Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 6, No. 7, Juli 2026, Hal. 1380-1390
DOI: https://doi.org/10.52436/1.jpti.2018
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis Sentimen dan Perbandingan Ulasan Pengguna ChatGPT dan Grok Menggunakan *Support Vector Machine* (SVM)

I Nyoman Yuda Sugiantara[*1], Kadek Agus Arya Purnawan[2], Kadek Riko Redita[3], Rizal Jamalul Lail[4], Indriyani[5]

[1,2,3,4,5]Program Studi Sistem Informasi, Fakultas Informatika dan Ilmu Komputer, ITB STIKOM Bali
Email: [1]230030448@stikom-bali.ac.id, [2]230030486@stikom-bali.ac.id
[3]230030160@stikom-bali.ac.id, [4]230030467@stikom-bali.ac.id, [5]indriyani@stikom-bali.ac.id

Abstrak

Adopsi aplikasi *chatbot* berbasis AI, seperti ChatGPT dan Grok, yang menghasilkan berbagai ulasan pengguna di platform digital, telah meningkat karena kemajuan kecerdasan buatan. Analisis sentimen dari evaluasi ini dapat digunakan untuk memastikan tingkat kebahagiaan dan persepsi pengguna. Studi ini menggunakan teknik berbasis leksikon dan Support Vector Machine (SVM) untuk memeriksa dan membandingkan sentimen evaluasi pengguna ChatGPT dan Grok. Data penelitian terdiri dari 5.000 ulasan pelanggan yang dikumpulkan dari Google Play Store menggunakan teknik web scraping. Pra-pemrosesan data, pelabelan sentimen berbasis leksikon, pembagian 80:20 antara data pelatihan dan data uji, dan klasifikasi algoritma kernel SVM merupakan beberapa langkah dalam proses penelitian. Matriks kebingungan dan laporan klasifikasi dengan metrik termasuk akurasi, presisi, recall, dan F1-score digunakan untuk mengevaluasi model. Menurut temuan, model SVM memperoleh akurasi 89% pada dataset ChatGPT dan 88% pada dataset Grok. Kelas sentimen positif memperoleh nilai presisi dan F1-score tertinggi, masing-masing sebesar 0,93 dan 0,92, pada kedua dataset. Selain itu, model dataset Grok mengungguli ChatGPT dalam mengidentifikasi sentimen negatif. Secara keseluruhan, pendekatan kernel SVM menunjukkan efektivitas dalam menilai sentimen secara akurat dalam ulasan pengguna aplikasi *chatbot* AI.

Kata kunci: *analisis sentimen, SVM, lexicon-based, chatgpt, grok, machine learning, Google Play Store reviews.*

Sentiment Analysis and Comparative Study of User Reviews on ChatGPT and Grok Using Support Vector Machine and Lexicon-Based Approaches

Abstract

The adoption of AI-based chatbot applications, such as ChatGPT and Grok, which generate various user reviews on digital platforms, has increased due to advances in artificial intelligence. Sentiment analysis of these evaluations can be used to determine user happiness and perception. This study uses lexicon-based and Support Vector Machine (SVM) techniques to analyse and compare the sentiment of ChatGPT and Grok user evaluations. The research data consisted of 5,000 customer reviews collected from the Google Play Store using web scraping techniques. Pre-processing of the data, sentiment labelling based on lexicons, a 80:20 split between training and test data, and kernel SVM algorithm classification are some of the steps in the research process A confusion matrix and classification report with metrics such as accuracy, precision, recall and F1-score are used to evaluate the model. According to the findings, the SVM model achieved an accuracy of 89% on the ChatGPT dataset and 88% on the Grok dataset. The positive sentiment class obtained the highest precision and F1-score, at 0.93 and 0.92 respectively, on both datasets. Additionally, the Grok dataset model outperformed ChatGPT in identifying negative sentiment. Overall, the kernel SVM approach demonstrates effective performance in accurately evaluating sentiment in user reviews of AI chatbot applications.

Keywords: *analisis sentimen, SVM, lexicon-based, chatgpt, grok, machine learning, Google Play Store reviews.*

1. PENDAHULUAN

Kecerdasan buatan (*Artificial Intelligence*) mendorong munculnya berbagai aplikasi berbasis *chatbot* yang mampu membantu aktivitas pengguna dalam berbagai bidang. Aplikasi berbasis AI seperti *chatbot* kini banyak digunakan dalam komunikasi digital, pembelajaran, dan pencarian informasi [1]. Tingginya penggunaan aplikasi tersebut menghasilkan banyak ulasan pengguna yang menunjukan pengalaman serta persepsi terhadap kualitas layanan aplikasi. Bagi para pengembang yang ingin meningkatkan kualitas sistem, ulasan-ulasan ini merupakan

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sumber informasi yang berharga. Untuk memiliki pemahaman yang lebih baik tentang kebutuhan dan tingkat kepuasan konsumen, menganalisis ulasan pengguna sangatlah penting [2].

Analisis sentimen adalah cabang dari pemrosesan bahasa alami atau *Natural Language Processing* yang menganalisis data teks untuk menentukan sikap, perasaan, dan opini pengguna tentang suatu objek [3]. Analisis sentimen telah digunakan dalam sejumlah studi untuk meningkatkan pengambilan keputusan berbasis data di berbagai sektor, termasuk media sosial, aplikasi seluler, dan ulasan produk. Karena analisis sentimen dapat secara efisien menangani volume data yang sangat besar dan menghasilkan wawasan yang sulit dicapai melalui analisis manual, penerapannya menjadi semakin penting. [4]. Analisis sentimen dapat dilakukan menggunakan berbagai pendekatan, salah satunya adalah metode berbasis leksikon (*lexicon-based*). Pendekatan ini bekerja dengan memanfaatkan kamus kata yang telah memiliki nilai polaritas (positif atau negatif) untuk menentukan sentimen suatu teks. Metode ini memiliki keunggulan karena tidak membutuhkan data latih dan hasilnya relatif mudah untuk dipahami. Namun, metode tersebut masih memiliki keterbatasan dalam menangkap konteks bahasa yang kompleks. Sejumlah penelitian juga menunjukkan bahwa pendekatan lexicon-based cukup efektif digunakan sebagai tahap awal dalam proses pelabelan data sebelum diterapkan pada model machine learning [5]. Pendekatan *machine learning* seperti *Support Vector Machine* (SVM) banyak digunakan dalam analisis sentimen ulasan pengguna aplikasi karena memiliki kemampuan klasifikasi yang baik, khususnya pada data teks berdimensi tinggi seperti ulasan di Google Play Store [2]. Dalam penelitian ini, sentimen dalam ulasan pengguna aplikasi ChatGPT dan Grok dikategorikan menggunakan metode SVM. Cara kerja SVM adalah menemukan hyperplane terbaik untuk membagi data ke dalam setiap kelas sentimen. [6]. SVM dipilih sebagai salah satu metode utama dalam penelitian ini untuk dibandingkan dengan pendekatan *lexicon-based* dalam mengklasifikasikan sentimen ulasan pengguna.

Kombinasi antara metode *lexicon-based* dan machine learning seperti SVM dapat meningkatkan performa analisis sentimen. Metode *lexicon-based* dapat digunakan untuk memberikan label awal pada data, sedangkan SVM digunakan untuk melakukan klasifikasi yang lebih akurat. Pendekatan *hybrid* ini terbukti mampu meningkatkan akurasi dan efisiensi dalam pengolahan data teks, terutama pada dataset ulasan pengguna aplikasi. Selain itu, beberapa studi juga menekankan pentingnya membandingkan performa metode yang digunakan untuk mengetahui keunggulan dan keterbatasan masing-masing pendekatan [7]-[8].

Seiring dengan meningkatnya popularitas *chatbot* berbasis kecerdasan buatan, ChatGPT dan Grok menjadi dua aplikasi yang banyak digunakan serta memperoleh ulasan dalam jumlah besar dari pengguna di Google Play Store. Kedua aplikasi tersebut dikembangkan oleh perusahaan yang berbeda dan menawarkan karakteristik layanan yang berbeda, sehingga berpotensi menghasilkan pola sentimen pengguna yang tidak sama. Perbedaan karakteristik layanan tersebut menjadikan kedua aplikasi menarik untuk dibandingkan karena dapat memberikan gambaran mengenai variasi persepsi pengguna terhadap kualitas layanan *chatbot* berbasis AI. Analisis dan perbandingan sentimen pada ulasan pengguna ChatGPT dan Grok menjadi penting untuk mengevaluasi efektivitas pendekatan *lexicon-based* dan *Support Vector Machine* (SVM) dalam proses klasifikasi sentimen serta mengidentifikasi kecenderungan opini pengguna terhadap masing-masing aplikasi.

Penelitian terdahulu umumnya berfokus pada analisis sentimen terhadap ulasan pengguna aplikasi tertentu, seperti WhatsApp, OVO, dan berbagai aplikasi seluler lainnya, dengan tujuan mengukur persepsi pengguna terhadap layanan yang diberikan. Namun, penelitian yang membandingkan hasil pelabelan sentimen dari *chatbot* berbasis kecerdasan buatan masih relatif terbatas. Selain itu, terbatasnya literatur penelitian yang secara khusus membandingkan karakteristik sentimen pengguna antara ChatGPT dan Grok, meskipun kedua aplikasi tersebut merupakan *chatbot* AI populer yang memiliki karakteristik layanan yang berbeda. Oleh karena itu, penelitian ini bertujuan untuk menerapkan pendekatan berbasis leksikon dan algoritma *Support Vector Machine* (SVM) untuk menganalisis serta membandingkan sentimen ulasan pengguna ChatGPT dan Grok. Kebaruan penelitian ini terletak pada perbandingan dua *chatbot* AI populer sebagai sumber pelabelan sentimen, serta penerapan pendekatan *hybrid* yang menggabungkan metode *lexicon-based* dan SVM dalam proses klasifikasi. Dengan membandingkan distribusi sentimen dan performa klasifikasi yang dihasilkan dari kedua *chatbot* tersebut, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan analisis sentimen berbasis pembelajaran mesin serta memberikan pemahaman yang lebih mendalam mengenai efektivitas penggunaan *chatbot* AI dalam proses pelabelan dan kategorisasi sentimen.

2. METODE PENELITIAN

2.1. Sumber dan Jenis Data Penelitain

Penulis menggunakan analisis sentimen bersamaan dengan metode kuantitatif deskriptif. Berdasarkan ulasan yang diterima, tujuan utamanya adalah untuk menentukan, mengkategorikan, dan membandingkan sentimen pengguna terhadap ChatGPT dan Grok. Data sekunder dari ulasan pengguna di Google Play Store digunakan dalam penelitian ini. Metode ini menggabungkan SVM dengan leksikon berbasis kamus sentimen.

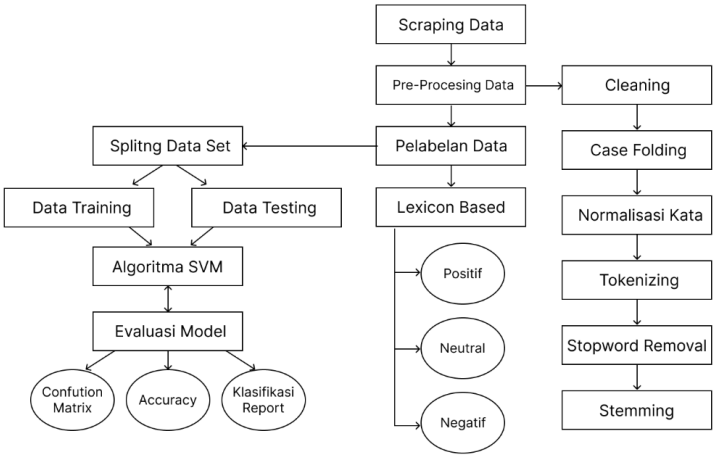
p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.2. Teknik Pengumpulan Data

Gambar 1. Teknik Pengumpulan Data

Proses pengumpulan data terdiri dari beberapa tahap. Tahap pertama adalah pengumpulan data (data scraping), yaitu proses pengambilan ulasan pengguna secara otomatis dari platform digital menggunakan bahasa pemrograman Python. Pada penelitian ini, proses scraping memanfaatkan library *google-play-scraper* untuk mengambil ulasan pengguna aplikasi ChatGPT dan Grok yang tersedia di Google Play Store. Data yang dikumpulkan meliputi isi ulasan, rating, tanggal ulasan, serta informasi relevan lainnya. Pengambilan data dilakukan melalui *Application Programming Interface* (API) yang disediakan oleh library tersebut, sehingga memungkinkan pengumpulan data dalam jumlah besar secara efektif dan efisien. Tahap berikutnya adalah seleksi dan pembersihan data (*data cleaning*), yang mencakup penghapusan data duplikat, spam, serta data yang tidak lengkap atau kosong guna memastikan kualitas dan validitas dataset yang digunakan dalam analisis. Setelah melalui proses pembersihan, data disimpan dalam format terstruktur, seperti CSV atau Excel, sehingga dapat digunakan pada tahap prapemrosesan dan analisis sentimen selanjutnya [9].

2.3. Alur Kerja Analisis



Gambar 2. Workflow Analisis Sentimen

Proses penelitian ini diawali dengan tahap pengumpulan data yang dilakukan melalui teknik *web scraping*. Data yang sudah dikumpulkan akan diproses pada tahap praproses, yang meliputi pembersihan, pengubahan huruf besar/kecil, normalisasi kata, tokenisasi, penghapusan kata penghenti (*stopword*), dan *stemming*, kemudian diterapkan pada data yang telah dikumpulkan. Prosedur ini digunakan untuk meningkatkan kualitas dan konsistensi data teks sehingga dapat dimanfaatkan seefektif mungkin pada tahap analisis selanjutnya. [10]. Setelah tahap pra-pemrosesan, pendekatan berbasis leksikon digunakan untuk memberi label sentimen pada data dimana dilakukanya proses pelabelan sentimen menggunakan pendekatan berbasis leksikon (*lexicon-based*). Pelabelan sentimen memanfaatkan InSet Lexicon Bahasa Indonesia yang terdiri atas kamus kata positif (*positive.tsv*) dan kamus kata negatif (*negative.tsv*). Proses pelabelan dilakukan dengan menghitung jumlah kata yang terdapat pada kamus positif dan kamus negatif di setiap ulasan. Dengan menggunakan teknik ini, ulasan dibagi menjadi tiga kelompok berdasarkan nilai polaritas kata-kata dalam teks: positif, negatif, dan netral [11]. Dataset yang telah dilabeli kemudian dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) guna mendukung proses pembangunan dan evaluasi model klasifikasi dengan rasio 80:20. Pemilihan rasio tersebut bertujuan untuk menyediakan jumlah data pelatihan yang cukup agar model mampu mempelajari pola sentimen secara optimal, sekaligus mempertahankan sebagian data sebagai data pengujian untuk mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah diproses sebelumnya.

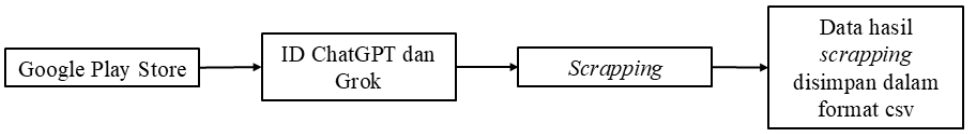
Proses klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel linear sebagai model klasifikasi. Sebelum proses klasifikasi, data teks terlebih dahulu dikonversi ke dalam bentuk representasi numerik menggunakan CountVectorizer sehingga dapat diproses oleh algoritma SVM. Algoritma SVM bekerja dengan mencari *hyperplane* optimal yang mampu memisahkan data ke dalam masing-masing kelas sentimen dengan margin terbesar [12]-[13]. Penggunaan kernel linear dipilih karena memiliki kinerja yang baik pada data teks berdimensi tinggi, menghasilkan proses komputasi yang lebih efisien, serta mampu memberikan performa klasifikasi yang optimal pada analisis sentimen berbasis teks [14].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3. HASIL DAN PEMBAHASAN

3.1. Scrapping Data

Ulasan pengguna terhadap aplikasi ChatGPT dan Grok yang dikumpulkan dari Google Play Store menjadi sumber data penelitian ini. Pengambilan data dilakukan dengan metode website scraping, dengan batasan maksimal 5.000 ulasan pengguna. Alur proses *scraping* data tersebut ditunjukkan pada Gambar 3.



Gambar 3. Alur Proses *Scrapping*

Semua data ulasan yang terkumpul disimpan dalam format CSV untuk memungkinkan pemrosesan lebih lanjut setelah operasi pengambilan data selesai. Fungsi export_to_csv digunakan untuk menyimpan data, sehingga mudah diakses dan terstruktur. Berikut merupakan hasil *scrapping* dari aplikasi Chat GPT dan Grok tersebut disajikan pada Gambar 4 dan Gambar 5.

	Review ID	Username	Rating	Review Text	Date
0	8a9a3f01-a484-4243-8a7e-b87ee807f761	Pengguna Google	5	aPKnya bagus banget	2026-05-17 01:20:54
1	c3d40ca0-794a-4fde-8d5e-4a233bc9f7eb	Pengguna Google	5	Aplikasi keren luar biasa secara otomatis bisa...	2026-05-17 01:19:00
2	c644d8ad-3951-4e51-8533-b1b52be989d3	Pengguna Google	5	👍👍👍👍👍👍👍	2026-05-17 01:17:23
3	35754caa-7039-4cad-854f-981ee263c878	Pengguna Google	5	sangat membantu	2026-05-17 01:16:15
4	015ed944-8f7c-4137-9876-54273369c458	Pengguna Google	4	aplikasinya bagus sekali saya puas gunakan ini...	2026-05-17 01:15:29

Gambar 4. Hasil *Scrapping* ChatGPT

	Review ID	Username	Rating	Review Text	Date
0	94bd4473-b404-4d1a-ba3f-3339d1d59647	Pengguna Google	5	baik dan ramah	2026-05-17 01:26:29
1	4c25a77c-c580-445c-b435-1ed617794ea6	Pengguna Google	5	keren	2026-05-17 01:17:51
2	31158bcb-b641-4e05-98f5-75e0f98a5fd0	Pengguna Google	1	kecewa	2026-05-17 01:10:15
3	3fb0a944-1050-42cd-84ea-6596de90d7c4	Pengguna Google	5	keren banget, tapi sayang sekarang berbayar	2026-05-17 00:40:52
4	8d80ce10-9952-49e6-b14e-ee2d6651c7df	Pengguna Google	5	asli roasting nya gacor	2026-05-17 00:10:27

Gambar 5. Hasil *Scrapping* Grok

3.2. *Pre-Processing*

Langkah pertama dalam analisis sentimen adalah pra-pemrosesan, yang bertujuan untuk membersihkan dan mempersiapkan data teks agar sistem dapat menganalisisnya dengan lebih efisien. Pada tahap ini, sejumlah prosedur, termasuk *cleaning*, *case folding*, normalisasi, tokenisasi, penghapusan kata-kata penghenti (stopword), dan stemming, digunakan untuk mengubah input teks yang tidak terstruktur menjadi format yang lebih terstruktur. [15]. Berikut merupakan hasil *pre-proccesing* ulasan ChatGPT dan Grok pada Gambar 6 dan Gambar 7.

	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi	tokenize	stopword removal	steming_data
0	2026-05-30 07:14:52	Pengguna Google	5	bagus	bagus	bagus	bagus	[bagus]	[bagus]	bagus
1	2026-05-30 07:13:17	Pengguna Google	3	bring back gpt-4	bring back gpt	bring back gpt	bring back gpt	[bring, back, gpt]	[bring, back, gpt]	bring back gpt
2	2026-05-30 07:12:33	Pengguna Google	5	bgus	bgus	bgus	bagus	[bagus]	[bagus]	bagus
3	2026-05-30 07:06:42	Pengguna Google	5	bijak dan santun	bijak dan santun	bijak dan santun	bijak dan santun	[bijak, dan, santun]	[bijak, santun]	bijak santun
4	2026-05-30 07:06:15	Pengguna Google	1	BUTUT	BUTUT	butut	butut	[butut]	[butut]	butut

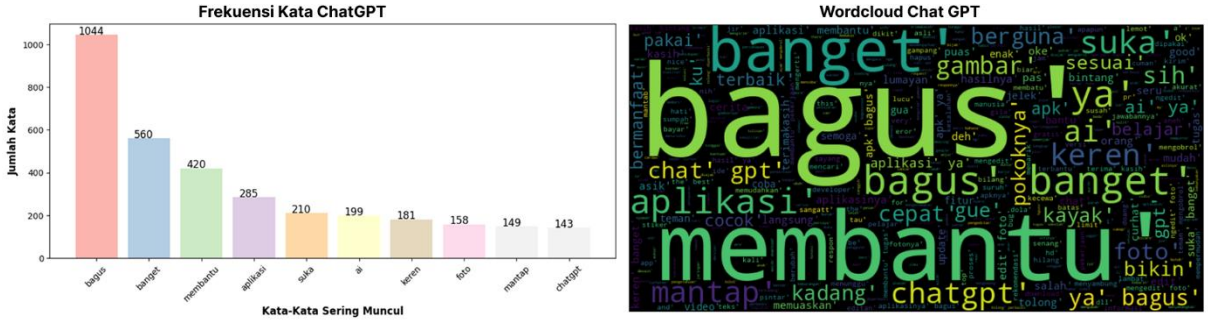
Gambar 6. Hasil *Pra-Scrapping* ChatGPT

p-ISSN: 2775-4227
e-ISSN: 2775-4219

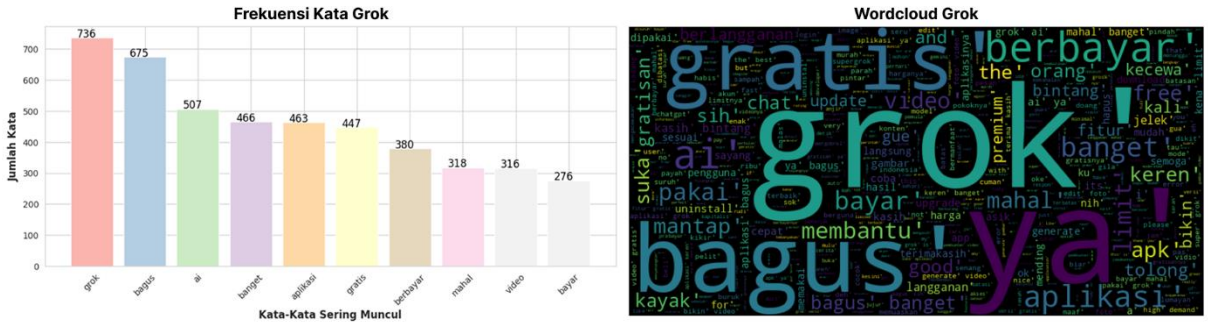
	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi	tokenize	stopword removal	steming_data
0	2026-05-30 06:05:19	Pengguna Google	5	sangat membantu	sangat membantu	sangat membantu	sangat membantu	[sangat, membantu]	[membantu]	bantu
1	2026-05-30 05:52:10	Pengguna Google	5	nice	nice	nice	nice	[nice]	[nice]	nice
2	2026-05-30 05:39:05	Pengguna Google	5	kerennn	kerennn	kerennn	keren	[keren]	[keren]	keren
3	2026-05-30 05:19:16	Pengguna Google	5	berpikir kritis. ai sebagus ini kok pemerintah...	berpikir kritis ai sebagus ini kok pemerintah ...	berpikir kritis ai sebagus ini kok pemerintah ...	berpikir kritis ai sebagus ini kok pemerintah ...	[berpikir, kritis, ai, sebagus, ini, kok, peme...	[berpikir, kritis, ai, sebagus, pemerintah, ng...	pikir kritis ai bagus perintah ngide blokir sa...
4	2026-05-30 05:05:51	Pengguna Google	5	gk tau	gk tau	gk tau	tidak tau	[tidak, tau]	[tau]	tau

Gambar 7. Hasil *Pra-Scrapping* Grok

Berikut merupakan hasil frekuensi kata dan worldcloud yang muncul setelah *pre-proccesing* ulasan ChatGPT dan Grok pada Gambar 8 dan Gambar 9.



Gambar 8. Frekuensi Kata dan Wordcloud *Pre-proccesing* ChatGPT

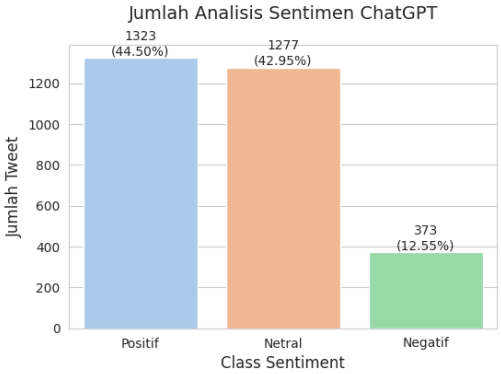


Gambar 9. Frekuensi Kata dan Wordcloud *Pre-proccesing* Grok

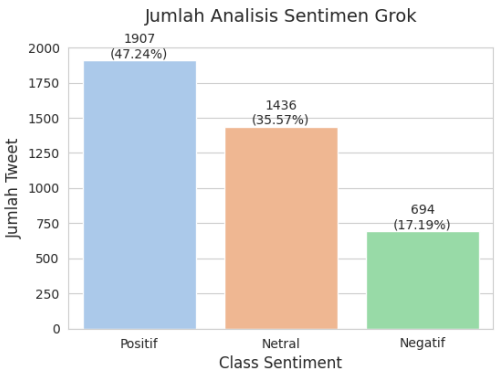
3.3. Pelabelan Sentimen

Pelabelan pada data menggunakan *lexicon-based* dimana pelabelan adalah salah satu proses pemberian kategori atau kelas pada data berupa teks berdasarkan sentimen yang terkandung di dalamnya, seperti sentimen positif, negatif, atau netral [16]. Metode *Lexicon-based* menghasilkan label sentimen tanpa memerlukan perlatihan model, sehingga dapat digunakan sebagai *ground truth* untuk tahap *supervised machine learning.* Tujuan dari langkah ini adalah untuk menyediakan data yang lebih terstruktur untuk digunakan dalam pengujian dan pelatihan model analisis sentimen. Kualitas pelabelan sangat berpengaruh terhadap performa model klasifikasi, karena label yang tepat dapat membantu sistem mengenali pola sentimen dengan lebih akurat. Berikut merupakan hasil dari pelabelan sentimen pada ulasan ChatGPT dan Grok pada Gambar 10 dan Gambar 11.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 10. Hasil Pelabelan ChatGPT



Gambar 11. Hasil Pelabelan Grok

Hasil pelabelan menunjukkan bahwa kedua chatbot tersebut mengekspresikan sentimen dengan cara yang berbeda. ChatGPT menghasilkan proporsi sentimen netral yang lebih tinggi, sedangkan Grok cenderung mengkategorikan ulasan sebagai positif atau negatif. Hal ini menunjukkan bahwa ChatGPT lebih berhati-hati dalam menentukan apakah suatu ulasan bersifat positif atau negatif, terutama jika ulasan tersebut ambigu. Sebaliknya, Grok lebih baik dalam membedakan antara ulasan positif dan negatif. Perbedaan ini menunjukkan bahwa pendekatan masing-masing model dalam memberi label sentimen sangat dipengaruhi oleh pemahaman mereka terhadap konteks ulasan pengguna. Hal ini mengindikasikan adanya perbedaan karakteristik dan sensitivitas kedua model dalam melakukan klasifikasi sentimen terhadap data tweet yang dianalisis.

Gambar 12. *Wordcloud* Sentimen Negatif dan Positif dan Negatif ChatGPT

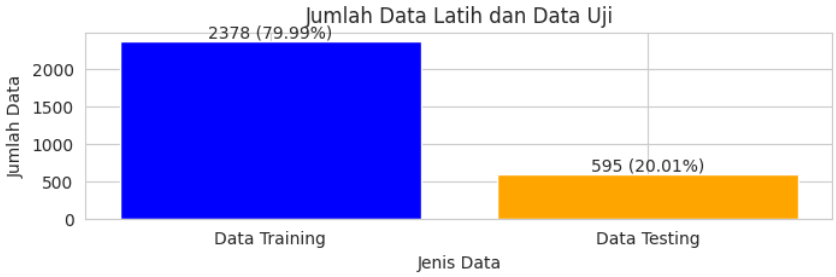
Gambar 13. *Wordcloud* Sentimen Negatif dan Positif dan Negatif Grok

Wordcloud menunjukkan bahwa sentimen positif terhadap kedua aplikasi tersebut sebagian besar berasal dari kata-kata yang membahas manfaat dan seberapa baik layanannya. Pada ulasan ChatGPT, kata-kata “bantuan” dan “bagus” menunjukkan bahwa pengguna sangat menyukai cara aplikasi ini membantu mereka dalam pekerjaan dan studi. Sementara itu, untuk Grok, penggunaan kata “gratis” menunjukkan bahwa kemudahan akses menjadi alasan utama kepuasan pengguna. Di sisi lain, sentimen negatif terhadap kedua aplikasi tersebut terutama disebabkan oleh kata-kata yang berkaitan dengan masalah layanan dan kendala teknis, yang menunjukkan bahwa pengalaman pengguna dan seberapa stabil aplikasi tersebu. Secara keseluruhan, kedua model menunjukkan pola sentimen yang berbeda, di mana ChatGPT lebih banyak dikaitkan dengan kualitas hasil dan bantuan penggunaan, sedangkan Grok lebih sering dikaitkan dengan aspek fitur layanan dan kendala teknis.

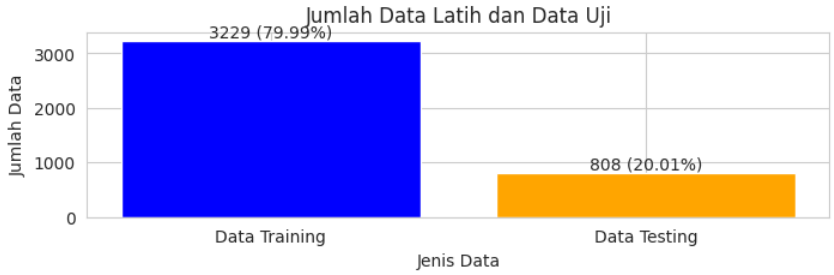
p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.4. *Data Splitting*

Setelah proses pelabelan selesai, dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Untuk membantu model mengidentifikasi pola dan karakteristik sentimen yang ada dalam dataset, 80% data digunakan sebagai data pelatihan. Sementara itu, 20% sisanya digunakan sebagai data pengujian untuk mengukur seberapa baik kinerja model pada data baru yang belum dianalisis. Karena dianggap menghasilkan proses pelatihan dan evaluasi model yang lebih efisien dan mengurangi kemungkinan bias dalam temuan pengujian, rasio partisi data 80:20 banyak digunakan dalam penelitian analisis sentimen [17].



Gambar 14. Pembagian Jumlah Data Latih dan Data Uji ChatGPT



Gambar 15. Pembagian Jumlah Data Latih dan Data Uji Grok

3.5. Hasil Evaluasi ChatGPT dan Grok dengan *Support Vector Machine* (SVM)

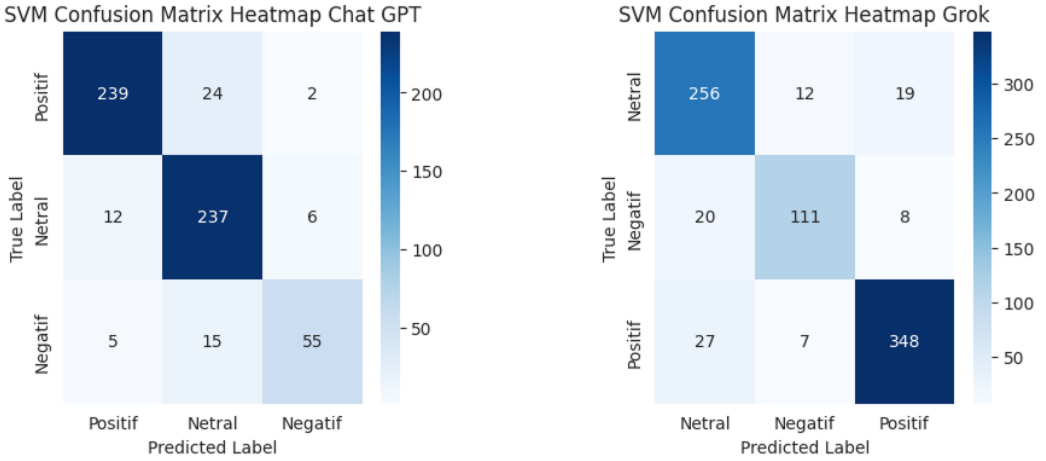
Evaluasi kinerja model dilakukan dengan menggunakan *confusion matrix* yang mencakup parameter *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* merupakan metode evaluasi yang digunakan untuk membandingkan hasil prediksi model dengan data sebenarnya pada proses klasifikasi. *Accuracy* adalah suatu model ditentukan oleh seberapa baik model tersebut memprediksi semua data, *Precision* merupakan ukuran yang digunakan untuk mengetahui tingkat ketepatan prediksi positif yang dihasilkan oleh model. *Recall* adalah ukuran yang menunjukkan kemampuan model dalam mengidentifikasi seluruh data positif yang tersedia. Sementara itu, *F1-Score* merupakan metrik evaluasi yang menilai keseimbangan kinerja model klasifikasi dengan menggabungkan presisi dan recall [6]. Pada penelitian ini, algoritma Support Vector Machine (SVM) menggunakan kernel linear (kernel='linear'). Pemilihan kernel linear didasarkan pada karakteristik data teks yang telah direpresentasikan dalam bentuk fitur numerik menggunakan metode CountVectorizer. Kernel linear banyak digunakan pada klasifikasi teks karena mampu menangani data berdimensi tinggi secara efektif serta memiliki kompleksitas komputasi yang lebih rendah dibandingkan kernel nonlinier. Dengan penggunaan kernel linear, model SVM dapat membentuk hyperplane optimal untuk memisahkan kelas sentimen positif, netral, dan negatif berdasarkan fitur-fitur yang dihasilkan dari proses ekstraksi teks. Selain itu, penelitian ini juga melakukan perbandingan hasil sentimen antara aplikasi ChatGPT dan Grok untuk mengidentifikasi kecenderungan opini pengguna terhadap masing-masing layanan AI *chatbot*. Hasil evaluasi tersebut digunakan untuk menilai efektivitas algoritma SVM dengan kernel linear dalam melakukan klasifikasi sentimen, sekaligus memberikan gambaran mengenai tingkat kepuasan pengguna terhadap kedua aplikasi berdasarkan ulasan yang diberikan.

a. *Confusion Matrix*

Confusion matrix merupakan metode evaluasi yang digunakan untuk membandingkan hasil prediksi model dengan data sebenarnya. Pada penelitian ini, *confusion matrix* diterapkan untuk mengevaluasi kemampuan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan pengguna ke dalam tiga kategori, yaitu positif, negatif, dan netral. Melalui *confusion matrix*, dapat diperoleh informasi mengenai komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada masing-masing kelas sentimen, yang selanjutnya digunakan untuk menghitung nilai *accuracy*, *precision*, *recall*, *dan F1-*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

score. Pada visualisasi *confusion matrix*, nilai yang berada pada diagonal utama menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar oleh model, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi, yaitu data yang diprediksi ke kelas sentimen yang berbeda dari label sebenarnya. Dengan adanya tiga kelas sentimen, *confusion matrix* mampu memberikan gambaran yang lebih rinci mengenai kemampuan model dalam membedakan setiap kategori sentimen. Semakin besar nilai pada diagonal utama dan semakin kecil nilai di luar diagonal, maka semakin baik performa model klasifikasi yang dihasilkan.



Gambar 16. *Confusion Matrix* ChatGPT dan Grok

Berdasarkan perbandingan kedua *confusion matrix* tersebut, model klasifikasi sentimen pada aplikasi Grok memiliki performa yang lebih baik dibandingkan ChatGPT karena mampu menghasilkan jumlah prediksi benar yang lebih tinggi pada seluruh kategori sentimen, khususnya pada kelas positif dan netral. Selain itu, nilai kesalahan prediksi pada Grok relatif lebih kecil dibandingkan jumlah data yang berhasil diklasifikasikan dengan benar, sehingga menunjukkan bahwa model SVM kernel bekerja lebih optimal pada dataset ulasan Grok. Namun, kedua model masih mengalami kesalahan klasifikasi terutama antara sentimen positif dan netral karena karakteristik opini pengguna yang memiliki makna hampir serupa.

b. *Classification Report* dan *Accuracy*

Classification report merupakan model klasifikasi yang menyajikan nilai *precision*, *recall*, *F1-score*, dan *support* setiap kategori sentimen. klasifikasi [10]. Pada penelitian ini, *classification report* digunakan untuk mengevaluasi kinerja algoritma *Support Vector Machine* dalam mengelompokkan sentimen positif, netral, dan negatif pada ulasan pengguna aplikasi ChatGPT dan Grok. Dengan menggunakan *classification report*, peneliti dapat memahami kemampuan model dalam memprediksi setiap kategori sentimen secara lebih mendalam dibandingkan hanya mengandalkan nilai *accuracy*.

Classification report memiliki keterkaitan yang dengan *confusion matrix* karena seluruh metrik evaluasi yang ditampilkan diperoleh berdasarkan hasil perhitungan pada *confusion matrix*. Dengan demikian, *confusion matrix* berperan sebagai representasi evaluasi dasar yang menggambarkan distribusi hasil prediksi model terhadap data aktual, sedangkan classification report menyajikan interpretasi kuantitatif dari hasil tersebut dalam bentuk metrik evaluasi yang lebih terstruktur. Kombinasi kedua metode evaluasi ini memungkinkan analisis performa model klasifikasi dilakukan secara lebih komprehensif, objektif, dan sistematis.

Tabel 1. Hasil *Classification Report* Model ChatGPT

Class	Precision	Recall	F1-Score	Support
Negatif	0.87	0.73	0.80	75
Netral	0.86	0.93	0.89	225
Positif	0.93	0.90	0.92	265
Accuracy			0.89	595
Macro Avg	0.89	0.85	0.87	595
Weighted Avg	0.89	0.89	0.89	595

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 2. Hasil *Classification Report* Model Grok

Class	Precision	Recall	F1-Score	Support
Negatif	0.85	0.80	0.83	139
Netral	0.84	0.89	0.87	287
Positif	0.93	0.91	0.92	382
Accuracy			0.88	808
Macro Avg	0.88	0.87	0.87	808
Weighted Avg	0.89	0.88	0.88	808

Berdasarkan *classification report*, kedua model menunjukkan performa yang stabil dengan nilai *precision*, *recall*, dan *F1-score* yang sebagian besar berada di atas 0,80. Kelas sentimen positif memperoleh performa terbaik pada kedua dataset, menunjukkan bahwa pola bahasa pada ulasan positif relatif lebih konsisten sehingga lebih mudah dipelajari oleh model SVM. Sebaliknya, performa pada kelas negatif masih lebih rendah karena variasi ekspresi keluhan pengguna lebih beragam dibandingkan sentimen positif. Meskipun demikian, model pada dataset Grok menunjukkan kemampuan yang lebih baik dalam mengenali sentimen negatif dibandingkan model pada dataset ChatGPT. Pada akurasi, terdapat selish sebesar 1% menunjukkan bahwa algoritma SVM dengan kernel linear memiliki performa yang relatif konsisten pada kedua dataset. Perbedaan tersebut mengindikasikan bahwa karakteristik ulasan pengguna ChatGPT dan Grok tidak memberikan pengaruh yang signifikan terhadap kemampuan model dalam membentuk hyperplane pemisah antar kelas. Oleh karena itu, perbedaan utama kedua model lebih terlihat pada kemampuan mengenali kelas sentimen tertentu dibandingkan pada akurasi keseluruhan. Kemampuan model pada dataset Grok dalam mengenali sentimen negatif yang lebih baik diduga dipengaruhi oleh distribusi data negatif yang lebih besar dibandingkan dataset ChatGPT. Jumlah sampel negatif yang lebih banyak memungkinkan model mempelajari variasi pola bahasa yang digunakan pengguna ketika menyampaikan keluhan, sehingga meningkatkan kemampuan model dalam membedakan sentimen negatif dari kelas lainnya. Sebaliknya, proporsi sentimen negatif yang lebih sedikit pada dataset ChatGPT menyebabkan model memiliki informasi yang lebih terbatas untuk mengenali karakteristik kelas tersebut.

4. KESIMPULAN

4.1. Kesimpulan

Berdasarkan hasil penggunaan matriks kebingungan dan laporan klasifikasi, model *Support Vector Machine* (SVM) pada kedua dataset dapat memberikan hasil klasifikasi terbaik untuk pengguna aplikasi ChatGPT dan Grok. Tingkat akurasi yang diperoleh pada dataset ChatGPT mencapai 89%, sedangkan dataset Grok memperoleh nilai sebesar 88%. Hal ini menunjukkan bahwa metode kernel SVM memiliki kemampuan untuk secara efektif mengekstrak sentimen positif, negatif, dan netral dari input pengguna. Hasil terbaik terlihat pada kategori sentimen positif dataset ChatGPT, dengan *precision* sekitar 0,93, *recall* sekitar 0,90, dan *F1-Score* sekitar 0,92. Hasil ini menunjukkan bahwa model dapat mengidentifikasi ulasan positif dengan tingkat akurasi yang tinggi. Selain itu, kategori netral pada dataset ini memiliki nilai *recall* yang tinggi yaitu 0,93, menunjukkan bahwa model sangat akurat dalam mengidentifikasi opini netral. Di sisi lain, dataset Grok juga menunjukkan kinerja klasifikasi yang stabil. Kategori positif memiliki akurasi sekitar 0,93, *recall* 0,91, dan *F1-Score* 0,92. Sebaliknya, pada kategori negatif, model Grok menghasilkan *F1-Score* 0,83 dan *recall* sekitar 0,80. Nilai ini lebih tinggi daripada model pada dataset ChatGPT, yang hanya memiliki *recall* 0,73 dan *F1-Score* 0,80. Kondisi ini menunjukkan bahwa model yang menggunakan dataset Grok memiliki kemampuan yang lebih baik dalam mendeteksi ulasan negatif.

Hasil pada *confusion matrix* juga memperlihatkan bahwa sebagian besar data berhasil diprediksi sesuai kelas sebenarnya. Walaupun demikian, masih ditemukan beberapa kesalahan klasifikasi, khususnya antara sentimen positif dan netral. Hal tersebut terjadi karena adanya kemiripan konteks maupun penggunaan kata pada opini pengguna yang menyebabkan model sulit membedakan kedua kategori tersebut secara sempurna. Secara umum, penelitian ini menunjukkan bahwa algoritma *Support Vector Machine*, yang berbasis pada kernel yang efisien, digunakan dalam analisis sentimen untuk aplikasi *chatbot* AI. Meskipun akurasi dataset ChatGPT agak lebih tinggi, model pada dataset Grok menunjukkan kinerja yang lebih konsisten dan seimbang, terutama dalam menangani sentimen negatif pengguna.

4.2. Diskusi

Hasil penelitian ini mendukung temuan dari berbagai penelitian sebelumnya yang menunjukkan bahwa metode *Support Vector Machine* (*SVM*) memiliki kemampuan yang efektif dalam proses klasifikasi teks maupun analisis sentimen. Hal tersebut disebabkan karena *SVM* mampu membentuk pemisah antar kelas secara optimal

p-ISSN: 2775-4227
e-ISSN: 2775-4219

sehingga proses pengelompokan data menjadi lebih akurat. Berdasarkan hasil *confusion matrix*, sebagian besar data ulasan berhasil diprediksi sesuai dengan label aslinya, walaupun masih ditemukan beberapa kesalahan klasifikasi antara sentimen positif dan netral karena adanya kemiripan konteks dan makna dalam ulasan pengguna. Pada kategori sentimen negatif, model yang diterapkan pada dataset Grok memperlihatkan nilai *recall* dan *F1-Score* yang lebih tinggi dibandingkan model pada dataset ChatGPT, sehingga kemampuan model dalam mendeteksi ulasan bernada negatif terlihat lebih konsisten.

Perbedaan kinerja tersebut kemungkinan disebabkan oleh karakteristik bahasa dalam masing-masing kumpulan data. Ulasan negatif tentang aplikasi Grok sering kali menggunakan kata-kata yang lebih langsung seperti "kesalahan," "batasan," "bug," atau "macet," yang membantu memberikan gambaran yang lebih jelas mengenai fitur-fitur aplikasi saat menggunakan *CountVectorizer* untuk analisis. Di sisi lain, ulasan negatif tentang ChatGPT biasanya menyertakan kritik bersama dengan saran, perbandingan, atau pujian terhadap fitur tertentu. Pola bahasa ini membuat kata-kata negatif muncul bersamaan dengan kata-kata positif atau netral, sehingga lebih sulit untuk membedakan antara berbagai jenis perasaan. Kondisi ini membuat model SVM lebih sulit untuk membuat garis pemisah terbaik pada dataset ChatGPT dibandingkan pada dataset Grok.

Penelitian ini menunjukkan bahwa analisis sentimen dapat dimanfaatkan sebagai sarana untuk mengukur tingkat kepuasan pengguna terhadap aplikasi *AI chatbot* sekaligus menjadi bahan pertimbangan bagi pengembang dalam melakukan peningkatan kualitas layanan aplikasi. Temuan penelitian ini juga menunjukkan bahwa performa algoritma Support Vector Machine (SVM) tidak hanya dipengaruhi oleh pemilihan kernel, tetapi juga oleh karakteristik fitur bahasa pada dataset yang direpresentasikan menggunakan CountVectorizer. Representasi berbasis frekuensi kata memungkinkan model membedakan kelas sentimen secara lebih efektif apabila suatu kategori memiliki kosakata yang lebih konsisten dan polaritas yang lebih tegas. Sebaliknya, keberadaan kosakata yang ambigu atau tumpang tindih antar kelas sentimen dapat menurunkan kemampuan model dalam membentuk hyperplane pemisah secara optimal. Dengan demikian, kualitas representasi fitur dan karakteristik linguistik dataset merupakan faktor penting yang turut menentukan performa klasifikasi selain pemilihan algoritma yang digunakan. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, di antaranya jumlah dataset yang digunakan belum terlalu besar, penggunaan algoritma klasifikasi yang masih terbatas pada satu metode, serta proses pelabelan sentimen yang berpotensi dipengaruhi oleh subjektivitas. Oleh karena itu, penelitian lebih lanjut didorong untuk menggunakan kumpulan data yang lebih besar dan membandingkan berbagai teknik pembelajaran mesin sehingga hasil klasifikasi dapat lebih akurat dan ideal.

DAFTAR PUSTAKA

[1] D. Fitriono, R. Indriati, and A. Ristyawan, "Analisis Sentimen Ulasan Aplikasi Chatgpt Menggunakan Algoritma Support Vector Machine dan Lexicon Based," vol. 3, no. 2, pp. 101–111, 2025, doi: https://doi.org/10.53624/jsitik.v3i2.719.

[2] A. Suharman, M. K. Sulaeman, T. Industri, U. Muhammadiyah, and P. Hamka, "Analisis Sentimen Pengguna Aplikasi Livin' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec," vol. 5, no. 8, pp. 2201–2212, 2025, doi: https://doi.org/10.52436/1.jpti.941.

[3] W. Andriyani, Y. Astuti, and B. A. Wisesa, "Analisis Sentimen pada Ulasan Produk dengan SVM dan Word2Vec Sentiment Analysis on Product Reviews with SVM and Word2Vec," vol. 8, no. 1, pp. 173–185, 2024, doi: 10.26798/jiko.v8i1.1498.

[4] A. Tujahidah, M. Julkarnain, S. A. Oktavia, W. Yunanri, and S. Esabella, "Analisis Sentimen Facebook & Instagram tentang Pilgub NTB 2024 dengan Algoritma SVM," vol. 4, no. 1, pp. 16–22, 2025, doi: 10.47065/mis.v2i3.

[5] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest," no. 54, 2023, doi: https://doi.org/10.33884/jif.v11i02.7987.

[6] S. Ernawati and R. Wati, "Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon," vol. 15, no. 1, pp. 40–49, 2024, doi: https://doi.org/10.24002/jbi.v15i1.7925.

[7] E. D. Shanty, B. A. Jayadi, and V. C. Mawardi, "Analisa Sentimen Ulasan Aplikasi Transportasi Menggunakan Metode Svm Dengan Pendekatan Inset Lexicon-Based," vol. 9, no. 1, pp. 99–108, 2025, doi: https://doi.org/10.24912/19y15a17.

[8] M. R. A. Yudianto, A. Rahim, P. Sukmasetya, and R. A. Hasani, "Perbandingan Metode Support Vector Machine Dengan Metode Lexicon Dalam Analisis Sentimen Bahasa," vol. 6, no. 1, 2022, doi: 10.36294/jurti.v6i1.2537.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[9] M. J. Setiawan, V. Rahmayanti, and S. Nastiti, "DANA App Sentiment Analysis: Comparison of XGBoost, SVM, and Extra Trees," vol. 13, no. 3, pp. 337–345, 2024, doi: 10.32736/sisfokom.v13i3.2239.

[10] A. R. Sitorus, W. John, P. Ziraluho, and I. M. S. S, "Analisis Sentimen Game Mobile Legends Berdasarkan Review Pengguna Di Playstore Menggunakan Algoritma Naive Bayes," vol. 4, no. 2, pp. 108–113, 2024, doi: https://doi.org/10.29407/avw8b190.

[11] M. Iqbal, M. Afdal, and R. Novita, "Implementasi Algoritma Support Vector Machine untuk Analisa Sentimen Data Ulasan Aplikasi Pinjaman Online Di Google Play Store," vol. 4, no. 4, pp. 1244–1252, 2024, doi: https://doi.org/10.57152/malcom.v4i4.1435.

[12] F. Fitriana and H. Setiawan, "Performance Analysis of SVM In Emotion Classification : A Comparative Study Of TF-IDF and Countvectorizer," vol. 6, no. 2, pp. 133–145, 2025, doi: https://doi.org/10.59562/jessi.v6i2.8396.

[13] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksmana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer Dike : Jurnal Ilmu Multidisiplin," vol. 1, no. 2, pp. 37–43, 2023, doi: https://doi.org/10.69688/dike.v1i2.35.

[14] A. Faradisia and M. A. I. Pakereng, "Analisis Komparatif Kernel Linear, Polynomial, RBF, dan Sigmoid pada Support Vector Machine untuk Klasifikasi Penyakit Jantung," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 4, pp. 1531–1537, 2025, doi: https://doi.org/10.57152/malcom.v5i4.2321.

[15] A. Salman and Kusnawi, "Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee," vol. 12, no. 1, pp. 2766–2776, 2023, doi: https://doi.org/10.33022/ijcs.v12i5.3304.

[16] S. A. Nugraha, "Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara," vol. 9, no. 3, pp. 4949–4957, 2025, doi: 10.36040/jati.v9i3.13836.

[17] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," vol. 15, no. 2, pp. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.