



Komparasi Analisis Sentimen Pengguna Aplikasi ChatGPT dan Gemini di Google Play Store menggunakan Metode IndoBERT

Okan Rahadin *, Erna Dwi Astuti, Nur Hasanah

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Sains Al-Qur'an Wonosobo, Indonesia
 *Email (Penulis Korespondensi): okanrahadian69@gmail.com

Abstrak. Perkembangan aplikasi kecerdasan buatan generatif seperti ChatGPT dan Gemini telah menarik perhatian jutaan pengguna di Indonesia, sehingga menghasilkan ulasan dalam jumlah besar di Google Play Store yang sulit dianalisis secara manual. Penelitian ini bertujuan membandingkan sentimen pengguna terhadap ChatGPT dan Gemini menggunakan metode IndoBERT. Data dikumpulkan melalui scraping menggunakan Google Play Scraper sebanyak 5.000 ulasan untuk masing-masing aplikasi, kemudian melalui tahap preprocessing (case folding, cleaning, tokenizing, stopword removal, dan stemming), pelabelan otomatis berdasarkan rating menjadi kelas positif dan negatif, serta fine-tuning model indobenchmark/indobert-base-p1 dengan pembagian data latih dan uji 80:20. Hasil penelitian menunjukkan model ChatGPT memperoleh akurasi 91,32% dan model Gemini memperoleh akurasi 86,98%. Distribusi sentimen menunjukkan ulasan positif ChatGPT mencapai 81,51%, jauh lebih tinggi dibandingkan Gemini sebesar 65,55%, sedangkan proporsi sentimen negatif Gemini (34,45%) hampir dua kali lipat dibandingkan ChatGPT (18,51%). Dapat disimpulkan bahwa persepsi pengguna terhadap ChatGPT secara umum lebih positif dibandingkan Gemini, dan IndoBERT terbukti efektif digunakan untuk analisis sentimen ulasan berbahasa Indonesia yang bersifat informal. Penelitian ini turut mendukung pencapaian Tujuan Pembangunan Berkelanjutan (SDGs), khususnya SDG 4, SDG 9, dan SDG 16, melalui peningkatan literasi teknologi, kapasitas riset, serta transparansi dan perlindungan data pengguna.

Kata kunci: Analisis Sentimen; ChatGPT; Gemini; IndoBERT; Google Play Store

Abstract. The development of generative artificial intelligence applications such as ChatGPT and Gemini has attracted millions of users in Indonesia, generating a large volume of reviews on the Google Play Store that are difficult to analyze manually. This study aims to compare user sentiment toward ChatGPT and Gemini using the IndoBERT method. Data were collected through scraping using Google Play Scraper, obtaining 5,000 reviews for each application, followed by preprocessing (case folding, cleaning, tokenizing, stopword removal, and stemming), automatic labeling based on rating into positive and negative classes, and fine-tuning of the indobenchmark/indobert-base-p1 model with an 80:20 train-test split. The results show that the ChatGPT model achieved an accuracy of 91.32%, while the Gemini model achieved 86.98%. The sentiment distribution shows that positive reviews for ChatGPT reached 81.51%, considerably higher than Gemini at 65.55%, while the proportion of negative sentiment for Gemini (34.45%) was almost twice that of ChatGPT (18.51%). It can be concluded that user perception of ChatGPT is generally more positive than that of Gemini, and IndoBERT proved effective for sentiment analysis of informal Indonesian-language reviews. This study also supports the achievement of the Sustainable Development Goals (SDGs), particularly SDG 4, SDG 9, and SDG 16, through improved technology literacy, research capacity, and transparency and protection of user data.

1. Pendahuluan

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) dalam beberapa tahun terakhir telah membawa perubahan signifikan dalam berbagai aspek kehidupan, yang sebagian besar didorong oleh kemajuan pesat pada bidang *deep learning* (Goodfellow dkk., 2016). Salah satu wujud nyata perkembangan tersebut adalah kemunculan aplikasi berbasis AI yang berfungsi sebagai asisten virtual dan mampu berinteraksi secara komunikatif dengan pengguna. Di antara berbagai aplikasi AI generatif yang populer, ChatGPT yang dikembangkan oleh OpenAI dan Gemini yang dikembangkan oleh Google merupakan dua platform yang paling banyak digunakan oleh masyarakat di berbagai negara, termasuk Indonesia (Dwivedi dkk., 2023). Persaingan yang semakin ketat di pasar AI generatif menjadikan pemahaman terhadap persepsi pengguna sebagai hal yang sangat penting, sebab pengembang membutuhkan umpan balik berbasis data untuk terus meningkatkan kualitas layanan agar tetap kompetitif. Selain itu, maraknya penggunaan aplikasi AI generatif di Indonesia menghasilkan volume ulasan pengguna yang terus bertambah setiap harinya, sehingga diperlukan evaluasi berbasis data yang sistematis untuk mengukur bagaimana masyarakat sesungguhnya menerima dan menilai teknologi tersebut, alih-alih hanya mengandalkan asumsi maupun opini yang tidak terukur.

ChatGPT merupakan model bahasa generatif berbasis arsitektur *Generative Pre-trained Transformer* (GPT) yang mampu memahami konteks bahasa dan menghasilkan teks secara alami melalui proses *pre-training* menggunakan data dalam jumlah besar serta *fine-tuning* untuk berbagai tugas pemrosesan bahasa alami (Liu dkk., 2023). Sejak diluncurkan pada November 2022, ChatGPT berhasil menarik perhatian jutaan pengguna di seluruh dunia karena kemampuannya dalam menjawab pertanyaan, menulis konten, membantu pengkodean, hingga mendukung berbagai kebutuhan akademis dan profesional (Hassani & Silva, 2023). Sementara itu, Gemini merupakan model *Large Language Model* (LLM) multimodal yang dikembangkan oleh *Google DeepMind* dengan kemampuan memproses berbagai jenis data seperti teks, gambar, audio, video, dan kode secara terpadu, sehingga mampu memberikan pemahaman dan penalaran yang lebih menyeluruh (Gemini Team, 2023). Gemini dirancang sebagai generasi baru kecerdasan buatan yang memiliki integrasi kuat dengan berbagai layanan digital Google untuk mendukung produktivitas dan interaksi pengguna (Imran & Almusharraf, 2024).

Popularitas kedua aplikasi tersebut dapat dilihat dari jumlah unduhan pada platform *Google Play Store*. Sebagai platform distribusi aplikasi Android terbesar di dunia, *Google Play Store* menjadi salah satu sumber data yang kaya akan ulasan pengguna yang mencerminkan persepsi, kepuasan, serta keluhan masyarakat terhadap suatu aplikasi. Ulasan-ulasan tersebut mengandung informasi berharga yang dapat dimanfaatkan oleh pengembang maupun peneliti untuk memahami pengalaman pengguna secara lebih mendalam (Herlinawati dkk., 2020).

Analisis sentimen merupakan salah satu pendekatan dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau perasaan yang terkandung dalam teks ke dalam kategori tertentu, seperti positif, negatif, dan netral (Medhat dkk., 2014). Analisis sentimen telah menjadi salah satu teknik yang

banyak digunakan untuk memahami opini publik dari data teks tidak terstruktur dan mendukung pengambilan keputusan berbasis data (Kotelnikova dkk., 2022). Meskipun analisis sentimen terhadap ulasan *Google Play Store* telah banyak diterapkan pada berbagai domain aplikasi (Liu, 2015), penelitian yang secara khusus membandingkan sentimen pengguna terhadap dua aplikasi AI generatif sekaligus, yaitu ChatGPT dan Gemini, dalam konteks bahasa Indonesia masih sangat terbatas.

Beberapa penelitian terdahulu telah mengeksplorasi analisis sentimen pada ulasan ChatGPT maupun Gemini secara terpisah. Mursidah dkk. (2025) melakukan klasifikasi sentimen terhadap ulasan pengguna ChatGPT berbahasa Indonesia menggunakan IndoBERT dan mencapai akurasi 89%. Fahlapu & Rosandi (2025) menganalisis sentimen masyarakat terhadap Gemini AI menggunakan beberapa metode machine learning dan menemukan bahwa model Support Vector Machine (SVM) memberikan performa tertinggi dengan akurasi 96,34%. Subhan dkk. (2025) meneliti sentimen publik terhadap ChatGPT, Gemini, dan DeepSeek menggunakan metode Naive Bayes dan *Logistic Regression*, yang mengungkapkan bahwa ChatGPT menerima sentimen positif tertinggi dibandingkan dua aplikasi lainnya. Namun, penelitian-penelitian tersebut memiliki keterbatasan yang menjadi celah penelitian (*research gap*): belum ada penelitian yang secara khusus membandingkan sentimen pengguna berbahasa Indonesia terhadap ChatGPT dan Gemini secara bersamaan menggunakan metode yang sama; sebagian besar penelitian sebelumnya menggunakan metode klasik seperti Naive Bayes, SVM, atau *Logistic Regression* yang memiliki keterbatasan dalam memahami konteks kalimat secara mendalam dibandingkan model berbasis *transformer* seperti BERT; serta penelitian yang menggunakan IndoBERT untuk analisis komparatif sentimen antar platform AI masih sangat terbatas.

IndoBERT merupakan model *Bidirectional Encoder Representations from Transformers* (BERT) yang dikembangkan dan dilatih menggunakan data teks berbahasa Indonesia dalam skala besar, sehingga terbukti unggul dalam memahami konteks bahasa Indonesia yang kaya akan variasi linguistik dibandingkan model BERT umum (Wirayudha dkk., 2025). Beberapa riset terdahulu membuktikan efektivitas IndoBERT dalam tugas-tugas NLP berbahasa Indonesia, termasuk analisis sentimen pada ulasan aplikasi di *Google Play Store*, misalnya pada aplikasi *e-commerce* (Putri dkk., 2025), platform transportasi dan layanan publik (Rosdiana dkk., 2025), aplikasi Mobile JKN (Tarwoto dkk., 2025), serta aplikasi perbankan digital BRI Mo (Simarmata & Sasongko, 2025). Berdasarkan rekam jejak keberhasilan tersebut, IndoBERT dipilih sebagai metode utama dalam penelitian ini karena kemampuannya yang teruji dalam menangani karakteristik ulasan informal, singkatan, dan ekspresi emosional pengguna.

Berdasarkan uraian di atas, penelitian ini merupakan studi komparatif pertama yang menggunakan metode IndoBERT untuk menganalisis sentimen pengguna ChatGPT dan Gemini secara bersamaan pada ulasan berbahasa Indonesia, sehingga memberikan kontribusi dalam memahami preferensi pengguna Indonesia terhadap dua platform AI generatif terbesar tersebut. Penelitian ini bertujuan untuk mengisi *research gap* yang ada dengan melakukan komparasi analisis sentimen pengguna aplikasi ChatGPT dan Gemini di *Google Play Store* menggunakan metode IndoBERT. Data penelitian diperoleh melalui proses *scraping* ulasan pengguna kedua aplikasi pada *Google Play Store*. Dengan melakukan komparasi menggunakan metode yang sama, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan analisis sentimen berbahasa Indonesia, khususnya pada

kategori aplikasi AI generatif, serta menjadi referensi bagi pengembang aplikasi dalam memahami persepsi pengguna.

Selain kontribusi keilmuan, penelitian ini juga diharapkan memberikan kontribusi terhadap pencapaian Tujuan Pembangunan Berkelanjutan (Sustainable Development Goals/SDGs). Penelitian ini mendukung SDG 4 (Pendidikan Berkualitas) melalui peningkatan literasi teknologi dan kemudahan akses informasi masyarakat terhadap pemanfaatan aplikasi kecerdasan buatan, SDG 9 (Industri, Inovasi, dan Infrastruktur) khususnya target 9.5 yang mendorong peningkatan kapasitas riset ilmiah dan teknologi di negara berkembang termasuk Indonesia, serta SDG 16 (Perdamaian, Keadilan, dan Kelembagaan yang Tangguh) melalui dukungan terhadap transparansi akses informasi dan perlindungan data pengguna dalam pemanfaatan teknologi digital.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis fine-tuning model bahasa pra-latih. Objek penelitian adalah ulasan pengguna terhadap aplikasi ChatGPT dan Gemini pada *Google Play Store*, karena keduanya merupakan aplikasi AI chatbot dengan jumlah pengguna dan unduhan terbesar sehingga ulasannya representatif untuk menggambarkan persepsi publik. Tahapan penelitian dilakukan secara bertahap, mulai dari pengumpulan data, preprocessing, pelabelan, *fine-tuning* model, evaluasi, hingga analisis komparatif.

2.1. Pengumpulan Data

Data dikumpulkan melalui teknik scraping menggunakan pustaka *google-play-scraper* berbasis Python yang dijalankan pada lingkungan Google Colaboratory. Fungsi `reviews()` digunakan untuk mengambil ulasan berbahasa Indonesia (`lang="id"`, `country="id"`) dengan urutan Most Relevant sebanyak 5.000 ulasan untuk masing-masing aplikasi, yaitu ChatGPT (ID paket `com.openai.chatgpt`) dan Gemini (ID paket `com.google.android.apps.bard`), sehingga diperoleh total 10.000 ulasan. Atribut data yang digunakan meliputi nama pengguna (`userName`), rating (`score`), isi ulasan (`content`), dan tanggal ulasan (`at`).

2.2. Preprocessing Data

Tahap *preprocessing* dilakukan melalui lima langkah, yaitu *case folding* (pengubahan seluruh teks menjadi huruf kecil), *cleaning* (penghapusan tanda baca, angka, emoji, dan karakter tidak relevan menggunakan *Regular Expression*), *tokenizing* (pemecahan kalimat menjadi token menggunakan *word_tokenize* dari pustaka NLTK), *stopword removal* (penghapusan kata umum menggunakan daftar *stopwords* Bahasa Indonesia dari NLTK), dan *stemming* (pengubahan kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi). Seluruh proses ini menghasilkan dataset teks bersih yang siap digunakan pada tahap pelabelan dan pemodelan.

2.3. Pelabelan Data

Pelabelan sentimen dilakukan secara otomatis berdasarkan nilai rating yang diberikan pengguna. Ulasan dengan rating 1 dan 2 diberi label Negatif, ulasan dengan rating 4 dan 5 diberi label Positif, sedangkan ulasan dengan rating 3 dikategorikan netral dan tidak

digunakan pada tahap pelatihan model. Setelah proses ini, jumlah data yang digunakan untuk *fine-tuning* adalah 4.664 ulasan untuk ChatGPT dan 4.567 ulasan untuk Gemini.

2.4. Fine-Tuning Model IndoBERT

Data yang telah diberi label dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi *train_test_split* dari *scikit-learn* secara *stratified* (*stratify* berdasarkan label) dengan *random_state*=42, dilakukan secara terpisah untuk masing-masing aplikasi. Proses *fine-tuning* dilakukan terhadap model pra-latih *indobenchmark/indobert-base-p1* menggunakan pustaka *Hugging Face Transformers* dan *framework PyTorch* pada lingkungan *Google Colaboratory* dengan akselerasi GPU Tesla T4. Teks diproses melalui tokenizer IndoBERT dengan *padding*="max_length", *truncation*=True, dan *max_length*=128. Konfigurasi pelatihan (*TrainingArguments*) meliputi *learning_rate* 2e-5, *batch_size* 16, jumlah epoch sebanyak 3, dan *weight_decay* 0,01, menggunakan *Trainer* API dengan strategi evaluasi dan penyimpanan model pada setiap akhir epoch.

2.5. Evaluasi Model dan Analisis Komparatif

Evaluasi performa model dilakukan menggunakan data uji (20% dari total data) dengan mengukur nilai *accuracy*, *precision*, *recall*, dan *F1-score* berdasarkan *confusion matrix*. Selanjutnya dilakukan analisis komparatif antara hasil klasifikasi sentimen ChatGPT dan Gemini, yang meliputi perbandingan distribusi sentimen dan perbandingan performa model, untuk memperoleh gambaran perbedaan persepsi pengguna terhadap kedua aplikasi. Untuk menguji apakah perbedaan akurasi antara model ChatGPT dan model Gemini bersifat signifikan secara statistik, dilakukan uji beda dua proporsi (*two-proportion z-test*) terhadap jumlah prediksi benar pada masing-masing data uji. Uji McNemar tidak digunakan karena mensyaratkan kedua model dievaluasi pada set data uji yang sama (berpasangan), sedangkan pada penelitian ini model ChatGPT dan model Gemini dievaluasi pada set data uji yang berbeda dan saling independen, sehingga uji beda dua proporsi lebih sesuai untuk membandingkan performa keduanya.

3. Hasil dan Pembahasan

Hasil penelitian disajikan berdasarkan tahapan yang telah dilakukan pada bagian metode, mulai dari hasil pengumpulan data, hasil pelabelan sentimen, hasil evaluasi model, hingga analisis komparatif antara ChatGPT dan Gemini.

3.1. Hasil Pengumpulan Data

Proses scraping menghasilkan 5.000 ulasan untuk aplikasi ChatGPT dan 5.000 ulasan untuk aplikasi Gemini, sehingga total data awal berjumlah 10.000 ulasan. Pemeriksaan data menunjukkan tidak ditemukan nilai kosong maupun data duplikat, sehingga seluruh data dapat dilanjutkan ke tahap preprocessing.

Tabel 1. Jumlah Data Hasil Scraping

Aplikasi	Jumlah Data
ChatGPT	5.000
Gemini	5.000
Total	10.000

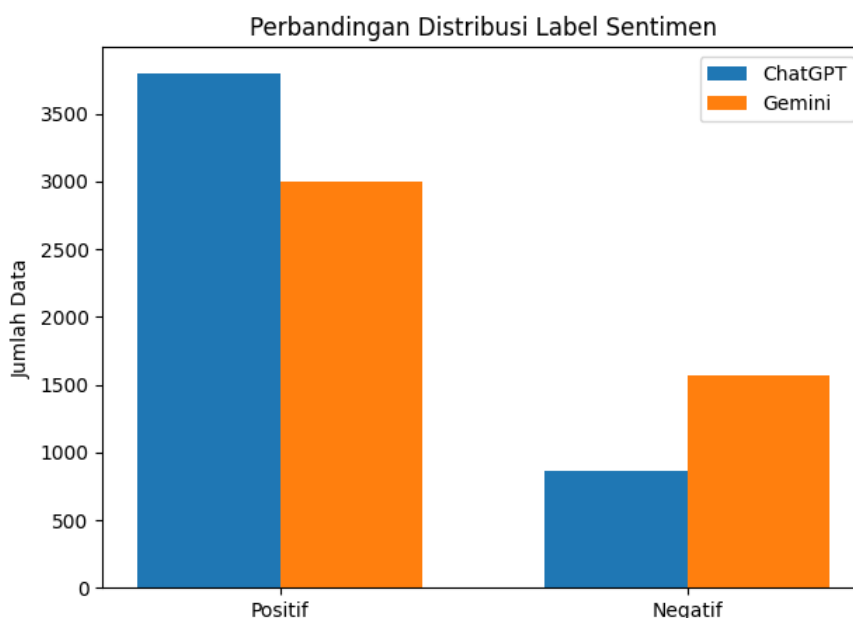
Distribusi rating awal menunjukkan bahwa ulasan ChatGPT didominasi rating 5 bintang (3.199 ulasan) dengan rating 1 bintang hanya 667 ulasan, sedangkan ulasan Gemini meskipun juga didominasi rating 5 bintang (2.418 ulasan), memiliki jumlah rating 1 bintang yang jauh lebih besar yaitu 1.217 ulasan. Hal ini menjadi indikasi awal bahwa proporsi pengguna yang memberikan penilaian rendah pada Gemini relatif lebih tinggi dibandingkan ChatGPT.

3.2. Hasil Pelabelan Sentimen

Setelah proses pelabelan berdasarkan rating dan penghapusan kelas netral, diperoleh distribusi label sentimen.

Tabel 2. Distribusi Label Sentimen ChatGPT dan Gemini

Label	ChatGPT	Persentase	Gemini	Persentase
Positif	3.801	81,51%	2.994	65,55%
Negatif	863	18,51%	1.573	34,45%
Total	4.664	100%	4.567	100%



Gambar 1. Distribusi Label Sentimen ChatGPT dan Gemini

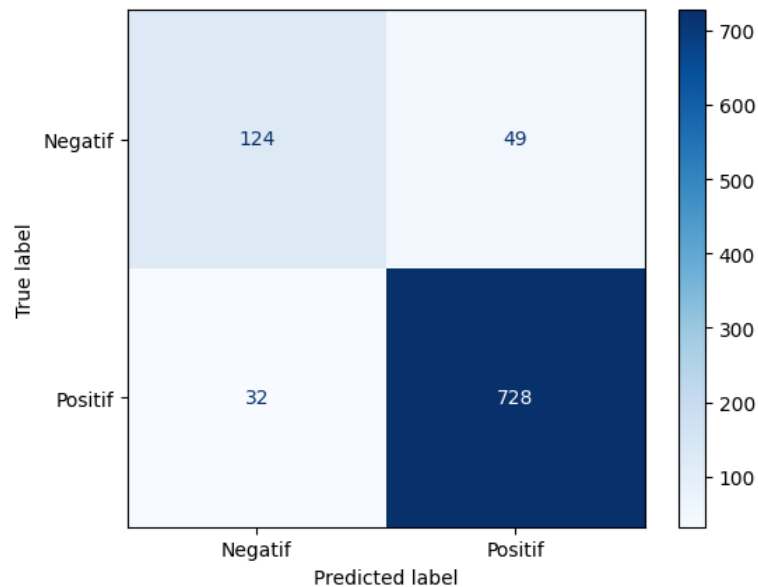
Kedua aplikasi memiliki karakteristik distribusi kelas yang tidak seimbang (*imbalanced*), dengan kelas Positif mendominasi pada kedua aplikasi. Namun demikian, proporsi kelas Negatif pada Gemini (34,45%) jauh lebih besar dibandingkan ChatGPT (18,51%), yang sejalan dengan temuan distribusi rating dan mengindikasikan bahwa secara umum pengguna Gemini pada sampel data ini melaporkan pengalaman yang relatif kurang memuaskan dibandingkan pengguna ChatGPT.

3.3. Hasil Evaluasi Model

Model IndoBERT untuk ChatGPT diuji terhadap 933 data uji, sedangkan model Gemini diuji terhadap 914 data uji. Hasil confusion matrix dan classification report masing-masing model disajikan pada Tabel 3, Tabel 4, Gambar 2, dan Gambar 3.

Tabel 3. *Classification Report* Model ChatGPT

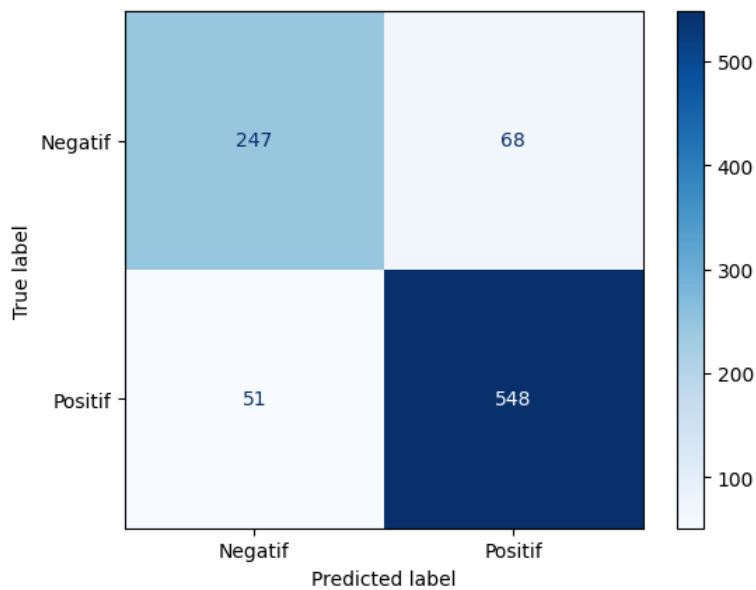
Kelas	Precision	Recall	F1-score	Support
Negatif	0,7949	0,7168	0,7538	173
Positif	0,9369	0,9579	0,9473	760
Accuracy			0,9132	933
Macro avg	0,8659	0,8373	0,8505	933
Weighted avg	0,9106	0,9132	0,9114	933



Gambar 2. *Confusion Matrix* Model ChatGPT

Tabel 4. *Classification Report* Model Gemini

Kelas	Precision	Recall	F1-score	Support
Negatif	0,8289	0,7841	0,8059	315
Positif	0,8896	0,9149	0,9021	599
Accuracy			0,8698	914
Macro avg	0,8592	0,8495	0,8540	914
Weighted avg	0,8687	0,8698	0,8689	914



Gambar 3. *Confusion Matrix* Model Gemini

Model ChatGPT memperoleh akurasi sebesar 91,32%. Model menunjukkan performa yang sangat baik dalam mengenali kelas Positif dengan precision 93,69% dan recall 95,79%, namun performa pada kelas Negatif relatif lebih rendah dengan precision 79,49% dan recall 71,68%. Rendahnya recall kelas Negatif menunjukkan bahwa masih terdapat cukup banyak ulasan negatif yang salah diklasifikasikan sebagai positif, sejalan dengan jumlah data latih kelas Negatif yang jauh lebih sedikit dibandingkan kelas Positif.

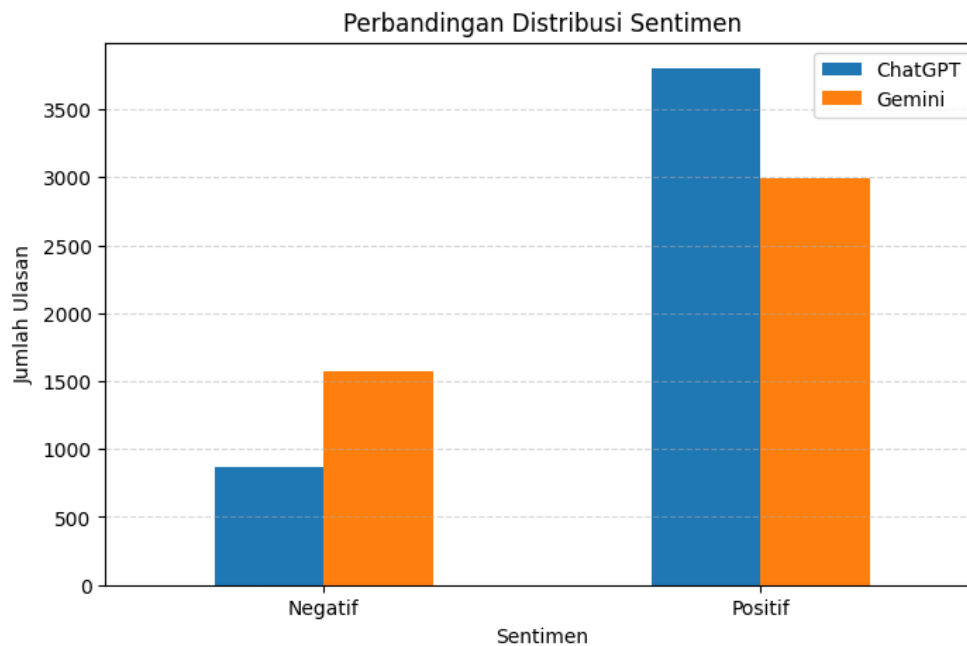
Model Gemini memperoleh akurasi sebesar 86,98%, dengan precision dan recall pada kelas Positif masing-masing 88,96% dan 91,49%, sedangkan pada kelas Negatif sebesar 82,89% dan 78,41%. Meskipun akurasi model Gemini lebih rendah dibandingkan ChatGPT, model ini menunjukkan kemampuan mengenali kelas Negatif yang relatif lebih baik, sejalan dengan proporsi data latih kelas Negatif pada Gemini yang lebih besar.

3.4. Analisis Komparatif Hasil Sentimen ChatGPT dan Gemini

3.4.1. Perbandingan Distribusi Sentimen

Tabel 5. Perbandingan Distribusi Sentimen ChatGPT dan Gemini

Sentimen	ChatGPT	Gemini
Negatif	863	1.573
Positif	3.801	2.994



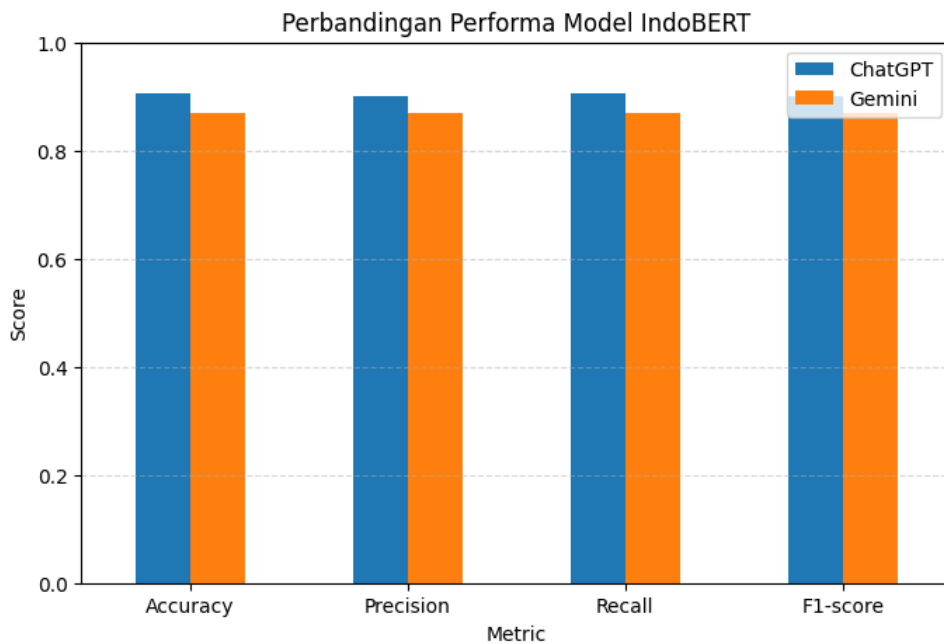
Gambar 4. Perbandingan Distribusi Sentimen ChatGPT dan Gemini

Jumlah ulasan bersentimen Positif pada ChatGPT (3.801) lebih tinggi dibandingkan Gemini (2.994), sedangkan jumlah ulasan bersentimen Negatif pada Gemini (1.573) hampir dua kali lipat dibandingkan ChatGPT (863). Temuan ini mengindikasikan bahwa secara umum persepsi pengguna terhadap ChatGPT pada periode data yang dianalisis cenderung lebih positif dibandingkan persepsi pengguna terhadap Gemini.

3.4.2. Perbandingan Performa Model

Tabel 6. Perbandingan Performa Model IndoBERT ChatGPT dan Gemini

Metrik	ChatGPT	Gemini
Accuracy	91,32%	86,98%
Precision	91,06%	86,87%
Recall	91,32%	86,98%
F1-score	91,14%	86,89%



Gambar 5. Perbandingan Performa Model IndoBERT ChatGPT dan Gemini

Model IndoBERT yang di *fine-tuning* pada data ulasan ChatGPT menghasilkan performa yang lebih unggul pada seluruh metrik evaluasi dibandingkan model pada data ulasan Gemini, dengan selisih accuracy sebesar 3,48 poin persentase. Selisih performa ini sejalan dengan karakteristik distribusi kelas pada masing-masing dataset, di mana proporsi kelas Positif pada ChatGPT jauh lebih dominan sehingga pola sentimen menjadi lebih mudah dipelajari oleh model, sedangkan data Gemini memiliki proporsi kelas Negatif yang lebih besar dan lebih beragam sehingga tingkat kesalahan klasifikasi menjadi relatif lebih tinggi. Distribusi kelas yang berbeda tersebut menyebabkan model pada dataset Gemini harus mempelajari variasi ulasan negatif yang lebih banyak, sehingga proses klasifikasi menjadi lebih kompleks dibandingkan model pada dataset ChatGPT. Hasil uji beda dua proporsi (two-proportion z-test) terhadap jumlah prediksi benar pada kedua data uji (852 dari 933 data pada ChatGPT dan 795 dari 914 data pada Gemini) menunjukkan nilai z sebesar 3,00 dengan p-value sebesar 0,003 ($p < 0,05$). Nilai ini menunjukkan bahwa perbedaan akurasi antara model ChatGPT dan model Gemini bersifat signifikan secara statistik, sehingga dapat disimpulkan bahwa keunggulan performa model ChatGPT bukan sekadar kebetulan, melainkan sejalan dengan perbedaan karakteristik distribusi kelas pada data latih kedua aplikasi.

Secara keseluruhan, hasil analisis komparatif menunjukkan bahwa meskipun kedua model mampu mengklasifikasikan sentimen ulasan dengan performa yang tergolong baik (accuracy di atas 85%), terdapat perbedaan persepsi pengguna maupun performa model antara ChatGPT dan Gemini. Aplikasi ChatGPT memperoleh proporsi sentimen positif yang lebih tinggi serta performa model klasifikasi yang lebih baik, sedangkan aplikasi Gemini memiliki proporsi sentimen negatif yang lebih besar sehingga performa model klasifikasinya relatif lebih rendah. Temuan ini konsisten dengan hasil penelitian Mursidah dkk. (2025) yang juga melaporkan dominasi sentimen positif pada ulasan ChatGPT menggunakan IndoBERT, serta memperkuat efektivitas IndoBERT dalam menangani karakteristik ulasan berbahasa Indonesia yang informal dibandingkan metode klasik seperti pada penelitian Fahlapi & Rosandi (2025).

Kesimpulan

Penelitian ini berhasil menerapkan metode IndoBERT untuk melakukan analisis dan komparasi sentimen pengguna terhadap aplikasi ChatGPT dan Gemini di Google Play Store berbahasa Indonesia. Berdasarkan hasil pengujian, model IndoBERT untuk ChatGPT memperoleh akurasi sebesar 91,32%, sedangkan model untuk Gemini memperoleh akurasi sebesar 86,98%, yang menunjukkan bahwa kedua model mampu mengklasifikasikan sentimen ulasan dengan performa yang tergolong baik. Hasil analisis distribusi sentimen menunjukkan bahwa persepsi pengguna terhadap ChatGPT cenderung lebih positif (81,51%) dibandingkan Gemini (65,55%), sementara proporsi sentimen negatif pada Gemini (34,45%) hampir dua kali lipat dibandingkan ChatGPT (18,51%). Perbedaan ini turut memengaruhi performa klasifikasi model, di mana model ChatGPT unggul pada seluruh metrik evaluasi dibandingkan model Gemini. Hasil penelitian ini dapat dijadikan acuan praktis bagi masyarakat dalam memilih platform AI generatif yang sesuai dengan kebutuhan mereka, serta menjadi referensi bagi pengembang aplikasi dalam meningkatkan kualitas layanan berdasarkan persepsi pengguna. Penelitian selanjutnya disarankan untuk memperluas cakupan data dengan periode waktu yang lebih panjang, menambahkan kelas sentimen netral, serta membandingkan performa IndoBERT dengan model transformer berbahasa Indonesia lainnya.

Daftar Pustaka

- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., dkk. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Fahlapi, R. F., & Rosandi, Y. (2025). Sentiment analysis of Gemini AI reviews on Google Play Store using machine learning methods. *Jurnal Informatika dan Rekayasa Perangkat Lunak (JIRPL)*, 7(1), 45-56.
- Gemini Team. (2023). Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*. <https://arxiv.org/abs/2312.11805>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hassani, H., & Silva, E. S. (2023). The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the field. *Big Data and Cognitive Computing*, 7(2), 62. <https://doi.org/10.3390/bdcc7020062>
- Herlinawati, H., Marlina, S., & Kurniawan, A. (2020). Analisis sentimen ulasan pengguna aplikasi berbasis Play Store menggunakan metode Naive Bayes. *Jurnal Ilmiah Teknologi Informasi*, 8(2), 112-120.
- Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, 11(1), 22. <https://doi.org/10.1186/s40561-024-00310-z>
- Kotelnikova, A., Paschenko, D., Bochenina, K., & Kotelnikov, E. (2022). Lexicon-based methods vs. BERT for text sentiment analysis. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 71-83). Springer.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

-
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., dkk. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2), 100017. <https://doi.org/10.1016/j.metrad.2023.100017>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Mursidah, S., Salim, A., & Rahmat, R. (2025). Analisis sentimen ulasan pengguna aplikasi ChatGPT di Google Play Store menggunakan IndoBERT. *Jurnal Ilmiah Teknik Informatika dan Sistem Informasi (JITSI)*, 6(1), 33-44.
- Putri, D. A., Wijaya, K., & Santoso, H. (2025). Klasifikasi sentimen ulasan aplikasi e-commerce makanan menggunakan IndoBERT. *Jurnal Sistem Informasi dan Teknologi*, 7(1), 55-66.
- Rosdiana, R., Fauzan, A., & Nugroho, W. (2025). Penerapan IndoBERT untuk analisis sentimen ulasan aplikasi transportasi online berbahasa Indonesia. *Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 78-89.
- Simarmata, J., & Sasongko, D. (2025). Analisis sentimen ulasan aplikasi perbankan digital BRImo menggunakan model IndoBERT. *Jurnal Riset Sistem Informasi*, 9(2), 101-112.
- Subhan, M., Ramadhani, F., & Kurnia, A. (2025). Analisis sentimen publik terhadap ChatGPT, Gemini, dan DeepSeek menggunakan Naive Bayes dan Logistic Regression. *Jurnal Teknik Informatika*, 17(1), 20-31.
- Tarwoto, A. R., Pradhisa, I., & Fajriyah, L. (2025). Analisis sentimen ulasan aplikasi berbasis AI di Google Play Store menggunakan model BERT berbahasa Indonesia. *Jurnal Nasional Teknologi dan Sistem Informasi (TEKNOSI)*, 11(1), 15-27.
- Wirayudha, T., Kurniawan, A., & Prasetyo, B. (2025). Efektivitas IndoBERT dalam pemrosesan bahasa alami berbahasa Indonesia: Sebuah tinjauan sistematis. *Jurnal Ilmu Komputer dan Informasi*, 18(1), 1-14.

CC BY-SA 4.0 (Attribution-ShareAlike 4.0 International).

This license allows users to share and adapt an article, even commercially, as long as appropriate credit is given and the distribution of derivative works is under the same license as the original. That is, this license lets others copy, distribute, modify and reproduce the Article, provided the original source and Authors are credited under the same license as the original.

