Jurnal Artificial, Vol 4 No 1, April 2026

Evaluasi Sentimen Ulasan Aplikasi DeepSeek di Google Play Store menggunakan Algoritma Decision Tree

Muhammad Ishak [1*], Andi Wafda [2], Ryan Alghazali Pakkaja [3]

[1, 2, 3] Politeknik Dewantara, Indonesia

* ishak@polidewa.ac.id

Abstrak

Aplikasi kecerdasan buatan (AI) D*eepSeek* telah menjadi fenomena baru dalam teknologi asisten virtual, bersaing ketat dengan platform mapan lainnya. Ulasan pengguna di Google Play Store menyediakan data krusial untuk mengevaluasi persepsi publik terhadap kinerja aplikasi ini. Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna *DeepSeek* serta mengidentifikasi faktor dominan penyebab kepuasan maupun ketidakpuasan. Penelitian menggunakan pendekatan eksperimen komputasi, yaitu melakukan pengujian model klasifikasi sentimen berbasis *supervised learning* (*Decision Tree*) pada data ulasan yang diperoleh melalui scraping Google Play Store dan direpresentasikan menggunakan fitur TF-IDF. Metodologi penelitian dimulai dengan pengumpulan data sebanyak 3506 ulasan melalui teknik *scraping*, diikuti tahap pra-pemrosesan teks yang meliputi *case* folding, *cleaning,* normalisasi kata tidak baku dan *stopword removal* untuk menangani variasi bahasa informal. Klasifikasi sentimen dilakukan menggunakan algoritma Decision Tree dengan ekstraksi fitur TF-IDF, menerapkan pelabelan otomatis berbasis skor/rating. Hasil evaluasi menunjukkan model mampu mengklasifikasikan sentimen dengan akurasi sebesar 75,93%. Analisis mendalam menggunakan N-Gram mengungkapkan temuan signifikan: sentimen negatif didominasi oleh kendala infrastruktur teknis dengan tingginya frekuensi frasa "server busy please" dan "try again later", sedangkan sentimen positif berpusat pada aspek fungsionalitas dengan kata kunci "sangat membantu" dan "bagus". Temuan ini mengindikasikan bahwa perbaikan stabilitas server dan skalabilitas layanan berpotensi menjadi strategi utama untuk meningkatkan persepsi pengguna.

Keywords: *Analisis Sentimen, Algoritma Decision Tree, Aplikasi DeepSeek*

Pendahuluan

Perkembangan teknologi kecerdasan buatan atau *Artificial Intelligence* (AI) telah mengalami lonjakan eksponensial dalam beberapa tahun terakhir, mengubah cara manusia berinteraksi dengan informasi. Inovasi-inovasi yang dibawa oleh AI, termasuk asisten virtual yang cerdas, semakin mengukuhkan peran teknologi dalam kehidupan sehari-hari. Salah satu platform yang baru-baru ini menarik perhatian publik adalah DeepSeek, sebuah aplikasi berbasis *Large Language Model* (LLM) yang menawarkan kemampuan pemrosesan bahasa alami/*Natural Language Processing* (NLP) yang canggih (Darulanda et al, 2024).

Aplikasi ini memungkinkan penggunanya untuk berinteraksi dengan teknologi secara lebih intuitif dan efisien, namun juga menghadapi tantangan untuk terus mempertahankan relevansi di pasar yang semakin kompetitif. Dalam industri aplikasi, umpan balik dari pengguna menjadi salah satu sumber informasi yang sangat penting bagi pengembang (Lisdiyanto et al., 2025).

https://doi.org/10.54065/artificial.4.1.2026.1129

Platform distribusi aplikasi seperti Google Play Store memungkinkan pengguna untuk memberikan ulasan tentang aplikasi yang mereka gunakan (Husain et al., 2024). Ulasan ini, yang berupa *User Generated Content* (UGC), memberikan gambaran yang sangat jelas mengenai pengalaman pengguna, baik yang positif maupun negatif (Umbara, 2021). Ulasan tersebut mencerminkan tingkat kepuasan, keluhan, serta ekspektasi pengguna terhadap aplikasi secara jujur dan *real-time*, dan menjadi bahan evaluasi yang berharga bagi pengembang untuk meningkatkan kualitas produk mereka (Wily et al., 2025). Namun, dengan volume ulasan yang terus bertambah setiap harinya, pengumpulan dan analisis informasi ini secara manual menjadi tidak mungkin dilakukan.

Namun, volume ulasan yang terus bertambah setiap harinya menghadirkan tantangan tersendiri. Data ulasan tersebut bersifat tidak terstruktur (*unstructured text*), mengandung banyak variasi bahasa tidak baku, dan berjumlah ribuan, sehingga mustahil untuk dianalisis secara manual. Oleh karena itu, diperlukan penerapan teknik Analisis Sentimen (*Sentiment Analysis*) atau *Opinion Mining*. Analisis sentimen memungkinkan ekstraksi informasi subjektif dari teks untuk menentukan polaritas opini, apakah bersifat positif atau negatif, yang kemudian dapat digunakan sebagai dasar pengambilan keputusan strategis bagi pengembang aplikasi (Wati et al., 2025).

Sehingga, pengembang aplikasi dapat memperoleh wawasan yang lebih terstruktur mengenai persepsi pengguna terhadap aplikasi mereka. Teknik ini sangat penting karena dapat memberikan dasar yang kuat untuk pengambilan keputusan strategis yang berbasis pada data pengguna yang nyata dan relevan. Selain itu, dengan menggunakan analisis sentimen, pengembang dapat memahami aspek apa saja dari aplikasi yang perlu diperbaiki atau ditingkatkan agar dapat lebih memenuhi harapan pengguna. Namun, analisis sentimen terhadap ulasan aplikasi tidaklah mudah. Data yang dihasilkan berupa teks tidak terstruktur yang mengandung banyak variasi bahasa tidak baku dan sering kali memiliki konteks yang ambigu (Karim et al, 2025).

Oleh karena itu, diperlukan penerapan teknik yang mampu mengelola data teks ini secara efisien. Salah satu pendekatan yang populer dalam analisis sentimen adalah penggunaan algoritma pembelajaran mesin (*Machine Learning*), Dalam melakukan klasifikasi sentimen teks, pemilihan algoritma memegang peranan krusial. dengan salah satu yang paling banyak digunakan adalah Decision Tree. Salah satu algoritma *Machine Learning* yang populer dan terbukti efektif untuk klasifikasi teks adalah *Decision Tree*. Algoritma ini memiliki keunggulan struktur *white-box*, yang berarti model keputusan yang dihasilkan mudah diinterpretasikan oleh manusia dalam bentuk aturan *if-then*, berbeda dengan algoritma *black-box* seperti *Neural Networks* yang sulit dilacak alur keputusannya (Prasetyo et al., 2025).

Selain pemilihan algoritma yang tepat, metode pembobotan kata juga sangat penting dalam meningkatkan akurasi model analisis sentimen. Salah satu metode pembobotan yang sering digunakan dalam penelitian sebelumnya adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini berfungsi untuk menilai pentingnya suatu kata dalam suatu dokumen dibandingkan dengan keseluruhan koleksi dokumen, sehingga membantu mengurangi pengaruh kata-kata yang sering muncul namun tidak memiliki makna sentimen yang signifikan (Ulfa et al., 2024).

Sejumlah penelitian terdahulu menunjukkan efektivitas kombinasi pra-pemrosesan dan algoritma klasifikasi pada ulasan aplikasi, misalnya pada domain *e-commerce*, e-*government, finance*, dan sebagainya (Tanggraeni et al, 2022; Muzaki et al, 2024; Larasati et

al, 2022). Sedangkan studi yang mengevaluasi sentimen ulasan aplikasi Deepsek telah menggunakan metode *Naive Bayes*, *Random Forest*, KNN, SVM, hingga LSTM dan IndoBERT (Prilindaputra et al, 2025; Mahendra et al., 2025; Riady, 2021). Akan tetapi, belum terdapat studi yang secara spesifik mengevaluasi sentimen ulasan aplikasi AI generatif pendatang baru seperti *DeepSeek* menggunakan Decision Tree, sekaligus mengaitkannya dengan kata dominan dan pola frasa (N-gram) yang menjelaskan sumber kepuasan maupun keluhan pengguna (Maulana et al, 2025; Fathoni et al, 2025; Baihaqi et al., 2025).

Berdasarkan observasi awal terhadap ulasan pengguna, terlihat adanya kecenderungan bahwa meskipun aplikasi DeepSeek memiliki kapabilitas kecerdasan tinggi, banyak pengguna yang mengeluhkan aspek teknis, khususnya terkait infrastruktur dan performa aplikasi. Kebaruan atau *novelty* pada penerapan Decision Tree berbasis fitur TF-IDF untuk klasifikasi sentimen ulasan aplikasi AI generatif pendatang baru (*DeepSeek*) dengan pendekatan *white-box* yang mudah diinterpretasikan, serta dilengkapi analisis kata dominan dan pola frasa (N-gram) untuk menjelaskan sumber kepuasan maupun keluhan pengguna secara lebih operasional, bukan sekadar memberikan label positif/negatif. Berdasarkan kondisi tersebut, penelitian ini bertujuan menerapkan Decision Tree berbasis fitur TF-IDF untuk mengklasifikasikan sentimen ulasan pengguna DeepSeek serta mengidentifikasi kata dan frasa dominan pada ulasan positif maupun negatif. Temuan penelitian diharapkan menjadi masukan berbasis data bagi pengembang untuk memprioritaskan perbaikan teknis dan pengembangan fitur sesuai kebutuhan pengguna.

Metode

Pengumpulan dan Persiapan data penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen komputasi untuk menganalisis sentimen pengguna. Data utama diperoleh melalui teknik *scraping* pada ulasan aplikasi *DeepSeek* di *Google Play Store*. Dari proses akuisisi data tersebut, dataset awal yang berhasil dikumpulkan kemudian divalidasi kelengkapannya.

Teks ulasan dibersihkan melalui tahapan pra-pemrosesan untuk mengurangi noise dan meningkatkan kualitas fitur. Tahapan data mencakup proses pembersihan data (*data cleaning*) untuk menghapus baris yang kosong (*missing values*) dan duplikasi, *case folding, cleaning*, *tokenization, normalization,* hingga *stopword removal*. Setelah itu, diperoleh total data bersih sebanyak 3.506 ulasan unik yang siap untuk dianalisis. Dataset ini terdiri dari kolom teks ulasan yang telah melalui tahap *stopword removal* (penghapusan kata umum yang tidak bermakna) dan kolom skor pengguna yang berkisar antara 1 hingga 5. Integritas data diverifikasi dengan memastikan tipe data kolom teks sebagai objek string dan kolom skor sebagai integer untuk menjamin kelancaran proses komputasi selanjutnya.

Automatic Labelling of Sentiment

Pelabelan data otomatis untuk keperluan klasifikasi dengan *machine learning* (*supervised learning*), setiap ulasan memerlukan label sentimen sebagai target prediksi. Penelitian ini menerapkan teknik pelabelan otomatis berdasarkan skor/rating pengguna. Mekanisme pelabelan membagi data ke dalam dua kelas sentimen (*biner*). Ulasan dengan skor/rating 1, 2, dan 3 dikategorikan ke dalam kelas Negatif, merepresentasikan ketidakpuasan pengguna. Sementara itu, ulasan dengan skor 4 dan 5 dikelompokkan ke dalam kelas Positif, merepresentasikan kepuasan pengguna. Label kategori teks ini kemudian dikonversi menjadi format numerik (0 dan 1) menggunakan teknik *Label Encoding* agar dapat diproses oleh algoritma klasifikasi.

Feature Extraction using TF-IDF

Ekstraksi Fitur *Teks* (TF-IDF) Mengingat algoritma Decision Tree tidak dapat memproses data teks mentah secara langsung, dilakukan proses transformasi data menjadi representasi vektor numerik menggunakan metode TF-IDF. Konfigurasi vektorisasi dalam penelitian ini diatur secara spesifik untuk meningkatkan kekayaan fitur. Parameter *ngram_range* diatur pada skala (1,2), yang berarti model tidak hanya mempelajari kata tunggal (*unigram*), tetapi juga frasa yang terdiri dari dua kata berurutan (*bigram*) untuk menangkap konteks kalimat yang lebih baik. Selain itu, untuk menjaga efisiensi komputasi dan mengurangi *noise*, jumlah fitur dibatasi hanya pada 5.000 kata dengan bobot tertinggi (*max_features=5000*). Proses ini mengubah kolom teks ulasan menjadi matriks fitur (X) dan label sentimen menjadi vektor target (y).

Data Splitting

Skenario pembagian data (*Data Splitting*) validasi model dilakukan dengan membagi dataset menjadi dua bagian terpisah, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian dilakukan dengan rasio 80:20, di mana 80% data digunakan untuk melatih model mempelajari pola sentimen, dan 20% sisanya digunakan untuk menguji performa model pada data baru. Proses pembagian ini menerapkan teknik *Stratified Sampling* (*stratify=y*), sebuah metode statistik yang memastikan bahwa proporsi rasio antara sentimen positif dan negatif pada data latih dan data uji tetap seimbang dan sama dengan proporsi pada dataset asli. Pengacakan data dikunci menggunakan parameter *random_state=42* untuk memastikan hasil penelitian yang konsisten dan dapat direplikasi (*reproducible*) oleh peneliti lain.

Evaluation Metrics

Konfigurasi Model dan Evaluasi Kinerja Klasifikasi sentimen dilakukan menggunakan algoritma Decision Tree. Algoritma ini dipilih karena karakteristiknya yang transparan (*white-box model*), memungkinkan visualisasi struktur keputusan yang jelas. Model dilatih (*fitting*) menggunakan data latih untuk membangun pohon keputusan yang memetakan fitur kata ke label sentimen. Evaluasi kinerja model diukur secara komprehensif menggunakan metrik *Confusion Matrix* yang menghasilkan nilai Akurasi, Presisi, *Recall*, dan *F1-Score*. Selain itu, analisis hasil juga diperdalam dengan membandingkan label prediksi terhadap label aktual, serta visualisasi struktur pohon keputusan (*Tree Plot*) dengan kedalaman terbatas (*max_depth*) untuk menginterpretasikan aturan keputusan (*decision rules*) terpenting yang dipelajari oleh model.

Hasil dan Pembahasan

Tahapan ini menguraikan tentang hasil eksperimen yang mencakup analisis statistik deskriptif terhadap data, evaluasi kinerja model klasifikasi Decision Tree, serta analisis mendalam mengenai karakteristik sentimen pengguna melalui pendekatan visualisasi kata dan analisis kontekstual (*N-Gram*).

Hasil Pelabelan Sentimen Otomatis (Automatic Labeling)

Tahap awal dalam analisis hasil adalah transformasi atribut skor pengguna menjadi label kelas sentimen. Mengingat data ulasan mentah yang dikumpulkan tidak memiliki label sentimen eksplisit, penelitian ini menerapkan teknik pelabelan otomatis (*rule-based labeling*) dengan menetapkan batasan ambang (*threshold*) pada skor (skala 1-5).

Berdasarkan fungsi algoritma yang diterapkan, skema pengelompokan dibagi menjadi dua kelas (klasifikasi biner). Ulasan dengan skor 1, 2, dan 3 dikategorikan sebagai Negatif, sedangkan ulasan dengan skor 4 dan 5 dikategorikan sebagai Positif. Keputusan untuk memasukkan skor 3 (netral) ke dalam kelas negatif bertujuan untuk mempertegas batasan kepuasan, di mana nilai tengah dianggap belum memenuhi standar kepuasan pengguna yang optimal dalam konteks pengembangan aplikasi. Tabel berikut menampilkan sampel data setelah proses pelabelan dilakukan:

Tabel 1. *Sampel Hasil Pelabelan Data Berdasarkan Skor/rating*

Stopword Removal	Score	Sentiment
Tolong buatkan memory aplikasi	3	Negatif
Server down	1	Negatif
Top markotop	5	Positif
Aplikasi AI tahan banting diperbaiki laptop *virtual survive* tes humor berlapis fitur terkecoh belajar developernya fitur deteksi bercanda cepat fitur napas dalamdalam virtual dikerjai user perbaiki laptop chat tahan uji humor berlapis ala indonesia developernya kayaknya paham suka bercanda butuh solusi serius minus kadang terkecoh lucu overall ai humoris	5	Positif
Deepseek tolong batas limit chat menghambat ku menuangkan imajinasi kreativitas dikarnakan limit pendek	4	Positif

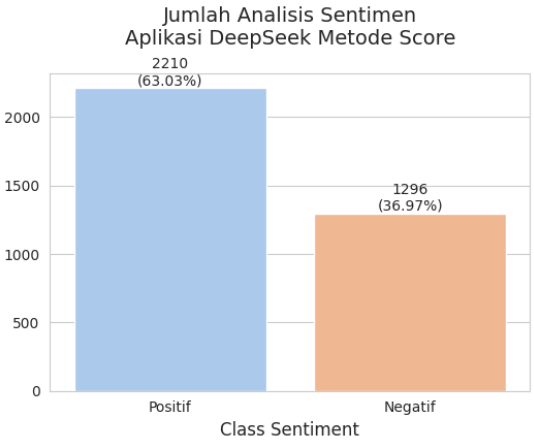
Analisis terhadap sampel data di atas menunjukkan keberhasilan proses pemetaan label. Ulasan dengan sentimen ketidakpuasan yang ekstrem, seperti pada data ke-2 ("server down", skor 1), secara tepat terlabeli sebagai Negatif. Menariknya, pada data ke-1 yang memiliki skor 3 ("tolong buatkan memory aplikasi"), sistem melabelinya sebagai Negatif. Hal ini konsisten dengan desain penelitian untuk menangkap saran perbaikan atau ketidakpuasan moderat sebagai sentimen negatif.

Sebaliknya, ulasan yang mengekspresikan kepuasan seperti pada data ke-3 ("top markotop", skor 5) dikategorikan sebagai Positif. Namun, terdapat tantangan tersendiri pada data ke-5, di mana pengguna memberikan skor 4 (Positif) namun isi teksnya mengandung keluhan terkait "batas limit chat". Hal ini menjadi catatan penting bahwa pelabelan berbasis skor memiliki sedikit bias inheren karena pengguna terkadang memberikan skor tinggi meskipun menyampaikan kritik.

Distribusi Data dan Keseimbangan Kelas

Sebelum melangkah pada proses pemodelan, dilakukan analisis statistik deskriptif terhadap 3.506 data ulasan bersih hasil pra-pemrosesan. Analisis distribusi kelas dilakukan untuk melihat proporsi jumlah data antara sentimen positif dan negatif. Pemahaman terhadap distribusi ini krusial karena ketidakseimbangan data (*imbalanced data*) seringkali mempengaruhi kinerja algoritma klasifikasi, di mana model cenderung bias memprediksi kelas mayoritas dan mengabaikan kelas minoritas (Yulvida et al., 2025). Visualisasi distribusi jumlah ulasan per kelas sentimen disajikan pada Gambar 1.

Berdasarkan visualisasi grafik batang pada gambar 1, terlihat bahwa dataset memiliki sebaran yang mewakili dua spektrum opini pengguna. Kelas sentimen Positif mendominasi dataset dengan jumlah 2.210 ulasan (63,03%), sementara kelas sentimen Negatif berjumlah 1.296 ulasan (36,97%). Dominasi sentimen positif memberikan indikasi awal bahwa mayoritas pengguna merasa puas dengan layanan yang diberikan. Namun, proporsi sentimen negatif yang mencapai hampir 37% menandakan adanya masalah substansial yang dialami oleh lebih dari sepertiga pengguna.

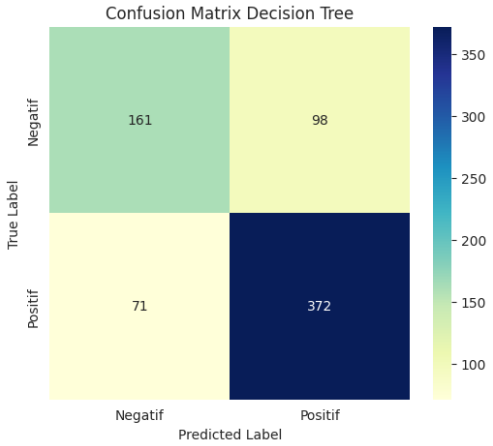


Gambar 1. *Jumlah Analisis Sentimen Aplikasi DeepSeek Metode Skor*

Guna menjaga validitas evaluasi, data ini kemudian dibagi (*split*) menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20 menggunakan teknik Stratified Sampling. Teknik ini dipilih untuk memastikan bahwa proporsi keseimbangan kelas pada data uji tetap terjaga konsistensinya sama dengan proporsi pada data aslinya, sehingga evaluasi model menjadi lebih adil dan representatif. Berdasarkan skenario pembagian tersebut, sebanyak 2.804 data dialokasikan untuk melatih model guna membentuk pola aturan keputusan, sedangkan 702 data disimpan secara terpisah sebagai data uji untuk evaluasi akhir.

Evaluasi Kinerja Model Klasifikasi (Decision Tree)

Setelah model *Decision Tree* dilatih, langkah selanjutnya adalah menguji kemampuannya dalam memprediksi sentimen pada 702 data uji. Evaluasi pertama dilakukan menggunakan *Confusion Matrix* untuk memetakan detail prediksi benar dan salah.



Gambar 2. *Hasil Confusion Matrix Decision Tree*

Gambar di atas menunjukkan peta persebaran hasil prediksi model. Dari total 259 data aktual sentimen negatif, model berhasil memprediksi dengan benar sebanyak 161 data (*True Negative*), namun terdapat kesalahan prediksi (*miss-classification*) sebanyak 98 data yang dianggap positif (*False Positive*). Sebaliknya, pada kelas positif yang berjumlah 443 data, model mampu memprediksi dengan benar sebanyak 372 data (True Positive), dengan tingkat kesalahan prediksi menjadi negatif sebanyak 71 data (*False Negative*). Untuk analisis kuantitatif yang lebih rinci, Tabel 2 menyajikan rangkuman kinerja model:

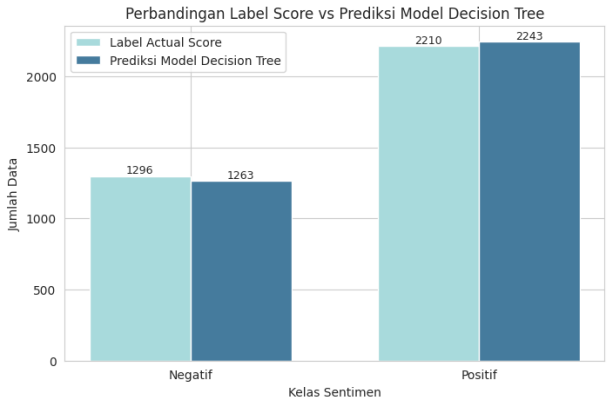
Tabel 2. *Laporan Klasifikasi (Classification Report) Model Decision Tree*

	Precision	Recall	F1-Score	Support
Negatif	0.6940	0.6216	0.6558	259
Positif	0.7915	0.8397	0.8149	443
Accuracy			0.7593	702
Macro Avg	0.7427	0.7307	0.7354	702
Weighted Avg	0.7555	0.7593	0.7562	702

Berdasarkan tabel 2, model menghasilkan tingkat akurasi sebesar 75,93%. Nilai ini menunjukkan kinerja yang moderat untuk klasifikasi teks ulasan yang memiliki variasi bahasa tinggi. Namun, terdapat disparitas pada nilai *F1-Score*. Kelas Positif memiliki *F1-Score* sebesar 81,49%, jauh lebih tinggi dibandingkan kelas Negatif yang hanya mencapai 65,58%. Hal ini mengonfirmasi bahwa model memiliki tingkat kepercayaan yang lebih tinggi saat mengklasifikasikan ulasan positif, namun masih kesulitan menangkap nuansa bahasa pada ulasan negatif yang mungkin lebih implisit atau sarkastik (Wafda et al., 2025).

Analisis Komparasi Aktual vs Prediksi

Membedah lebih lanjut bias yang terdeteksi pada evaluasi kinerja, dilakukan analisis komparatif antara distribusi jumlah label data Aktual dengan data hasil Prediksi model pada data uji. Dalam konteks ini, data "Aktual" adalah label hasil *score-based*, sedangkan "Prediksi" adalah output dari algoritma.



Gambar 3. *Perbandingan Label Score dan Prediksi Model*

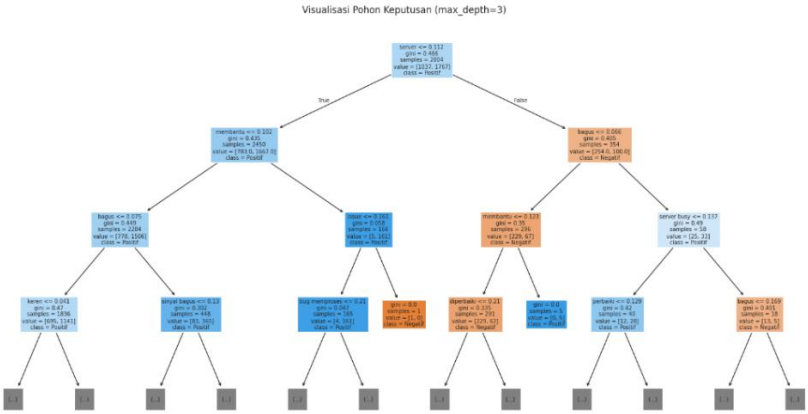
Visualisasi perbandingan antara Label *Actual Score* dan Prediksi Model Decision Tree di atas merepresentasikan hasil evaluasi model saat diterapkan kembali pada populasi data secara keseluruhan yang berjumlah 3.506 ulasan. Grafik tersebut menunjukkan tingkat konsistensi yang sangat tinggi antara pengelompokan manual berbasis skor pengguna dengan hasil klasifikasi otomatis yang dilakukan oleh algoritma, di mana model mampu memetakan distribusi sentimen yang hampir identik dengan kondisi aslinya.

Berdasarkan data yang tersaji, pada kelas sentimen negatif terlihat sedikit penurunan dari 1.296 data aktual menjadi 1.263 data prediksi, sedangkan sebaliknya pada kelas sentimen positif terjadi kenaikan proporsional dari 2.210 data aktual menjadi 2.243 data prediksi. Fenomena pergeseran minor sebanyak 33 data dari kategori negatif ke positif ini mengindikasikan adanya sejumlah kecil ulasan yang memiliki karakteristik ambiguitas, seperti pengguna yang memberikan rating rendah namun menggunakan struktur kalimat atau diksi yang secara tekstual terbaca positif oleh sistem, misalnya memuji kecerdasan AI namun memberikan bintang rendah karena masalah koneksi. Secara keseluruhan, kedekatan angka yang signifikan antara label aktual dan prediksi ini memvalidasi bahwa model *Decision Tree* yang dibangun memiliki performa yang stabil (*robust*) dan mampu

melakukan generalisasi pola dengan sangat baik tanpa mengalami kegagalan struktural saat dihadapkan pada seluruh variasi data ulasan.

Interpretasi Pohon Keputusan (Decision Tree Visualization)

Salah satu keunggulan algoritma *Decision Tree* adalah sifatnya yang *White Box*, yang memungkinkan peneliti melihat aturan logika di balik keputusan model.

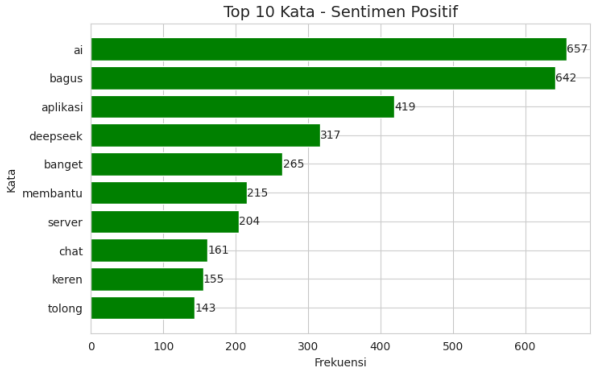


Gambar 4. *Visualisasi Pohon Keputusan*

Visualisasi pohon keputusan mengungkap fitur kata kunci penentu sentimen. Simpul akar (*root node*) membagi data berdasarkan fitur kata "server". Jika kata "server" muncul (nilai TF-IDF tinggi), pohon cenderung mengarahkan keputusan ke cabang kanan yang didominasi oleh kelas Negatif. Sebaliknya, cabang kiri didominasi oleh kata-kata seperti "membantu", "bagus", dan "keren", yang mengarah pada kelas Positif. Struktur ini secara valid merepresentasikan pola pikir manusia: adanya kata teknis (*server, error*) berkonotasi negatif, sedangkan kata sifat pujian berkonotasi positif.

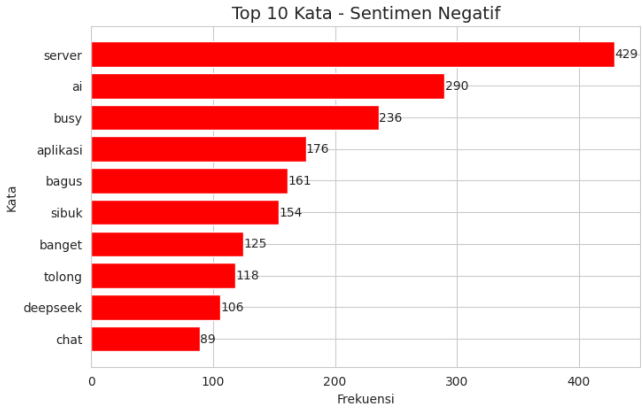
Analisis Karakteristik Kata Dominan

Berdasarkan memahami apa yang sebenarnya dibicarakan pengguna, dilakukan analisis visual dengan statistik frekuensi kata. Analisis ini bertujuan untuk menggali kata-kata dominan yang sering digunakan oleh pengguna, baik dalam konteks positif maupun negatif. Dengan menganalisis frekuensi kemunculan kata-kata tersebut, kita dapat memperoleh wawasan lebih dalam mengenai apa yang menjadi fokus dan perhatian utama pengguna, serta bagaimana mereka merespons aplikasi yang digunakan. Selain itu, analisis ini juga membantu untuk mengidentifikasi pola sentimen yang muncul dalam percakapan pengguna, yang dapat memberikan petunjuk tentang elemen-elemen aplikasi yang dianggap penting atau bermasalah (Maldini et al, 2025).



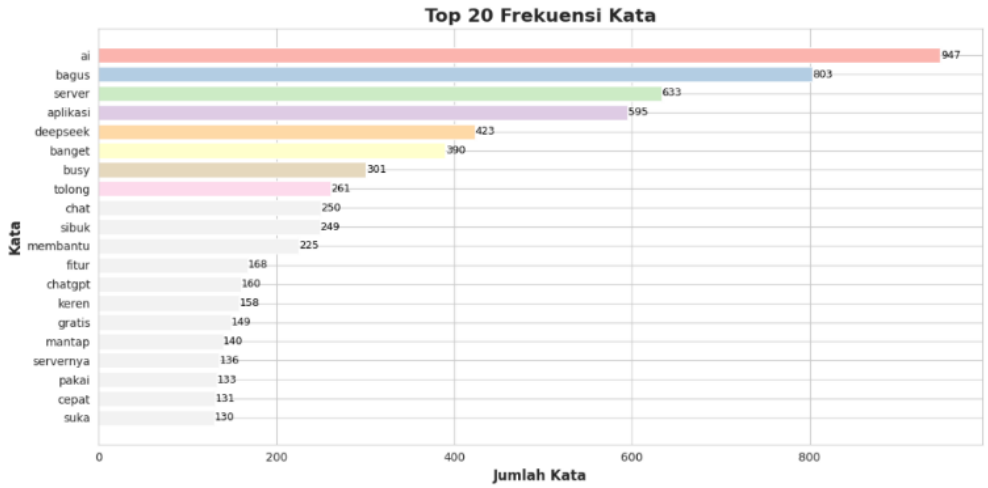
Gambar 5. *Top 10 Kata Sentimen Positif*

Grafik pada gambar 5 menunjukkan hasil analisis kata dominan berdasarkan sentimen positif dan negatif, serta frekuensi kemunculan kata secara keseluruhan. Grafik tersebut menampilkan sepuluh kata dengan frekuensi tertinggi dalam kategori sentimen positif. Sentimen Positif pada kelas positif, kata "ai" (657 kemunculan) dan "bagus" (642 kemunculan) mendominasi secara signifikan. Hal ini mengindikasikan bahwa kepuasan utama pengguna bersumber dari kecerdasan buatan itu sendiri. Kata kerja "membantu" (215) juga muncul di posisi tinggi, yang memvalidasi bahwa aplikasi ini dinilai fungsional dan bermanfaat dalam menyelesaikan tugas pengguna. Selanjutnya berikut grafik yang menampilkan frekuensi kategori sentimen negati.



Gambar 6. *Top 10 Kata Sentimen Negatif*

Grafik ini menampilkan sepuluh kata dengan frekuensi tertinggi dalam kategori sentimen negatif. Kata "Server" (429 kemunculan) menjadi kata yang paling banyak muncul, diikuti oleh "busy" (236) dan "sibuk" (154). Hal ini mengindikasikan bahwa masalah teknis terkait server dan keterbatasan waktu menjadi isu utama bagi pengguna dalam menggunakan aplikasi ini.



Gambar 7. *Top 20 Frekuensi Kata*

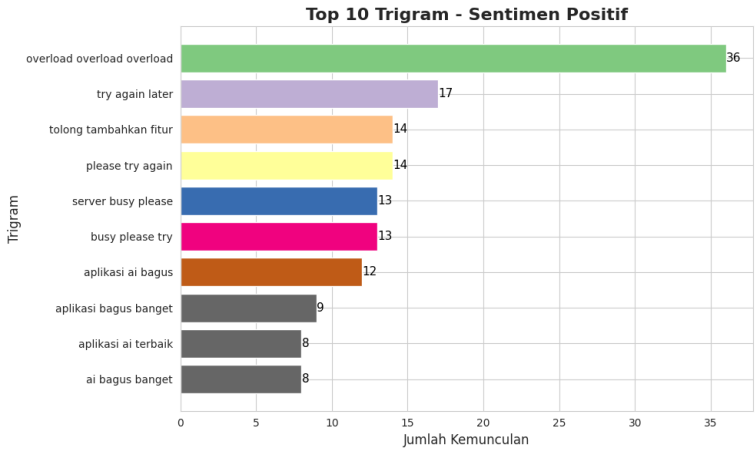
Grafik ini memberikan gambaran umum tentang dua puluh kata dengan frekuensi kemunculan tertinggi dalam seluruh analisis. Kata-kata seperti "ai", "bagus", dan "server" masih mendominasi, menunjukkan tema umum yang berfokus pada teknologi kecerdasan buatan dan kualitas aplikasi. Selain itu, kata-kata seperti "chat", "keres", dan "gratis" mencerminkan fitur interaktif dan elemen lain dari aplikasi yang juga menjadi perhatian pengguna.

ISSN 2808-7038

Analisis Kontekstual Berbasis N-Gram (Trigram)

Meskipun analisis frekuensi kata telah berhasil mengidentifikasi kata-kata dominan, metode tersebut memiliki keterbatasan karena memecah kalimat menjadi unit terkecil (kata tunggal/unigram) sehingga menghilangkan konteks semantiknya. Untuk mengatasi hal ini, penelitian melanjutkan eksplorasi menggunakan analisis N-Gram dengan skema Trigram (rangkaian tiga kata yang muncul berurutan). Pendekatan ini bertujuan untuk merekonstruksi struktur frasa agar dapat memahami *bagaimana* kata-kata tersebut digunakan dalam kalimat ulasan yang sebenarnya. Visual grafik batang (*Bar Chart*) berikut menampilkan 10 Trigram dengan frekuensi kemunculan tertinggi pada masing-masing kelas sentimen.

Analisis Trigram Sentimen Positif



Gambar 8. *Top 10 Trigram Sentimen Positif*

Visualisasi grafik batang horizontal pada Gambar *8 - Sentimen Positif* menyajikan sepuluh kombinasi tiga kata (*trigram*) yang paling sering muncul dalam ulasan berlabel positif. Analisis terhadap pola frasa ini menyingkap fenomena psikologis pengguna yang unik, yang tidak dapat terdeteksi hanya dengan melihat frekuensi kata tunggal.

Berdasarkan grafik tersebut, teridentifikasi tiga klaster pola utama. Pertama, Paradoks (Anomali Pesan *Error*), hal yang paling mencolok dari grafik ini adalah dominasi frasa yang bernada negatif atau teknis di urutan teratas, padahal label datanya adalah "Positif". Dominasi "Overload": Trigram "overload overload overload" menempati posisi pertama dengan frekuensi tertinggi (36 kemunculan). Pesan Error Server: Diikuti oleh frasa pesan kesalahan sistem seperti "try again later" (17), "server busy please" (13), dan "busy please try" (13). Interpretasi: Ini mengindikasikan bahwa pengguna yang memberikan rating tinggi (Positif) adalah pengguna yang sangat antusias ingin menggunakan aplikasi, namun merasa terganggu oleh masalah server.

Pengguna menulis ulasan seperti: *"servernya down overload booming lonjakan user semoga upgrade servernya selebihnya kemampuan chatbotnya unggul kompetitornya good job semoga berkembang"* ke dalam ulasan bintang 4 mereka sebagai laporan *bug* kepada pengembang. Kedua, umpan Balik Konstruktif (*Constructive Feedback*) pada ulasan pengguna trigram "tolong tambahkan fitur" muncul sebanyak 14 kali. Ini adalah ciri khas ulasan positif yang berkualitas. Pengguna tidak hanya memuji, tetapi memiliki ekspektasi tinggi dan keinginan agar aplikasi berkembang lebih jauh. Ini menandakan tingkat keterikatan (*engagement*) pengguna yang tinggi terhadap produk.

Ketiga, Apresiasi Inti Produk (*True Positive*) selanjutnya, pada urutan bawah visualisasi trigram (peringkat 6 hingga 10), pola distribusi mulai menunjukkan konsistensi yang murni dengan label sentimen positif, di mana frasa pujian terhadap kinerja produk mendominasi tanpa adanya ambiguitas keluhan teknis. Indikator kepuasan pengguna ini terekam jelas melalui kemunculan frasa *"aplikasi ai bagus"* sebanyak 12 kali, diikuti oleh *"aplikasi bagus banget"* dengan 9 kemunculan, serta frasa *"aplikasi ai terbaik"* dan *"ai bagus banget"* yang masing-masing muncul sebanyak 8 kali. Kehadiran kombinasi kata-kata apresiatif ini secara empiris memvalidasi bahwa di balik kendala infrastruktur yang ada, kapabilitas fundamental kecerdasan buatan (AI) yang ditawarkan aplikasi ini sebenarnya telah diakui kualitasnya dan diterima dengan sangat baik oleh pengguna.

Analisis Trigram Sentimen Negatif

Visualisasi grafik batang horizontal ini menampilkan 10 kombinasi tiga kata (*trigram*) yang paling sering muncul dalam ulasan berlabel negatif. Pola distribusi trigram ini memberikan wawasan krusial mengenai akar permasalahan utama yang memicu ketidakpuasan pengguna: Pertama, Dominasi Pesan Galat Sistem (*System Error Messages*), Hal yang paling mencolok dari visualisasi grafik trigram negatif ini adalah adanya homogenitas data yang kuat pada peringkat teratas, di mana empat trigram utama yang muncul bukanlah kalimat yang disusun secara natural oleh pengguna, melainkan merupakan kutipan langsung dari pesan *error* sistem aplikasi.

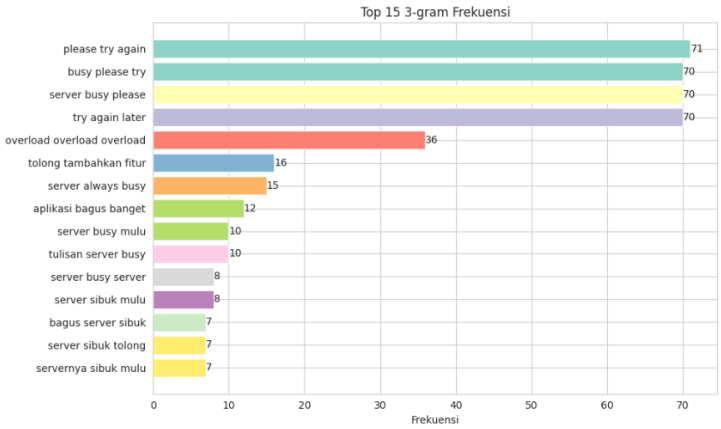
Fakta ini didukung oleh temuan pola statistik yang identik pada frasa "server busy please", "busy please try", dan "please try again" yang masing-masing memiliki jumlah kemunculan sama persis sebanyak 57 kali, serta diikuti oleh variasi pesan "try again later" dengan 53 kemunculan. Kedua, Penekanan pada durasi masalah (*Persistence*), Pada peringkat menengah (peringkat 5), muncul frasa "server always busy" sebanyak 11 kali. Penggunaan kata *"always"* (selalu) menjadi indikator penting bahwa gangguan ini tidak dirasakan sebagai kejadian sesekali, melainkan hambatan permanen yang dirasakan pengguna setiap kali mencoba mengakses aplikasi.

Ketiga, Ekspresi Kekesalan Natural (Slang "Mulu"), Beranjak ke bagian bawah visualisasi, terlihat pergeseran pola linguistik yang signifikan dari pesan sistem baku menjadi ekspresi keluhan natural pengguna yang menggunakan bahasa campuran. Fenomena ini ditandai oleh kemunculan variasi frasa seperti "server busy mulu" dengan frekuensi 9 kali, diikuti oleh "server sibuk mulu" sebanyak 7 kali, serta "servernya sibuk mulu" sebanyak 6 kali. Penggunaan kata slang "mulu" yang bermakna repetisi atau terus-menerus ini merefleksikan tingkat frustrasi yang mendalam dari pengguna, di mana kehadiran frasa-frasa tersebut secara implisit memvalidasi bahwa sebenarnya terdapat antusiasme dan keinginan kuat untuk memanfaatkan aplikasi, namun kesabaran pengguna akhirnya habis terkikis oleh hambatan teknis server yang bersifat persisten dan belum terselesaikan.

Analisis Global Top 15 Trigram Frekuensi

Visualisasi grafik batang horizontal (Gambar 9) yang menampilkan lima belas kombinasi tiga kata (*trigram*) dengan frekuensi tertinggi ini secara gamblang menyingkap realitas bahwa percakapan pengguna didominasi secara mutlak oleh kendala teknis, yang secara efektif menenggelamkan diskusi mengenai kualitas produk itu sendiri. Hal ini terbukti dari homogenitas data pada empat peringkat teratas, yakni "*please try again*", "*busy please try*", "*server busy please*", dan "*try again later*" yang memiliki frekuensi sangat

masif dan seragam di angka 70 hingga 71 kemunculan, mengindikasikan bahwa mayoritas input teks yang diterima sistem bukanlah opini organik, melainkan salinan pesan gangguan (*system error*) yang dialami pengguna secara massal.



Gambar 9. *Top 15 Trigram Frekuensi*

Tingkat frustrasi pengguna terekam jelas pada peringkat kelima melalui repetisi kata ekstrem "*overload overload overload*" sebanyak 36 kali, namun menariknya, di tengah badai keluhan tersebut, trigram "tolong tambahkan fitur" berhasil menempati peringkat keenam dengan 16 kemunculan, yang menandakan bahwa pengguna masih menaruh harapan dan memiliki loyalitas konstruktif terhadap pengembangan aplikasi. Selebihnya, dominasi isu infrastruktur kembali terlihat pada peringkat bawah yang dipenuhi variasi keluhan bahasa santai seperti "server busy mulu" dan "server sibuk mulu", menyisakan hanya satu ruang kecil untuk apresiasi murni pada frasa "aplikasi bagus banget" di peringkat ketujuh dengan 12 kemunculan, yang menegaskan kesimpulan akhir bahwa perbaikan stabilitas server adalah prioritas absolut di atas segalanya.

Secara keseluruhan, analisis trigram berhasil melengkapi analisis frekuensi kata dengan menghadirkan konteks frasa, sekaligus memperkuat bukti adanya fenomena “sentimen campuran” (rating tinggi namun berisi kritik), sehingga interpretasi isu utama menjadi lebih presisi. Dengan demikian, analisis sentimen ulasan aplikasi memerlukan pra-pemrosesan dan eksplorasi fitur teks untuk menangkap pola opini pengguna secara lebih akurat (Handayani et al., 2024), serta bahwa keluhan teknis sering muncul dalam bentuk frasa berulang yang menjadi indikator kuat ketidakpuasan. Selain itu, kemunculan kritik pada ulasan dengan rating tinggi sejalan dengan catatan bahwa pelabelan berbasis rating bersifat *weak supervision* dan dapat menghasilkan bias atau ketidaksesuaian antara label dan isi teks, sehingga analisis berbasis N-gram penting untuk memvalidasi konteks dan menjelaskan sumber ketidakpuasan secara lebih jelas (Wardana et al., 2025).

Kesimpulan

Berdasarkan analisis sentimen terhadap 3.506 ulasan DeepSeek di Google Play Store, distribusi label menunjukkan sentimen positif 2.210 (63,03%) dan negatif 1.296 (36,97%), sehingga porsi ketidakpuasan pengguna masih tergolong substansial. Hasil analisis kata dan frasa dominan mengonfirmasi bahwa sentimen negatif terutama dipicu oleh isu instabilitas infrastruktur, ditandai dominasi kata “server”, dan “sibuk”, serta kemunculan frasa galat seperti “server busy please” dan “please try again” yang sering disalin pengguna sebagai bentuk pelaporan gangguan layanan. Sebaliknya, ulasan positif didominasi apresiasi terhadap kualitas inti AI melalui kata “ai” dan “bagus” serta indikator

kebermanfaatan "membantu", sekaligus menunjukkan adanya "sentimen campuran" ketika pengguna tetap memberi rating tinggi meskipun menyisipkan keluhan teknis. Dengan fitur TF-IDF berbasis n-gram dan klasifikasi Decision Tree, model mencapai akurasi 75,93% pada data uji 702 ulasan dengan kinerja per kelas: Negatif (*Precision* 0,6940; *Recall* 0,6216; *F1-score* 0,6558) dan Positif (*Precision* 0,7915; *Recall* 0,8397; *F1-score* 0,8149). Hasil ini menegaskan bahwa klasifikasi sentimen ulasan dan identifikasi kata/frasa dominan pada ulasan positif maupun negatif telah tercapai secara terukur.

Implikasi praktisnya, hasil menunjukkan DeepSeek memiliki penerimaan yang kuat pada kualitas AI, namun peningkatan stabilitas server/skalabilitas layanan menjadi prioritas utama untuk menurunkan sentimen negatif yang didominasi keluhan teknis. Meski demikian, penelitian ini memiliki keterbatasan karena pelabelan sentimen berbasis rating bersifat *weak supervision* sehingga berpotensi menghasilkan "sentimen campuran" (rating tinggi dengan isi keluhan), serta hanya menggunakan skema dua kelas dan satu algoritma utama. Penelitian selanjutnya disarankan menambah anotasi manual sampel, menguji skema 3 kelas atau membuat kelas khusus "laporan gangguan server", menerapkan *k-fold cross validation*, serta membandingkan dengan model teks agar hasil lebih robust dan dapat digeneralisasi.

Acknowledgment

-

Daftar Pustaka

Baihaqi, M. F., Magdalena, L., & Fahrudin, R. (2025). Analisis Sentimen Aplikasi Deepseek Menggunakan Metode Naive Bayes dan Support Vector Machine. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(3), 4051-4062. https://doi.org/10.31004/riggs.v4i3.2511

Darulanda, H., Padangjati, A. N., & Al-Marami, Z. (2024). Narasi Populer tentang AI dalam Pendidikan: Studi Literatur Wacana Daring. *Jurnal Literasi Digital*, *3*(2), 99–109. https://doi.org/10.54065/jld.3.2.2023.601

Fathoni, F., Harahap, D. K., Pardede, E. T., Nachwa, S., Maulizidan, M. R., & Ibrahim, A. (2025). Komparasi Kinerja KNN dan Naïve Bayes Untuk Klasifikasi Sentimen Pengguna Aplikasi Deepseek AI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(4), 6021-6028. https://doi.org/10.36040/jati.v9i4.13887

Handayani, M. R., Yuniarti, W. D., & Umam, K. (2024). Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM: Indonesia. *SINTECH (Science and Information Technology) Journal*, *7*(2), 80-91. https://doi.org/10.31598/sintechjournal.v7i2.1581

Husain, N. P., Sukirman, S., & SAJIAH, S. (2024). Analisis Sentimen Ulasan Pengguna Tiktok pada Google Play Store Berbasis TF-IDF dan Support Vector Machine. *Journal of System and Computer Engineering*, *5*(1), 91-102. https://doi.org/10.61628/jsce.v5i1.1105

Karim, H. F., & Wibowo, A. P. (2025). Kinerja Metode Fine-Tuning IndoBERT untuk Klasifikasi Emosi Multi-Kelas pada Teks Informal Bahasa Indonesia. *Bulletin of Computer Science Research*, *6*(1), 63-74. https://doi.org/10.47065/bulletincsr.v6i1.850

Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(9), 4305-4313. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562

Lisdiyanto, A., Nugroho, R. A., Andhyka, A., Wibowo, A., Winarti, W., & Budiman, B. (2025). Pengembangan Aplikasi Bengkel Las di Kediri dengan Metode Extreme Programming. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *7*(1), 63-69. https://doi.org/10.47233/jteksis.v7i1.1740

Mahendra, L. S., Chrisnanto, Y. H., & Yuniarti, R. (2025). Analisis Sentimen Terhadap Ulasan Aplikasi Deepseek AI Menggunakan Model Bidirectional LSTM dan IndoBert. *Jurnal Algoritma*, *22*(2), 1739-1750. https://doi.org/10.33364/algoritma/v.22-2.2950

Maldini, M. A. A., & Andryana, S. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Perbankan Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(3), 4098-4105. https://doi.org/10.36040/jati.v9i3.13522

Maulana, M. R., Imran, A. Y., & Fathoni, F. (2025). Analisis Sentimen Ulasan Deepseek AI Di Google Playstore Menggunakan Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(4), 6473-6478. https://doi.org/10.36040/jati.v9i4.14070

Muzaki, A., Febriana, V., & Cholifah, W. N. (2024). Analisis Sentimen Pada Ulasan Produk di E-Commerce dengan Metode Naive Bayes. *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, *5*(4), 758–765. https://doi.org/10.30998/jrami.v5i4.9647

Prasetyo, T., Waskita, A. A., & Taryo, T. (2025). Analisis Sentimen Pengguna Mobil Listrik di Media Sosial Twitter Menggunakan Metode Klasifikasi Naïve Bayes, K-Nearest Neighbor (KNN), dan Decision Tree. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, *8*(2), 108-117. https://doi.org/10.47970/siskom-kb.v8i1.783

Prilindaputra, B., Putri, D. R. R., & Ulinnuha, N. (2025). Implementasi Support Vector Machine untuk Analisis Sentimen Aplikasi Deepseek. *INTEGER: Journal of Information Technology*, *10*(1). https://doi.org/10.31284/j.integer.2024.v10i1.7541

Riady, A. (2021). Pendidikan Berkualitas di Era Digital: (Fokus: Aplikasi Sebagai Media Pembelajaran). *Jurnal Literasi Digital*, *1*(2), 70–80. https://doi.org/10.54065/jld.1.2.2021.15

Tanggraeni, A. I., & Sitokdana, M. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, *9*(2), 785-795. https://doi.org/10.35957/jatisi.v9i2.1835

Ulfa, M., Kusumodestoni, R. H., & Sucipto, A. (2024). Analisis Sentimen Review Aplikasi Identitas Kependudukan Digital di Google Play Store Menggunakan KNN. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, *6*(4), 1155–1165. https://doi.org/10.51401/jinteks.v6i4.4963

Umbara, F. W. (2021). User Generated Content di Media Sosial Sebagai Strategi Promosi Bisnis. *Jurnal Manajemen Strategi dan Aplikasi Bisnis*, *4*(2), 572-581. https://doi.org/10.36407/jmsab.v4i2.366

Wafda, A., Fudholi, D. H., & Nugraha, J. (2025). Aspect-Based Sentiment Analysis On Twitter Tweets About The Merdeka Curriculum Using Indobert. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, *10*(3), 586-599. https://doi.org/10.33480/jitk.v10i3.5692

Wardana, R. K., Cahyono, N., Saputro, U. A., Negoro, A. H., & Aini, N. (2025). Implementasi Algoritma SVM dan Optimasi Menggunakan Komparasi N-Gram dan Glove Pada Sentimen Pengguna Aplikasi Access By KAI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(2), 3216-3223. https://doi.org/10.36040/jati.v9i2.13284

Wati, F. F., & Widodo, A. E. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Deepseek Menggunakan Algoritma Random Forest Dan Naive Bayes. *CONTEN: Computer and Network Technology,* 5(1), 8-15 https://doi.org/10.31294/hqpha267.

Wily, W. A., Anggai, S., & Tukiyat, T. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial X di Play Store Menggunakan Algoritma Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU): Studi Kasus pada Ulasan Pengguna di Google Play Store. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, *9*(1), 63-72. https://doi.org/10.47970/siskom-kb.v9i1.875

Yulvida, D., Quinevera, S., Mardianto, R., & Joses, S. (2025). Klasifikasi Pemohon Pinjaman dengan Hyperparameter Tuning dan Teknik Penyeimbangan Data. *Journal of Applied Computer Science and Technology*, *6*(2), 92-100. https://doi.org/10.52158/krjtrh05