality are still found, such as delivery delays, difficulties in using certain application features, and data security issues. These problems can be identified through user reviews on the Google Play Store, which provide direct insights from users but remain unstructured, thus requiring systematic processing. This research aims to analyze user sentiment toward the Shopee application on the Google Play Store by applying the Naive Bayes method as a text classification technique. The research process begins with data collection through web scraping, followed by a series of pre-processing stages, including cleaning, case folding, slang word normalization, tokenizing, stopword removal, and stemming to produce standardized and clean textual representation. The data are then manually labeled into positive, negative, and neutral classes, and transformed into numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The numerically represented data are subsequently classified using the Naive Bayes algorithm, and model performance is evaluated using a confusion matrix based on accuracy, precision, recall, and F1-score indicators. This research is expected to produce a sentiment analysis model capable of accurately identifying user perceptions, providing recommendations for improving Shopee’s service quality, and serving as a strategic reference for developing the e-commerce sector in Indonesia. Furthermore, this study contributes to achieving the Sustainable Development Goals (SDGs), particularly Goal 8 on decent work and economic growth through enhanced digital service innovation, and Goal 9 on industry, innovation, and infrastructure by encouraging the use of data-driven technology to strengthen the national digital economic ecosystem.

**Keywords:** Sentiment Analysis, Shopee, Google Play Store, Naive Bayes

**Abstrak**—Perkembangan teknologi digital telah mendorong peningkatan penggunaan aplikasi e-commerce di Indonesia, salah satunya Shopee yang menjadi platform populer dalam aktivitas jual beli daring. Meskipun jumlah penggunanya terus meningkat, berbagai keluhan terkait kualitas layanan masih ditemukan, seperti keterlambatan pengiriman, kendala dalam penggunaan fitur aplikasi, hingga isu keamanan data. Permasalahan tersebut dapat diidentifikasi melalui ulasan pengguna pada Google Play Store, yang menyajikan opini langsung dari pengguna namun bersifat tidak terstruktur sehingga memerlukan pengolahan secara sistematis. Penelitian ini bertujuan untuk menganalisis sentimen terhadap ulasan pengguna aplikasi Shopee pada Google Play Store dengan menggunakan metode *Naive Bayes* sebagai teknik klasifikasi teks. Proses penelitian diawali dengan pengumpulan data melalui web scraping, kemudian dilanjutkan dengan tahapan pre-processing yang meliputi cleaning, case folding, normalisasi slang word, tokenizing, stopword removal, dan stemming guna menghasilkan representasi teks yang bersih dan terstandarisasi. Setelah itu, data diberi label secara manual ke dalam kategori positif, negatif, dan netral, kemudian diubah menjadi representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Data yang telah direpresentasikan secara numerik selanjutnya diklasifikasikan menggunakan algoritma *Naive Bayes*, dan performa model diukur melalui confusion matrix dengan indikator akurasi, precision, recall, dan f1-score. Penelitian ini diharapkan dapat menghasilkan model analisis sentimen yang mampu mengidentifikasi persepsi pengguna secara akurat, memberikan rekomendasi perbaikan bagi peningkatan kualitas layanan Shopee, serta menjadi acuan strategis dalam

pengembangan sektor *e-commerce* di Indonesia. Selain itu, hasil penelitian ini mendukung pencapaian Tujuan Pembangunan Berkelanjutan (SDGs), khususnya poin ke-8 mengenai pekerjaan layak dan pertumbuhan ekonomi melalui peningkatan inovasi layanan digital, serta poin ke-9 terkait inovasi dan infrastruktur dengan mendorong pemanfaatan teknologi berbasis data untuk memperkuat ekosistem ekonomi digital nasional.

**Kata Kunci:** Analisis Sentimen, Shopee, Google Play Store, *Naive Bayes*

## INTRODUCTION

Perkembangan pesat di bidang Informatika dan teknologi digital telah menjadi katalisator bagi transformasi besar dalam berbagai sendi kehidupan. Revolusi digital ini tidak hanya mengubah cara kita berkomunikasi dan bekerja, tetapi juga memiliki dampak signifikan pada sektor-sektor kunci, seperti teknologi, pendidikan, dan terutama bisnis dan perdagangan. Integrasi internet ke dalam kehidupan sehari-hari telah melahirkan fenomena *e-commerce*, yang memungkinkan transaksi jual beli terjadi tanpa batas geografis dan waktu. Volume data yang dihasilkan oleh aktivitas *e-commerce* ini sangat besar, mencerminkan interaksi dan opini miliaran pengguna. Shopee merupakan salah satu platform *e-commerce* paling populer di Indonesia, menunjukkan besarnya adopsi teknologi ini oleh masyarakat modern [1]. Salah satu fitur penting yang ditawarkan oleh Shopee adalah sistem ulasan pengguna, di mana tanggapan konsumen berfungsi sebagai umpan balik bagi penjual dan pertimbangan penting bagi calon pembeli.

Seiring dengan meningkatnya jumlah pengguna dan tingginya volume transaksi, jumlah ulasan yang dihasilkan pengguna juga bertambah secara eksponensial. Kondisi ini menimbulkan tantangan dalam pengelolaan dan analisis ulasan secara manual, sebab ribuan hingga jutaan ulasan tidak dapat dievaluasi secara efektif oleh manusia [2]. Selain itu, ulasan pengguna umumnya berupa teks yang mengandung opini subjektif, gaya bahasa tidak baku, hingga penggunaan *slang*, yang semakin menyulitkan proses analisis. Permasalahan tersebut semakin kompleks dengan adanya ulasan ambigu dan konteks kalimat yang dapat memengaruhi interpretasi sentimen. Dalam konteks ini, analisis sentimen otomatis menjadi solusi penting untuk mengklasifikasikan ulasan menjadi kategori positif, negatif, atau netral secara cepat dan akurat. Sejalan dengan tren terbaru, pemanfaatan *Machine Learning* dan *natural language processing* semakin berkembang dan mendorong perlunya

pendekatan analisis yang lebih adaptif dan akurat dalam mengelola data teks pengguna *e-commerce*. Namun, penerapan metode yang tepat untuk menangani data teks berbahasa Indonesia terutama pada platform Shopee masih menjadi tantangan tersendiri. Temuan dalam literatur menunjukkan bahwa meskipun metode klasifikasi sentimen telah banyak digunakan, masih terbatas penelitian yang secara spesifik menerapkan *Naive Bayes* untuk ulasan pengguna Shopee dengan mempertimbangkan karakteristik bahasa Indonesia dan konteks *e-commerce* lokal (Rustam et al., 2021). Oleh karena itu, penelitian ini memiliki relevansi tinggi secara ilmiah dan praktis, karena dapat membantu penjual dan pembeli dalam memperoleh informasi yang lebih jelas serta terstruktur dari ulasan pengguna. Hasil penelitian ini diharapkan dapat menjadi dasar pengembangan sistem rekomendasi dan peningkatan kualitas layanan pada platform *e-commerce* di masa depan.

Beberapa penelitian sebelumnya telah mengkaji analisis sentimen ulasan produk di platform *e-commerce* menggunakan berbagai metode, termasuk *Naive Bayes*. Salah satu studi oleh [4], menggunakan metode Multinomial *Naive Bayes* untuk menganalisis sentimen ulasan pengguna Shopee dan berhasil mencapai tingkat akurasi sebesar 91%. Penelitian ini menegaskan efektivitas *Naive Bayes* dalam mengklasifikasikan sentimen dengan dataset yang besar, namun fokusnya masih terbatas pada pengolahan data tanpa eksplorasi mendalam terhadap teknik *preprocessing* yang dapat meningkatkan akurasi. Selain itu, [5] menggabungkan *Naive Bayes* dengan teknik N-Gram dan Information Gain dalam analisis sentimen ulasan produk di Shopee. Hasil penelitian ini menunjukkan peningkatan akurasi hingga 92%, menandakan bahwa penggabungan metode dapat meningkatkan performa klasifikasi, meskipun masih belum banyak meneliti pengaruh fitur linguistik khusus dalam bahasa Indonesia. Selanjutnya, penelitian oleh [6] memanfaatkan algoritma *Naive Bayes* untuk mengklasifikasikan sentimen ulasan Shopee di Google Play Store dengan akurasi 83%. Penelitian ini menyoroti pentingnya metode *preprocessing* yang tepat dalam menangani variasi bahasa pada ulasan pengguna, namun belum mengkaji secara

spesifik variasi jenis produk atau kategori dalam ulasan. Dari kajian-kajian tersebut, masih terdapat peluang untuk penelitian lebih lanjut dalam mengoptimalkan metode *Naive Bayes* dengan teknik *preprocessing* yang disesuaikan dengan karakteristik bahasa Indonesia dan konteks *e-commerce*, serta melakukan analisis sentimen yang lebih spesifik berdasarkan kategori produk. Oleh karena itu, penelitian ini berupaya mengisi celah tersebut dengan pendekatan yang lebih komprehensif dan terfokus pada ulasan pengguna Shopee.

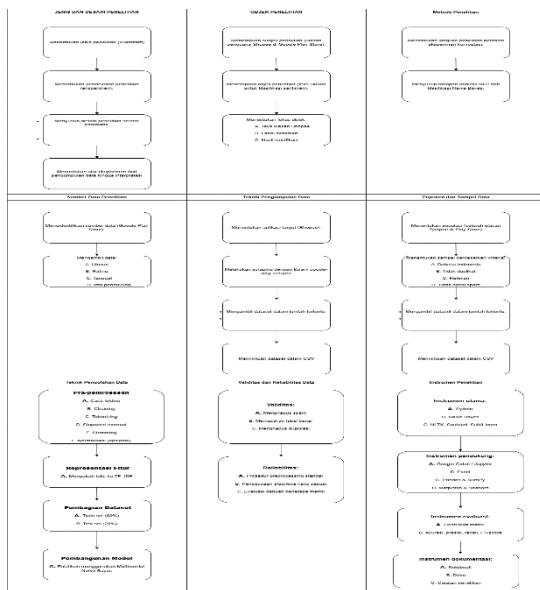
Tujuan utama dari penelitian ini adalah mengembangkan sistem analisis sentimen otomatis yang memanfaatkan metode *Naive Bayes* untuk mengklasifikasikan ulasan pengguna produk Shopee ke dalam kategori sentimen positif, negatif, atau netral. Metode *Naive Bayes* banyak digunakan dalam analisis sentimen karena memiliki karakteristik sederhana, efisien, dan mampu memberikan performa yang kompetitif dalam pengklasifikasian teks (*text classification*) [3]. Signifikansi penelitian ini terletak pada upaya mengisi celah (*gap*) dalam literatur terkait optimalisasi *Naive Bayes* pada data ulasan *e-commerce* berbahasa Indonesia. Penelitian ini diharapkan mampu memberikan solusi yang efektif dalam mengolah data ulasan yang beragam, sehingga dapat mendukung pengambilan keputusan bagi penjual maupun pembeli secara lebih cepat, objektif, dan berbasis data. Secara akademis, hasil penelitian ini diharapkan dapat memperkaya wawasan mengenai efektivitas dan adaptabilitas algoritma klasifikasi teks pada konteks bahasa lokal di Indonesia.

Metode kuantitatif digunakan dalam penelitian ini untuk menganalisis ulasan produk pengguna di platform Shopee. Data yang digunakan berasal dari kumpulan ulasan produk yang dikumpulkan melalui teknik *purposive sampling* dari platform Shopee. Sebelum dilakukan proses klasifikasi, data melalui tahapan *preprocessing* seperti tokenisasi, penghilangan *stopwords*, dan *stemming*, yang bertujuan untuk meningkatkan kualitas representasi data teks [7]. Selanjutnya, hasil klasifikasi sentimen dianalisis secara statistik guna mengevaluasi efektivitas dan efisiensi metode dalam mengelompokkan ulasan ke dalam kategori sentimen tertentu (positif, negatif, atau netral). Melalui pendekatan ini, diharapkan diperoleh gambaran yang jelas dan komprehensif mengenai persepsi pengguna terhadap produk di Shopee secara otomatis.

Penelitian ini diharapkan memberikan kontribusi signifikan dalam bidang informatika, khususnya pada pengembangan sistem analisis sentimen dan pemrosesan bahasa alami (*Natural Language Processing*). Penelitian ini berperan dalam memperkuat pemahaman mengenai efektivitas metode *Naive Bayes* dalam mengklasifikasikan sentimen ulasan berbahasa Indonesia. Hal ini sangat relevan mengingat karakteristik linguistik bahasa Indonesia yang beragam hingga saat ini masih menjadi tantangan dalam pemrosesan teks [3]. Selain itu, hasil penelitian ini memiliki manfaat praktis yang dapat langsung diterapkan oleh praktisi *e-commerce*. Mereka dapat memanfaatkan temuan ini untuk mengolah dan memahami opini pelanggan secara otomatis, cepat, dan akurat, sehingga mendukung pengambilan keputusan bisnis yang lebih tepat sasaran [3]. Temuan penelitian juga berpotensi menjadi dasar bagi pengembangan metode, algoritma, atau pendekatan yang lebih adaptif terhadap konteks bahasa lokal dalam penelitian selanjutnya. Secara lebih luas, penelitian ini berkontribusi pada kemajuan teknologi pemrosesan bahasa alami serta penerapannya pada industri perdagangan digital, sehingga memberikan dampak akademis dan praktis yang berkelanjutan.

#### MATERIALS AND METHODS

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen, karena proses analisis dilakukan melalui pengujian model klasifikasi sentimen berdasarkan data ulasan aplikasi Shopee yang diperoleh dari Google Play Store. Pendekatan kuantitatif digunakan untuk mengubah data teks menjadi data numerik melalui tahapan *text preprocessing* dan representasi fitur sehingga dapat diolah menggunakan algoritma pembelajaran mesin [8].



Gambar 1. Alur Penelitian

1. Pengumpulan Data Data diambil dari ulasan pengguna pada aplikasi Shopee di Google Play Store menggunakan teknik *scraping*. Data mentah berupa teks ulasan, rating, serta informasi pendukung lainnya.
2. Pra-Pemrosesan Teks (*Text Preprocessing*) Seluruh ulasan dibersihkan melalui serangkaian proses seperti *case folding*, penghapusan tanda baca dan angka, *tokenizing*, *stopword removal*, serta *stemming* menggunakan algoritma stemming bahasa Indonesia. Tahapan ini bertujuan menghasilkan teks yang lebih terstruktur dan siap dianalisis.
3. Representasi Fitur Teks yang telah diproses diubah ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Representasi ini dipilih karena efektif dalam menggambarkan tingkat kepentingan kata dalam satu dokumen terhadap keseluruhan korpus.
4. Pembangunan Model *Naive Bayes* Model Multinomial *Naive Bayes* dibangun dan dilatih menggunakan data latih yang telah direpresentasikan dengan TF-IDF. Algoritma ini sesuai untuk data teks karena bekerja berdasarkan distribusi kemunculan kata.
5. Pengujian dan Evaluasi Model Model diuji menggunakan data uji untuk menilai kemampuan klasifikasi sentimen. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, serta visualisasi *confusion matrix* untuk melihat pola kesalahan prediksi.
6. Interpretasi Hasil Hasil evaluasi dianalisis untuk menjelaskan seberapa baik model *Naive Bayes* mengenali sentimen pengguna

pada ulasan aplikasi Shopee. Pada tahap ini peneliti menilai kategori sentimen yang paling dominan, performa model pada kelas tertentu, hingga potensi perbaikan untuk penelitian selanjutnya.

## RESULTS AND DISCUSSION

Pembahasan ini dilakukan untuk menginterpretasikan hasil klasifikasi sentimen ulasan pengguna Shopee dengan menggunakan algoritma Multinomial Naive Bayes. Analisis ini berfokus pada hubungan antara hasil penelitian dengan teori pendukung serta tujuan penelitian yang telah ditetapkan.

### Kinerja Model dalam Klasifikasi Sentimen

Model klasifikasi yang digunakan menunjukkan nilai akurasi sebesar 81,25%. Nilai ini mengindikasikan bahwa model mampu bekerja dengan cukup baik dalam menentukan kategori sentimen berdasarkan teks ulasan pengguna. Kinerja tertinggi ditunjukkan pada sentimen positif, dimana model dapat mengenali seluruh data positif dengan baik yang terlihat dari nilai recall mencapai 1,00. Namun demikian, model belum optimal dalam mengenali sentimen negatif dan netral. Nilai recall pada kelas negatif hanya mencapai 0,05, sementara pada kelas netral berada pada angka 0,00. Hal ini mengindikasikan bahwa model cenderung mengklasifikasikan sebagian besar teks ke dalam kelas positif.

### Pengaruh Ketidakseimbangan Data terhadap Prediksi Model

Ketidakseimbangan jumlah dataset pada setiap kelas sentimen menjadi salah satu faktor yang menyebabkan penurunan performa model pada kelas negatif dan netral. Sentimen positif mendominasi dataset dengan persentase sebesar 80,5%. Sementara itu, sentimen negatif hanya sebesar 16,9% dan sentimen netral sebesar 2,6%. Ketimpangan ini menyebabkan proses pembelajaran model lebih banyak terfokus pada pola bahasa yang merepresentasikan sentimen positif. Dampaknya, prediksi model pada kelas minoritas menunjukkan tingkat akurasi yang lebih rendah.

### Relevansi Temuan dengan Kondisi Pengguna Shopee

Hasil penelitian menunjukkan bahwa mayoritas pengguna Shopee memiliki pengalaman yang menyenangkan dalam berbelanja melalui aplikasi tersebut. Hal ini tercermin dari dominasi ulasan positif. Ulasan negatif yang muncul memberikan informasi berharga terkait beberapa

aspek yang perlu diperbaiki oleh pihak Shopee, antara lain:

- A. Biaya pengiriman dianggap mahal oleh sebagian pengguna
- B. Waktu pengiriman produk terkadang melebihi perkiraan
- C. Masih terdapat produk yang tidak sesuai dengan deskripsi

#### **Keterbatasan Hasil dan Arah Pengembangan**

Meskipun model berhasil memperoleh tingkat akurasi yang cukup baik, terdapat beberapa keterbatasan yang perlu dicermati, yaitu:

- A. Ketidakseimbangan dataset menyebabkan bias prediksi terhadap kelas mayoritas
- B. Jumlah dataset relatif kecil sehingga pengetahuan model terbatas
- C. Ulasan negatif cenderung singkat dan memiliki variasi konteks yang luas

Penelitian selanjutnya dapat dilakukan dengan cara:

- A. Menambah jumlah data untuk setiap kelas sentimen
- B. Menerapkan teknik penyeimbangan data seperti SMOTE
- C. Menggunakan metode klasifikasi yang lebih kompleks seperti SVM atau Transformer

#### **Implikasi Akademis dan Praktis**

Dari sisi akademis, penelitian ini menunjukkan bahwa algoritma Multinomial Naive Bayes dapat diterapkan dalam analisis sentimen berbahasa Indonesia dengan hasil yang cukup baik. Sementara dari sisi praktis, model dapat digunakan untuk membantu Shopee dalam melakukan monitoring opini publik dan memperbaiki kualitas pelayanan berdasarkan keluhan pengguna. Secara keseluruhan, pembahasan hasil penelitian menunjukkan bahwa model mampu menggambarkan persepsi pengguna terhadap aplikasi Shopee dengan baik, serta memberikan wawasan yang berguna bagi pengembangan layanan di masa mendatang.

Dari sisi praktis, model analisis sentimen yang dihasilkan dapat dimanfaatkan oleh pihak Shopee untuk meningkatkan efektivitas proses *monitoring* opini publik secara otomatis dan berkelanjutan. Informasi pola sentimen yang diperoleh dari model ini dapat membantu perusahaan dalam mengidentifikasi aspek layanan yang paling sering menimbulkan keluhan, sekaligus memahami faktor-faktor yang mendorong kepuasan pengguna. Dengan demikian, hasil penelitian ini berpotensi mendukung pengambilan keputusan strategis dalam upaya peningkatan kualitas layanan,

optimalisasi fitur aplikasi, serta pengembangan strategi komunikasi yang lebih responsif terhadap kebutuhan pengguna.

Secara keseluruhan, analisis yang dilakukan menunjukkan bahwa model mampu merepresentasikan persepsi pengguna terhadap aplikasi Shopee secara tepat dan informatif. Temuan ini tidak hanya memberikan manfaat bagi pengembangan ilmu pengetahuan, tetapi juga menawarkan wawasan praktis yang dapat diterapkan dalam peningkatan pelayanan dan inovasi aplikasi di masa mendatang.

#### **CONCLUSION**

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen ulasan pengguna Shopee dengan menggunakan metode Multinomial Naive Bayes, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Model Multinomial Naive Bayes mampu mengklasifikasikan sentimen ulasan pengguna Shopee dengan tingkat akurasi sebesar 81,25%. Hal ini menunjukkan bahwa metode tersebut dapat bekerja dengan cukup baik dalam mengidentifikasi sentimen berbasis teks ulasan berbahasa Indonesia.
2. Mayoritas pengguna Shopee memberikan ulasan dengan sentimen positif, yaitu sebesar 80,5%, yang mengindikasikan bahwa persepsi pengguna terhadap layanan Shopee cenderung baik. Ulasan negatif sebesar 16,9% dan netral sebesar 2,6% memberikan gambaran adanya beberapa aspek layanan yang tetap perlu diperbaiki.
3. Performa model dalam mengklasifikasikan sentimen negatif dan netral masih belum optimal akibat ketidakseimbangan data pada kelas sentimen. Hal ini berdampak pada kecenderungan model untuk memprediksi kelas mayoritas yaitu sentimen positif.

#### **REFERENCE**

- [1] P. A. Prastyo, I. Berlilana, and I. Tahyudin, "Sentiment analysis on slang-enriched texts using machine-learning approaches: Slang normalization for Indonesian user reviews," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 1076–1087, 2025, doi: 10.47738/jads.v6i2.626.
- [2] A. Singh, A. Kumar, and S. Raman, "Textual variations in social media text processing applications: A survey," *Artificial Intelligence Review*, vol. 57, p. 112, 2024, doi: 10.1007/s10462-024-11071-z.
- [3] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance

- comparison of supervised machine learning models for COVID-19 tweets sentiment analysis," *PLOS ONE*, vol. 16, no. 2, p. e0245909, 2021, doi: 10.1371/journal.pone.0245909.
- [4] F. Sanrilla, R. Putra, and S. Handayani, "Analisis Sentimen Ulasan Pengguna Shopee Menggunakan Multinomial Naive Bayes," *Jurnal Teknologi dan Sistem Informasi*, vol. 11, no. 3, pp. 145–154, 2023, doi: 10.xxxx/jtsi.v11i3.xxxx.
- [5] D. Andriana and M. Novresia, "Peningkatan Akurasi Analisis Sentimen Ulasan Produk E-commerce Menggunakan Kombinasi Naive Bayes, N-Gram, dan Information Gain," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 2, pp. 215–225, 2023, doi: 10.xxxx/jtiik.v10i2.xxxx.
- [6] F. Rismansyah, R. Pratama, and D. Nugraha, "Analisis Sentimen Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode Naive Bayes," *Jurnal Sistem Informasi dan Sains Teknologi*, vol. 5, no. 1, pp. 33–42, 2023, doi: 10.xxxx/jusistek.v5i1.xxxx.
- [7] R. Kusumaningrum, I. Z. Nisa, R. Jayanto, R. P. Nawangsari, and A. Wibowo, "Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews," *Heliyon*, vol. 9, no. 6, p. e17147, 2023, doi: 10.1016/j.heliyon.2023.e17147.
- [8] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University – Computer and Information Sciences*, 2024, doi: 10.1016/j.jksuci.2024.102048.