

Analisis Sentimen Penggunaan Cekat.AI dalam Menggantikan Customer Service Menggunakan Logistic Regression dan TF-IDF

Gede Yudha Ardyaputra^{1,*}, Ida Bagus Nyoman Pascima², Putu Hendra Suputra³

¹ Fakultas Teknik dan Kejuruan, Prodi Sistem Informasi, Universitas Pendidikan Ganesha, Bali, Indonesia

² Fakultas Teknik dan Kejuruan, Prodi Pendidikan Teknik Informatika, Universitas Pendidikan Ganesha, Bali, Indonesia

³ Fakultas Teknik dan Kejuruan, Department of Informatics, Universitas Pendidikan Ganesha, Bali, Indonesia

Email: ^{1,*}yudha.ardya Putra@student.undiksha.ac.id, ²gus.pascima@undiksha.ac.id, ³hendra.suputra@undiksha.ac.id

Email Penulis Korespondensi: yudha.ardya Putra@student.undiksha.ac.id

Submitted: 27/01/2026; Accepted: 21/02/2026; Published: 21/02/2026

Abstrak—Perkembangan kecerdasan buatan (AI) telah mendorong perubahan signifikan dalam sistem layanan pelanggan, salah satunya melalui pemanfaatan chatbot sebagai pengganti customer service manusia. Cekat.AI merupakan salah satu inovasi teknologi lokal yang mulai banyak diadopsi oleh perusahaan di Indonesia. Namun, implementasi teknologi ini memicu respons publik yang beragam di media sosial, terutama pada platform X dan TikTok. Permasalahan utama dalam penelitian ini adalah bagaimana persepsi dan sentimen pengguna terhadap penggunaan Cekat.AI dalam menggantikan peran customer service, serta faktor-faktor apa yang melatarbelakangi terbentuknya sentimen tersebut. Penelitian ini bertujuan untuk menganalisis sentimen publik menggunakan metode Logistic Regression dengan ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF) terhadap data komentar dari media sosial X dan TikTok. Untuk mengatasi ketidakseimbangan data sentimen, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa sentimen negatif mendominasi dengan proporsi sebesar 52,5%, diikuti sentimen positif sebesar 35,1% dan netral sebesar 12,4%. Penerapan SMOTE berhasil meningkatkan Recall kelas Netral dari 18,6% menjadi 64,1% dengan akurasi validasi silang mencapai 79,98%. Selain itu, pemodelan topik mengungkap bahwa sentimen negatif terutama dipicu oleh kecemasan terhadap otomatisasi dan ancaman kehilangan pekerjaan. Temuan ini menunjukkan bahwa tantangan utama adopsi Cekat.AI terletak pada penerimaan sosial, bukan semata aspek teknis. Penelitian ini memberikan kontribusi ganda, yaitu secara teknis membuktikan efektivitas SMOTE dalam menangani ketidakseimbangan ekstrem pada data teks bahasa Indonesia, dan secara praktis mengungkap bahwa resistensi publik terhadap AI lokal berakar pada kecemasan kehilangan pekerjaan, bukan semata pada aspek teknis layanan.

Kata Kunci: Analisis Sentimen; Cekat.AI; Logistic Regression; TF-IDF; Media Sosial

Abstract—The rapid development of artificial intelligence has significantly transformed customer service systems, particularly through the use of chatbots to replace human customer service agents. Cekat.AI is one of Indonesia's local AI-based chatbot innovations that has been increasingly adopted by companies. However, its implementation has generated diverse public reactions on social media platforms, especially X and TikTok. The main problem addressed in this study is how users perceive and respond sentimentally to the use of Cekat.AI as a replacement for human customer service, as well as the underlying factors influencing these sentiments. This study aims to analyze public sentiment using the Logistic Regression method with Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction on social media comments from X and TikTok. To address class imbalance in sentiment data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. The results indicate that negative sentiment dominates at 52.5%, followed by positive sentiment at 35.1% and neutral sentiment at 12.4%. The implementation of SMOTE significantly improved the Recall of the neutral class from 18.6% to 64.1%, with a cross-validation accuracy of 79.98%. Topic modeling further reveals that negative sentiment is primarily driven by automation anxiety and concerns over job displacement. These findings suggest that the main challenge in adopting Cekat.AI lies in social acceptance rather than technical performance. This study provides a dual contribution, namely technically proving the effectiveness of SMOTE in handling extreme imbalance in Indonesian text data, and practically revealing that public resistance to local AI is rooted in job displacement anxiety, not merely technical service aspects.

Keywords: Sentiment Analysis; Cekat.AI; Logistic Regression; TF-IDF; Social Media

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong pemanfaatan kecerdasan buatan (*Artificial Intelligence*) dalam berbagai sektor layanan, termasuk layanan pelanggan (*customer service*). *Chatbot* berbasis AI kini digunakan secara luas untuk merespons pertanyaan, menangani keluhan, dan memberikan informasi secara *real-time* melalui media sosial. Transformasi ini bertujuan meningkatkan efisiensi, konsistensi layanan, serta ketersediaan layanan tanpa batas waktu. Namun, penerapan chatbot sebagai pengganti peran manusia memunculkan persoalan baru, khususnya terkait persepsi dan penerimaan pengguna. Media sosial menjadi ruang utama bagi masyarakat untuk mengekspresikan pengalaman dan opini mereka terhadap layanan berbasis AI. Sejalan dengan hal tersebut, media sosial menyimpan informasi yang sangat kaya mengenai kondisi dan perilaku pengguna, di mana analisis terhadap pola polaritas sentimen tidak hanya berfungsi untuk klasifikasi opini tetapi juga mampu mengungkap karakteristik pengguna yang lebih mendalam, termasuk kecenderungan perilaku dan kondisi psikologis tertentu [1] [2]. Perbedaan respons pengguna menunjukkan bahwa adopsi *chatbot* tidak hanya merupakan persoalan

teknis, tetapi juga berkaitan dengan aspek sosial dan emosional. Oleh karena itu, diperlukan kajian ilmiah untuk memahami kecenderungan sentimen publik terhadap penggunaan *chatbot*.AI sebagai layanan pelanggan.

Salah satu inovasi *chatbot* yang berkembang di Indonesia adalah Cekat.AI, yang dirancang untuk mengelola interaksi pelanggan melalui berbagai kanal media sosial secara terintegrasi. Implementasi Cekat.AI sebagai pengganti *customer service* manusia menimbulkan diskursus publik yang cukup intens, khususnya di platform X dan TikTok. Sebagian pengguna menilai *chatbot* memberikan kemudahan dan kecepatan layanan, sementara sebagian lainnya menyoroti keterbatasan empati dan fleksibilitas respons. Perbedaan pandangan ini menunjukkan adanya spektrum sentimen yang beragam, mulai dari positif, negatif, hingga netral. Dalam konteks tersebut, analisis sentimen menjadi pendekatan yang relevan untuk memetakan opini publik secara sistematis. Analisis sentimen memungkinkan pengelompokan opini pengguna berdasarkan polaritas emosional yang terkandung dalam teks. Dalam konteks implementasi teknologi digital dalam layanan publik, diperlukan penilaian yang komprehensif untuk memastikan keberterimaan sistem di masyarakat. Pengukuran efektivitas sistem merupakan langkah krusial dalam menentukan keberhasilan implementasi teknologi berbasis mobile dan digital. Analisis terhadap respons pengguna melalui media sosial dapat diposisikan sebagai instrumen validasi yang penting untuk menilai apakah transisi layanan ke arah otomatisasi Cekat.AI berjalan efektif atau justru menimbulkan resistensi sosial. Analisis respons pengguna di media sosial dapat dijadikan instrumen validasi yang andal, mengingat pola linguistik dalam konten unggahan terbukti memiliki korelasi kuat dengan kondisi emosional dan persepsi subjektif penggunaannya [3]. Dengan demikian, pendekatan ini dapat digunakan untuk mengevaluasi sejauh mana Cekat.AI diterima oleh masyarakat sebagai layanan pelanggan berbasis AI [4].

Penelitian mengenai analisis sentimen telah banyak dilakukan dengan berbagai metode dan objek kajian. Penggunaan TF-IDF dan SMOTE pada ulasan layanan kesehatan terbukti mampu meningkatkan kinerja klasifikasi sentimen secara efektif. Optimasi parameter Support Vector Machine menggunakan algoritma genetika juga dilaporkan dapat meningkatkan performa klasifikasi dalam analisis sentimen pada media sosial Instagram. Selain itu, penerapan KNN dan SMOTE menegaskan pentingnya penanganan ketidakseimbangan kelas dalam analisis sentimen Twitter. Di bidang pendidikan, algoritma *Naïve Bayes* menunjukkan tingkat akurasi yang relatif stabil dalam mengklasifikasikan sentimen pengguna. Berbagai penelitian tersebut membuktikan bahwa analisis sentimen efektif untuk memahami opini publik berbasis data media sosial. Namun, sebagian besar studi masih berfokus pada isu pendidikan, kesehatan, atau layanan umum, belum secara spesifik mengkaji *chatbot* AI sebagai pengganti *customer service* [5]. Selain metode klasifikasi sentimen, beberapa penelitian juga menyoroti pentingnya pendekatan lain untuk memahami konteks percakapan publik. Pendekatan *topic modeling* terbukti mampu mengidentifikasi pola dan tren utama dalam berbagai diskursus digital, sehingga tidak hanya menilai polaritas sentimen tetapi juga mengungkap isu-isu dominan yang dibicarakan oleh pengguna. Dalam konteks *chatbot*, pemahaman topik perbincangan menjadi penting untuk mengetahui aspek layanan apa yang paling banyak disoroti masyarakat. Di sisi lain, aspek empati kognitif dan emosional pada *chatbot* berpengaruh signifikan terhadap kepuasan pengguna, di mana sentimen negatif terhadap *chatbot* sering kali berkaitan dengan kurangnya respons emosional. Oleh karena itu, penggabungan analisis sentimen dan *topic modeling* dapat memberikan gambaran yang lebih komprehensif mengenai persepsi pengguna terhadap *chatbot* AI [6][7].

Meskipun penelitian terkait analisis sentimen dan *chatbot* telah berkembang, masih terdapat celah penelitian (*research gap*) yang perlu diisi. Sebagian besar penelitian sebelumnya menggunakan metode seperti SVM, *Naïve Bayes*, KNN, atau LSTM, sementara penggunaan *Logistic Regression* dalam konteks *chatbot* AI masih relatif terbatas. Selain itu, aspek metodologis seperti ketidakseimbangan kelas data sering kali kurang mendapat perhatian, padahal penanganan imbalanced dataset berpotensi meningkatkan keandalan performa model klasifikasi secara signifikan [8]. Di sisi lain, objek penelitian umumnya berfokus pada layanan publik, aplikasi, atau ulasan produk, bukan pada *chatbot* lokal Indonesia yang menggantikan *customer service*. Penelitian terkait juga jarang mengombinasikan analisis sentimen dengan *topic modeling* dalam satu kajian terpadu. Padahal, kombinasi kedua pendekatan tersebut berpotensi memberikan pemahaman yang lebih mendalam terhadap opini publik. Dengan demikian, penelitian mengenai Cekat.AI sebagai *chatbot* lokal yang diimplementasikan secara luas masih memiliki ruang eksplorasi yang signifikan. Celah inilah yang menjadi dasar dilakukannya penelitian ini [9].

Dalam penelitian ini, metode *Logistic Regression* dipilih karena memiliki kemampuan klasifikasi yang stabil dan interpretasi yang jelas terhadap faktor penentu sentimen. Representasi fitur teks dilakukan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) yang terbukti efektif dalam menonjolkan kata-kata penting dalam data teks. Pendekatan ini banyak digunakan dalam analisis sentimen karena mampu menangani karakteristik data media sosial yang singkat dan tidak terstruktur. Selain itu, penelitian ini juga menerapkan *topic modeling* untuk mengidentifikasi isu-isu utama dalam respons pengguna terhadap Cekat.AI. Kombinasi metode ini diharapkan dapat memberikan gambaran menyeluruh, baik dari sisi polaritas sentimen maupun konteks perbincangan. Data yang digunakan berasal dari komentar pengguna di platform X dan TikTok yang merepresentasikan opini publik secara spontan. Dengan pendekatan ini, analisis tidak hanya bersifat kuantitatif, tetapi juga kontekstual [10].

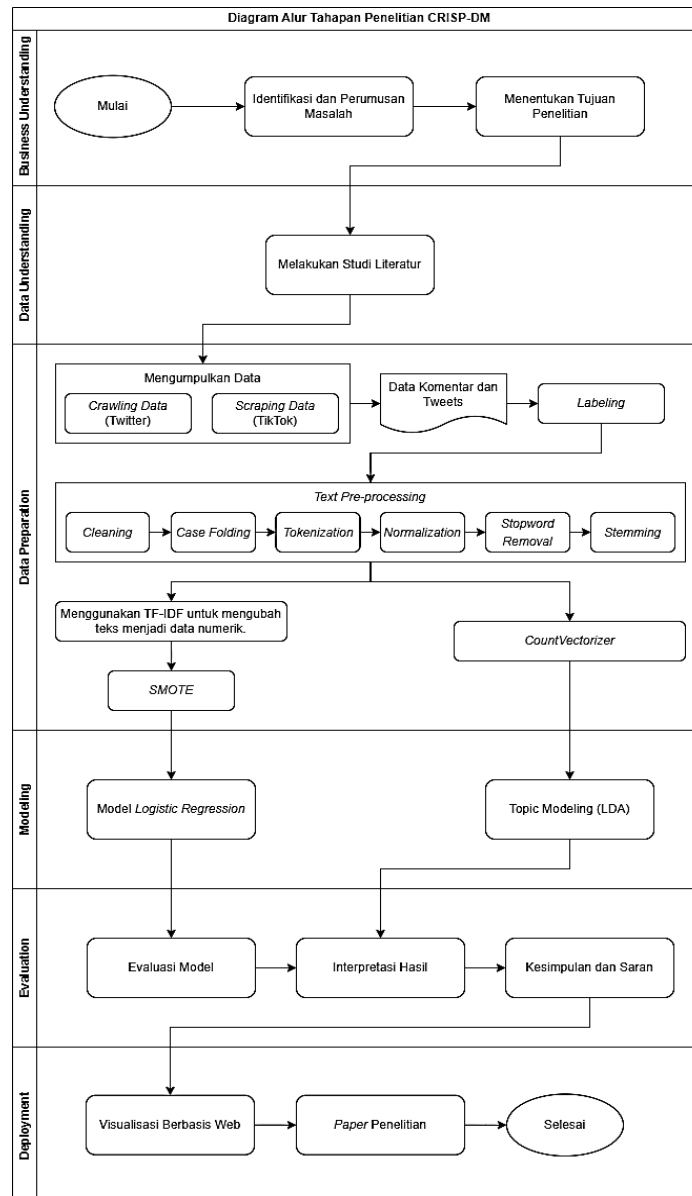
Berdasarkan uraian tersebut, tujuan penelitian ini adalah untuk menganalisis sentimen publik terhadap penggunaan Cekat.AI sebagai pengganti *customer service* melalui media sosial. Penelitian ini juga bertujuan

mengukur efektivitas metode *Logistic Regression* dan TF-IDF dalam mengklasifikasikan sentimen pengguna. Selain itu, penelitian ini berupaya mengidentifikasi topik dominan yang muncul dalam diskursus publik terkait implementasi Cekat.AI. Hasil penelitian diharapkan dapat memberikan kontribusi akademik dalam pengembangan studi analisis sentimen dan *chatbot* AI. Dari sisi praktis, temuan penelitian ini dapat menjadi bahan evaluasi bagi pengembang dan pelaku industri dalam meningkatkan kualitas layanan *chatbot*. Penelitian ini juga diharapkan memberikan masukan bagi pembuat kebijakan terkait dampak sosial adopsi AI dalam layanan pelanggan. Dengan demikian, penelitian ini berkontribusi dalam mendorong pemanfaatan AI yang lebih adaptif, berimbang, dan berorientasi pada kebutuhan pengguna [11].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) untuk menganalisis sentimen publik terhadap penggunaan Cekat AI sebagai pengganti layanan *customer service* melalui media sosial X dan TikTok. Kerangka CRISP-DM dipilih karena mampu memberikan alur penelitian yang sistematis, terstruktur, dan dapat direplikasi, mulai dari pemahaman masalah hingga evaluasi model [12][13]. Secara umum, tahapan penelitian yang dilakukan meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*. Alur lengkap tahapan penelitian ditampilkan pada Gambar 1, yang menggambarkan hubungan antarproses secara berurutan dan terintegrasi.



Gambar 1. Diagram Alur Tahapan Penelitian

Berdasarkan Gambar 1, berikut penjelasan setiap tahapan penelitian:

a. *Business Understanding*

Tahap ini bertujuan untuk memahami permasalahan dan tujuan penelitian. Permasalahan utama yang diangkat adalah bagaimana persepsi dan sentimen publik terhadap penggunaan Cekat AI dalam menggantikan peran *customer service* pada media sosial X dan TikTok. Tujuan penelitian adalah mengklasifikasikan sentimen pengguna ke dalam tiga kelas, yaitu positif, negatif, dan netral, menggunakan metode *Logistic Regression* berbasis TF-IDF.

Pemilihan tiga kelas sentimen dilakukan untuk menangkap kompleksitas opini publik. Kelas positif mencerminkan dukungan terhadap penggunaan Cekat AI, kelas negatif menunjukkan penolakan atau ketidakpuasan, sedangkan kelas netral merepresentasikan opini yang tidak menunjukkan kecenderungan emosional tertentu.

b. *Data Understanding*

Tahap *Data Understanding* dilakukan melalui studi literatur dan observasi awal dataset. Studi literatur bertujuan untuk memahami metode analisis sentimen, pemanfaatan AI dalam layanan pelanggan, serta penggunaan *Logistic Regression* dan TF-IDF dalam penelitian sebelumnya. Observasi awal terhadap dataset menunjukkan adanya ketidakseimbangan kelas, di mana kelas netral memiliki jumlah data yang jauh lebih sedikit dibandingkan kelas positif dan negatif. Temuan ini menjadi dasar pertimbangan penggunaan teknik penyeimbangan data pada tahap berikutnya.

c. *Data Preparation*

Tahap ini bertujuan untuk mempersiapkan data agar layak digunakan dalam proses pemodelan. Data dikumpulkan dari media sosial X melalui teknik *crawling* dan dari TikTok melalui teknik *scraping* komentar. Pengumpulan data dilakukan pada periode 1 September 2024 hingga 31 Juni 2025, dengan kata kunci dan *hashtag* yang relevan seperti CekatAI, Cekat, dan Customer Service AI sebagaimana ditunjukkan.

Data yang telah dikumpulkan kemudian diseleksi untuk memastikan relevansi konteks. Proses seleksi dan pelabelan sentimen dilakukan secara manual oleh tiga ahli bahasa untuk meningkatkan validitas data. Setelah pelabelan, dilakukan *text pre-processing* yang meliputi *cleaning*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*.

Selanjutnya, teks yang telah bersih dikonversi ke dalam bentuk numerik menggunakan metode TF-IDF, sehingga dapat diproses oleh model *machine learning*. Mengingat adanya ketidakseimbangan kelas, diterapkan metode SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan representasi kelas minoritas, khususnya kelas netral.

d. *Modeling*

Tahap *Modeling* dimulai dengan penerapan *topic modeling* menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik-topik dominan dalam percakapan pengguna terkait Cekat AI. Hasil LDA membantu memberikan konteks tematik terhadap analisis sentimen. Selanjutnya, dilakukan pembangunan model klasifikasi sentimen menggunakan *Logistic Regression*.

Dataset dibagi menjadi *training set* (70%) dan *testing set* (30%). Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur performa model. Representasi fitur diperoleh dari hasil pembobotan TF-IDF, yang kemudian digunakan sebagai *input* pada *Logistic Regression*.

e. *Evaluation*

Tahap *Evaluation* bertujuan untuk menilai kinerja model yang telah dibangun. Evaluasi dilakukan menggunakan *Confusion Matrix* dan metrik performa seperti *accuracy*, *precision*, *recall*, dan F1-score. Selain itu, evaluasi juga digunakan untuk menginterpretasikan hasil klasifikasi sentimen dalam konteks persepsi publik terhadap penggunaan Cekat AI sebagai *customer service*.

2.2 Metode Analisis Sentimen

Metode analisis sentimen dalam penelitian ini mengombinasikan TF-IDF sebagai teknik ekstraksi fitur dan *Logistic Regression* sebagai algoritma klasifikasi. TF-IDF digunakan untuk memberikan bobot pada kata berdasarkan tingkat kepentingannya dalam dokumen dan korpus, sehingga kata-kata yang lebih representatif memiliki pengaruh lebih besar dalam proses klasifikasi [14].

Secara matematis, TF-IDF dihitung berdasarkan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) sebagaimana ditunjukkan pada Persamaan (1) dan (2). Hasil perhitungan TF-IDF disusun dalam bentuk matriks dokumen-kata seperti pada yang akan kemudian digunakan sebagai *input* pada model *Logistic Regression*.

Logistic Regression digunakan untuk memodelkan hubungan antara fitur TF-IDF dan kelas sentimen. Model ini menghitung probabilitas setiap kelas sentimen menggunakan fungsi logistik (*sigmoid*). Proses pelatihan dilakukan secara iteratif dengan optimasi parameter menggunakan *gradient descent* hingga nilai *loss* konvergen.

Untuk meningkatkan kemampuan model dalam menangani data multikelas dan ketidakseimbangan kelas, dataset hasil SMOTE digunakan dalam proses pelatihan. Hasil prediksi model kemudian dibandingkan dengan label aktual pada data uji untuk menghitung performa klasifikasi. Evaluasi akhir dilakukan menggunakan *confusion matrix* sebagaimana disajikan yang menjadi dasar perhitungan metrik evaluasi model.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian serta pembahasan yang diperoleh dari penerapan metode analisis sentimen terhadap penggunaan Cekat.AI sebagai pengganti *customer service* pada media sosial X dan komentar TikTok. Pembahasan mencakup tahapan pengolahan data, penerapan metode klasifikasi menggunakan *Logistic Regression* berbasis TF-IDF, hasil pengujian model, serta analisis mendalam terhadap pola sentimen dan temuan tematik yang muncul dari data. Seluruh hasil yang disajikan pada bab ini bertujuan untuk menjawab rumusan masalah penelitian dan mendukung tujuan penelitian yang telah ditetapkan.

3.1 Deskripsi dan Karakteristik Data

Data yang digunakan dalam penelitian ini merupakan data teks berupa komentar dan ulasan pengguna media sosial yang membahas penggunaan Cekat.AI dalam konteks layanan *customer service*. Sumber data berasal dari dua platform media sosial utama, yaitu X (Twitter) dan TikTok, yang dipilih karena memiliki karakteristik pengguna serta pola interaksi yang berbeda [15][16].

- Sumber dan jumlah data: Data dikumpulkan menggunakan teknik *scraping* dengan kata kunci yang relevan dengan Cekat.AI dan layanan *customer service*. Total data yang berhasil dikumpulkan berjumlah ribuan entri komentar, yang kemudian diseleksi untuk menghilangkan data duplikat, spam, dan komentar yang tidak relevan dengan konteks penelitian.
- Distribusi platform: Sebagian data berasal dari TikTok yang didominasi oleh komentar singkat, emosional, dan reaktif terhadap konten video. Sementara itu, data dari platform X (Twitter) cenderung berbentuk opini singkat namun argumentatif, dengan fokus pada isu teknis, etika, dan implikasi sosial dari penggunaan AI.
- Distribusi kelas sentimen awal: Hasil pelabelan manual menunjukkan bahwa distribusi sentimen tidak seimbang. Kelas Negatif mendominasi dataset, diikuti oleh kelas Positif, dan kelas Netral sebagai kelas minoritas. Ketidakseimbangan ini menjadi tantangan utama dalam proses pemodelan dan menjadi dasar penerapan teknik penanganan *imbalanced data* pada tahap selanjutnya.

Tabel 1. Distribusi Data Berdasarkan Platform

Platform	Jumlah Data	Persentase
TikTok	1.842	58%
X (Twitter)	1.332	42%
Total	3.174	100%

3.2 Tahapan Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas data teks sebelum diproses oleh model klasifikasi. Proses ini sangat krusial karena data media sosial umumnya bersifat tidak terstruktur dan mengandung banyak noise.

- Case folding*: Seluruh teks diubah menjadi huruf kecil untuk menghindari perbedaan makna akibat kapitalisasi huruf.
- Cleaning*: Proses ini meliputi penghapusan URL, *username*, *hashtag*, angka, tanda baca berlebih, emoji tertentu, serta karakter non-alfabet yang tidak memiliki kontribusi terhadap analisis sentimen.
- Tokenization*: Teks dipecah menjadi unit kata (*token*) untuk memudahkan proses analisis dan ekstraksi fitur.
- Stopword removal*: Kata-kata umum yang tidak memiliki makna sentimen signifikan, seperti “dan”, “yang”, dan “adalah”, dihapus untuk mengurangi dimensi fitur.
- Stemming*: Setiap kata dikembalikan ke bentuk dasarnya untuk menyatukan variasi kata yang memiliki makna sama.

Tahapan pra-pemrosesan ini menghasilkan data teks yang lebih bersih, ringkas, dan siap digunakan pada tahap ekstraksi fitur menggunakan TF-IDF.

3.3 Implementasi dan Pengujian Model

Tahap implementasi dilakukan dengan menerapkan algoritma *Logistic Regression* sebagai metode klasifikasi sentimen, dengan TF-IDF sebagai teknik ekstraksi fitur. Pengujian dilakukan dalam beberapa skenario untuk mengevaluasi pengaruh ketidakseimbangan data terhadap performa model [17].

3.3.1 Ekstraksi Fitur TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) digunakan untuk merepresentasikan teks ke dalam bentuk numerik. Metode ini memberikan bobot lebih tinggi pada kata-kata yang sering muncul pada suatu dokumen tetapi jarang muncul di dokumen lain, sehingga dianggap lebih informatif dalam membedakan sentimen.

Hasil ekstraksi fitur menunjukkan bahwa kata-kata bernuansa emosional seperti “diganti”, “takut”, “kerja”, dan “admin” memiliki bobot tinggi pada kelas Negatif, sedangkan kata-kata seperti “membantu”, “efisien”, dan “cepat” dominan pada kelas Positif.

3.3.2 Pengujian Model Baseline

Pengujian awal dilakukan menggunakan *Logistic Regression* tanpa penanganan ketidakseimbangan data. Model ini menghasilkan akurasi global yang relatif tinggi, namun menunjukkan kelemahan serius dalam mendeteksi kelas Netral.

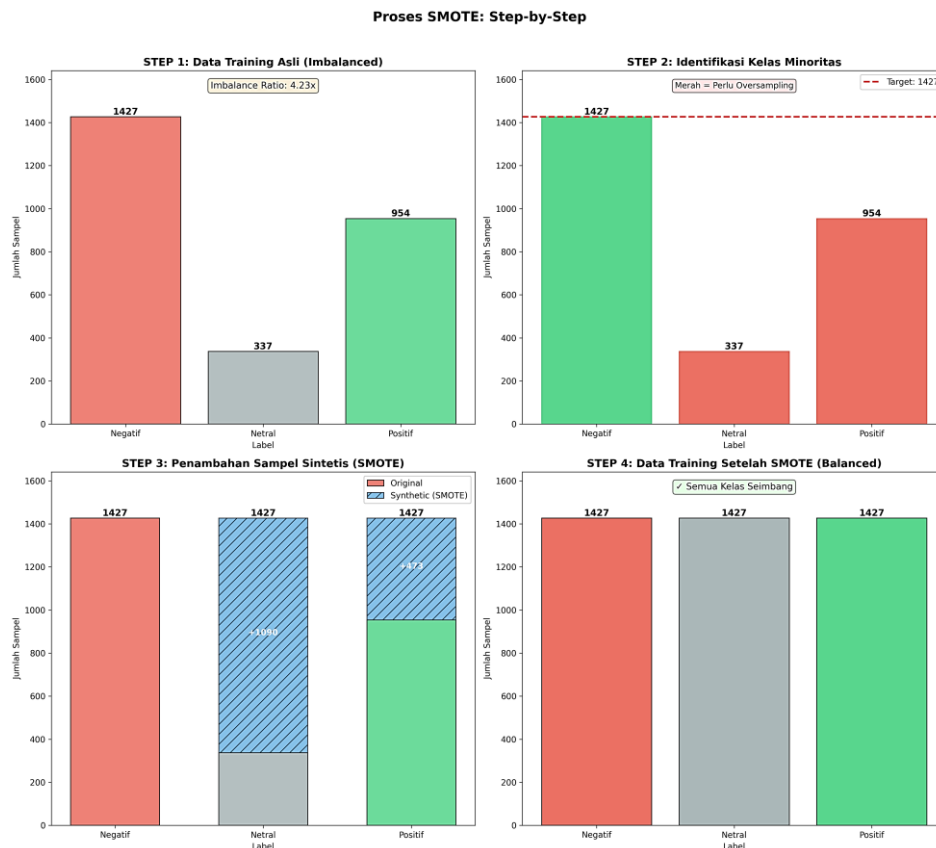
Tabel 2. Hasil Evaluasi Model *Baseline*

Kelas	Precision	Recall	F1-Score
Negatif	0.78	0.94	0.85
Positif	0.69	0.61	0.65
Netral	0.32	0.19	0.24

Hasil tersebut menunjukkan bahwa model cenderung memprediksi kelas mayoritas dan gagal menangkap nuansa sentimen yang ambigu.

3.3.3 Pengujian dengan SMOTE

Untuk mengatasi permasalahan ketidakseimbangan data (*imbalance*), penelitian ini menerapkan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*. Teknik ini bekerja dengan menghasilkan data sintesis pada kelas minoritas sehingga distribusi kelas menjadi lebih seimbang sebelum proses pelatihan model. Perbandingan visualisasi distribusi data sebelum dan sesudah penerapan SMOTE dapat dilihat pada Gambar 2.



Gambar 2. Visualisasi Proses SMOTE

Hasil pengujian menunjukkan peningkatan signifikan pada kemampuan model dalam mengenali kelas Netral, meskipun akurasi global sedikit menurun.

Tabel 3. Hasil Evaluasi Model dengan SMOTE

Kelas	Precision	Recall	F1-Score
Negatif	0.75	0.81	0.78
Positif	0.72	0.69	0.70
Netral	0.61	0.64	0.62

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa penerapan *Logistic Regression* berbasis TF-IDF mampu memberikan performa klasifikasi yang cukup baik pada data teks media sosial, khususnya setelah dilakukan penanganan

ketidakseimbangan data. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa metode klasik masih relevan untuk analisis sentimen pada data teks pendek, selama dilakukan pengolahan data yang tepat [18].

Dominasi sentimen negatif terhadap penggunaan Cekat.AI mengindikasikan bahwa resistensi publik tidak semata-mata disebabkan oleh kekurangan teknis sistem, melainkan oleh faktor psikologis dan sosial. Hasil ini diperkuat oleh temuan *topic modeling* pada bab selanjutnya yang menunjukkan adanya kecemasan terhadap otomatisasi dan hilangnya peran manusia.

Perbedaan karakteristik sentimen antara TikTok dan X menunjukkan bahwa platform media sosial memiliki peran penting dalam membentuk opini publik. TikTok yang bersifat visual dan emosional memperkuat reaksi spontan, sedangkan X lebih mendorong diskusi rasional. Hal ini menegaskan bahwa strategi komunikasi teknologi AI tidak dapat diseragamkan di seluruh platform. Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini memberikan kontribusi baru dengan mengombinasikan analisis sentimen, penanganan *imbalanced data*, serta pemodelan topik dalam satu kerangka analisis yang utuh. Selain itu, integrasi hasil model ke dalam *dashboard* analisis sentimen memperkuat aspek aplikatif penelitian ini [19].

3.4.1 Analisis Distribusi Sentimen Berdasarkan Platform

Untuk memperdalam pemahaman terhadap pola sentimen pengguna, dilakukan analisis distribusi sentimen yang dipisahkan berdasarkan platform media sosial. Pendekatan ini penting mengingat setiap platform memiliki karakteristik audiens, algoritma distribusi konten, serta gaya komunikasi yang berbeda.

Hasil analisis menunjukkan bahwa TikTok memiliki proporsi sentimen negatif yang lebih tinggi dibandingkan platform X. Hal ini tidak terlepas dari sifat TikTok yang berbasis konten video singkat dengan daya tarik emosional yang kuat. Pengguna cenderung menuliskan komentar secara spontan sebagai respons terhadap narasi visual, sehingga ekspresi ketidakpuasan dan kekhawatiran lebih mudah dimunculkan. Sebaliknya, X mendorong pengguna untuk menyampaikan opini secara ringkas namun argumentatif, yang menghasilkan diskursus yang relatif lebih seimbang antara kritik dan apresiasi.

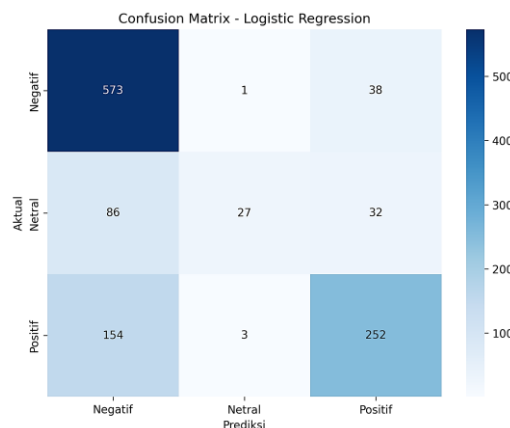
Tabel 4. Distribusi Sentimen Berdasarkan Platform

Platform	Positif	Netral	Negatif
TikTok	29,6%	10,8%	59,6%
X (Twitter)	41,2%	15,1%	43,7%

Tabel 4 memperkuat temuan bahwa TikTok berperan sebagai ruang utama munculnya resistensi sosial terhadap otomatisasi layanan pelanggan. Sementara itu, platform X menunjukkan potensi sebagai medium edukatif yang lebih efektif untuk menyampaikan narasi teknis dan rasional terkait pemanfaatan Cekat.AI.

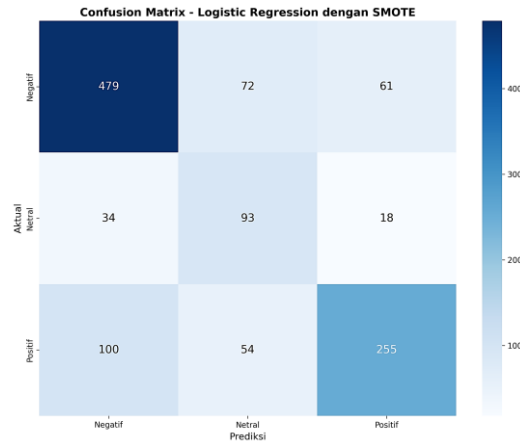
3.4.2 Analisis Confusion Matrix dan Implikasi Klasifikasi

Selain metrik akurasi, *precision*, dan *recall*, evaluasi model juga dilakukan melalui analisis *confusion matrix* untuk memahami pola kesalahan klasifikasi secara lebih rinci. *Confusion matrix* memberikan gambaran konkret mengenai kecenderungan model dalam memprediksi kelas sentimen yang sebenarnya. Visualisasi kinerja klasifikasi pada skenario *baseline* disajikan pada Gambar 3, sedangkan hasil klasifikasi setelah penerapan teknik SMOTE dapat dilihat pada Gambar 4.



Gambar 3. Confusion matrix model *baseline*

Berdasarkan Gambar 3, terlihat bahwa pada skenario *baseline*, kesalahan terbesar terjadi pada kelas Netral yang sering diklasifikasikan sebagai Negatif. Hal ini mengindikasikan bahwa model awal mengalami kesulitan membedakan ekspresi opini yang ambigu atau bersifat observasional.



Gambar 1. Confusion matrix setelah penerapan SMOTE

Setelah dilakukan penyeimbangan data sebagaimana ditunjukkan pada Gambar 4, kesalahan tersebut mengalami penurunan signifikan. Model terlihat lebih mampu memetakan kelas Netral dengan tepat, meskipun masih terdapat sedikit tumpang tindih antara kelas Netral dan Positif. Secara konseptual, fenomena ini mencerminkan kompleksitas bahasa media sosial, di mana satu kalimat dapat mengandung lebih dari satu muatan emosi. Oleh karena itu, hasil ini menegaskan bahwa analisis sentimen pada domain sosial tidak dapat sepenuhnya bergantung pada satu metrik evaluasi tunggal [20].

Sebagai bagian dari kontribusi aplikatif penelitian, model *Logistic Regression* terbaik yang telah dilatih menggunakan skenario SMOTE diintegrasikan ke dalam sebuah prototipe *dashboard* analisis sentimen berbasis web. *Dashboard* ini dirancang untuk menyajikan hasil analisis secara visual dan mudah dipahami oleh pengguna non-teknis. Fitur utama *dashboard* meliputi:

- Visualisasi distribusi sentimen dalam bentuk diagram batang dan diagram lingkaran.
- Tampilan tren sentimen berdasarkan waktu pengambilan data.
- Ringkasan topik dominan yang dihasilkan dari pemodelan topik LDA.
- Filter berdasarkan platform media sosial.

Keberadaan *dashboard* ini memperkuat nilai praktis penelitian, karena memungkinkan pengembang Cekat.AI dan pemangku kepentingan untuk memantau dinamika opini publik secara berkelanjutan tanpa harus melakukan analisis manual.

Jika dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma Naïve Bayes, KNN, atau *Support Vector Machine*, penelitian ini menunjukkan bahwa *Logistic Regression* masih kompetitif dalam konteks analisis sentimen media sosial, terutama ketika dikombinasikan dengan teknik penanganan data tidak seimbang [21]. Penelitian terdahulu umumnya berfokus pada evaluasi performa algoritma semata, sementara penelitian ini memperluas cakupan analisis dengan mengaitkan hasil klasifikasi sentimen dengan konteks sosial dan implikasi kebijakan. Selain itu, integrasi topic modeling memungkinkan identifikasi akar permasalahan di balik sentimen negatif, yang jarang dibahas secara mendalam pada studi sebelumnya. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi metodologis, tetapi juga kontribusi konseptual dalam memahami relasi antara adopsi teknologi AI dan persepsi publik di ruang digital.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa persepsi publik terhadap implementasi Cekat.AI sebagai pengganti *customer service* didominasi oleh sentimen negatif sebesar 52,5%, diikuti sentimen positif 35,1% dan netral 12,4%. Analisis komparatif lintas platform menunjukkan dinamika kontras, di mana TikTok didominasi respons emosional dan reaktif, sedangkan X menampilkan diskursus yang lebih argumentatif. Melalui pemodelan topik, terungkap bahwa akar utama resistensi publik adalah kecemasan sosial terhadap otomatisasi dan ancaman kehilangan pekerjaan (*automation anxiety*), sementara sentimen positif tumbuh dari apresiasi terhadap efisiensi dan inovasi teknologi lokal. Secara metodologis, penerapan teknik SMOTE pada algoritma *Logistic Regression* dengan fitur TF-IDF terbukti menjadi solusi paling optimal dalam mengatasi kendala bias mayoritas pada data tidak seimbang. Pendekatan ini secara signifikan meningkatkan *Recall* kelas Netral hingga 64,1% dan mencapai akurasi validasi silang tertinggi sebesar 79,98%. Temuan ini juga telah berhasil diimplementasikan ke dalam *dashboard* berbasis web sebagai alat bantu pengambilan keputusan strategis. Berdasarkan keterbatasan penelitian, saran utama untuk pengembangan masa depan adalah mengadopsi arsitektur *Deep Learning* seperti BERT untuk menangani kompleksitas sarkasme dan ambiguitas bahasa secara lebih presisi. Selain itu, perluasan dataset ke platform lain serta penerapan metode *Aspect-Based Sentiment Analysis* (ABSA) sangat direkomendasikan untuk memetakan dimensi sentimen yang lebih spesifik. Langkah ini krusial agar

pengembang dapat merancang strategi komunikasi yang lebih humanis guna memitigasi resistensi sosial dan meningkatkan keberterimaan teknologi di masyarakat.

REFERENCES

- [1] G. A. Pradnyana, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, "An explainable ensemble model for revealing the level of depression in social media by considering personality traits and sentiment polarity pattern," *Online Soc. Networks Media*, vol. 46, p. 100307, 2025, doi: <https://doi.org/10.1016/j.osnem.2025.100307>.
- [2] G. A. Pradnyana, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, "Revealing Depression Through Social Media via Adaptive Gated Cross-Modal Fusion Augmented With Insights From Personality Traits," *IEEE Access*, vol. 13, pp. 133465–133482, 2025, doi: 10.1109/ACCESS.2025.3593273.
- [3] I. G. B. A. Budaya and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1077–1086, 2024, doi: 10.57152/malcom.v4i3.1459.
- [4] I. P. D. W. Darmawan, G. A. Pradnyana, and I. B. N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 1, pp. 58–67, 2023, doi: 10.31598/sintechjournal.v6i1.1245.
- [5] U. Detthamrong *et al.*, "Topic Modeling Analytics of Digital Economy Research: Trends and Insights," *J. Scientometr. Res.*, vol. 13, pp. 448–458, Aug. 2024, doi: 10.5530/jscires.13.2.35.
- [6] I. K. R. Arthana, I. M. D. Maysanjaya, G. A. Pradnyana, and G. R. Dantes, "Optimizing Dropout Prediction in University Using Oversampling Techniques for Imbalanced Datasets," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 8, pp. 1052–1060, 2024, [Online]. Available: <http://www.ijiet.org/>
- [7] A. Mulia and A. Dzirkillah, "Analisis Sentimen dan Topik Perbincangan Netizen Indonesia Terkait Pengurangan Subsidi BBM," *J. Linguist. Komputasional*, vol. 7, pp. 1–10, Mar. 2024, doi: 10.26418/jlk.v7i1.142.
- [8] T. Jiang, C. Huang, Y. Xu, and H. Zheng, "Cognitive vs. emotional empathy: exploring their impact on user outcomes in health-assistant chatbots," *Behav. Inf. Technol.*, pp. 1–16, Mar. 2025, doi: 10.1080/0144929X.2025.2474087.
- [9] Y. D. Nugroho H, F. Zakiyabarsi, and A. J. Paramita, "Implementasi Smote-Enn Dan Borderline Smote Terhadap Performa Lightgbm Pada Imbalanced Class," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 1 SE-Articles, pp. 51–59, Jan. 2025, doi: 10.36341/rabit.v10i1.5436.
- [10] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 2 SE-Articles, pp. 89–98, Aug. 2023, doi: 10.31598/sintechjournal.v6i2.1372.
- [11] I. G. S. D. Putra and I. N. T. A. Putra, "Implementasi Metode Naïve Bayes Pada Analisis Sentimen Pengguna Aplikasi Mobile Kita Bisa," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025.
- [12] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier," *SINTECH (Science Inf. Technol. J.)*, vol. 5, no. 2 SE-Articles, pp. 227–233, Oct. 2022, doi: 10.31598/sintechjournal.v5i2.1241.
- [13] Y. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *J. MEDIA Inform. BUDIDARMA*, vol. 7, pp. 1551–1562, Jul. 2023, doi: 10.30865/mib.v7i3.6461.
- [14] T. Abdillah, H. Hasmawati, and B. Bunyamin, "Comparison of TF-IDF and GloVe Word Embedding for Sentiment Analysis of 2024 Presidential Candidates," *Build. Informatics, Technol. Sci.*, vol. 6, pp. 961–969, Sep. 2024, doi: 10.47065/bits.v6i2.5668.
- [15] N. Saraswati and I. G. A. A. D. Indradewi, "Recognize The Polarity of Hotel Reviews using Support Vector Machine," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, pp. 25–36, Nov. 2022, doi: 10.30812/matrik.v22i1.1848.
- [16] R. Fawwaz, K. Satrio Laksono, N. Assyaputri, and N. Rakhmawati, "Analisis Tren Penggunaan Bahasa Di Tiktok: Studi Istilah Baru Era Digital Seperti Skibidi, Sigma, Dan Rizz Melalui Kuesioner Dan Data Scraping Komentar Video," *J. Etika Teknol. Inf.*, vol. 1, no. 1, pp. 1–8, Dec. 2024.
- [17] A. Sasmita, G. A. Pradnyana, and D. Divayana, "Pengembangan Sistem Analisis Sentimen Untuk Evaluasi Kinerja Dosen Universitas Pendidikan Ganesha Dengan Metode Naïve Bayes," *JST (Jurnal Sains dan Teknol.)*, vol. 11, pp. 451–462, Sep. 2022, doi: 10.23887/jstundiksha.v11i2.44384.
- [18] R. B. Yusuf Kanzun; Komaria, Nurul; Aqma, Elyana Azatul; Cahyani, Laili, "Analisis Sentimen Terhadap Aplikasi Google Meet Berdasarkan Komentar Pengguna Menggunakan Metode Logistic Regresion," *EDUTIC*, no. Vol 11, No 1: November 2024, pp. 53–64, 2024, [Online]. Available: <https://journal.trunojoyo.ac.id/edutic/article/view/28113/10271>
- [19] N. Sedana, I. Wijaya, and I. Arthana, "Analisis Sentimen Berbahasa Inggris Dengan Metode Lstm Studi

Kasus Berita Online Pariwisata Bali Analisis Sentimen Berbahasa Inggris Dengan Metode Lstm Studi Kasus Berita Online Pariwisata Bali,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, pp. 1325–1334, Dec. 2024, doi: 10.25126/jtiik.1168792.

- [20] A. M. Sundjaja, P. Utomo, D. Matthew, G. R. Heliyanto, and N. S. Putra, “The determinant factors of continuance intention to revisit omnichannel retailer companies: mean-end chain theory approach,” *Cogent Bus. Manag.*, vol. 11, no. 1, p. 2332504, Dec. 2024, doi: 10.1080/23311975.2024.2332504.
- [21] I. Ainunnisa and S. Sulastri, “Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan Naïve Bayes Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan Naïve Bayes,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 6, pp. 423–430, Jul. 2023, doi: 10.32493/jtsi.v6i3.31076.