Analisis Sentimen pada Ulasan Pengguna Aplikasi Gojek untuk Mengidentifikasi Kepuasan Pelanggan dan Faktor Keluhan

**Muhammad Haikal Firdaus**  
*Information Systems Study Program*  
*Faculty of Engineering and Computer Science*  
*Universitas Komputer IndonesiaBandung, Indonesia*  
[haikal.10522035@mahasiswa.unikom.ac.id](mailto:haikal.10522035@mahasiswa.unikom.ac.id)

**Rahman Aditia**  
*Information Systems Study Program*  
*Faculty of Engineering and Computer Science*  
*Universitas Komputer IndonesiaBandung, Indonesia*  
[rahman.10522001@mahasiswa.unikom.ac.id](mailto:rahman.10522001@mahasiswa.unikom.ac.id)**Aditya Juliana**  
*Information Systems Study Program*  
*Faculty of Engineering and Computer Science*  
*Universitas Komputer IndonesiaBandung, Indonesia*  
[aditya.10522029@mahasiswa.unikom.ac.id](mailto:aditya.10522029@mahasiswa.unikom.ac.id)

**Andhika Wirayuda**  
*Information Systems Study Program*  
*Faculty of Engineering and Computer Science*  
*Universitas Komputer IndonesiaBandung, Indonesia*  
[andhika.10522037@mahasiswa.unikom.ac.id](mailto:andhika.10522037@mahasiswa.unikom.ac.id)

*Abstrak*—*Persaingan di industri layanan on-demand mengharuskan perusahaan seperti Gojek untuk terus meningkatkan kualitas layanan berdasarkan umpan balik pengguna. Ulasan di Google Play Store adalah sumber data yang melimpah untuk tujuan ini, tetapi analisis manual tidak efisien karena volume data yang sangat banyak. Penelitian ini mengusulkan penerapan data mining untuk menganalisis sentimen ulasan pengguna aplikasi Gojek secara otomatis. Penelitian ini menggunakan metode klasifikasi Naive Bayes yang diketahui efektif untuk tugas klasifikasi teks. Proses penelitian mencakup beberapa tahap utama: pengumpulan data, pra-pemrosesan data untuk pelabelan sentimen, pra-pemrosesan teks (tokenisasi, pelipatan huruf besar, pemfilteran kata henti), dan pembobotan fitur menggunakan TF-IDF. Kinerja model dievaluasi menggunakan skema validasi silang 10 kali lipat dengan metrik akurasi, presisi, dan penarikan. Dengan menganalisis ribuan ulasan, penelitian ini bertujuan untuk mengidentifikasi secara akurat faktor-faktor kunci yang mendorong kepuasan dan keluhan pelanggan, yang hasilnya dapat memberikan wawasan strategis dan dapat ditindaklanjuti untuk pengembangan produk dan layanan Gojek di masa mendatang*

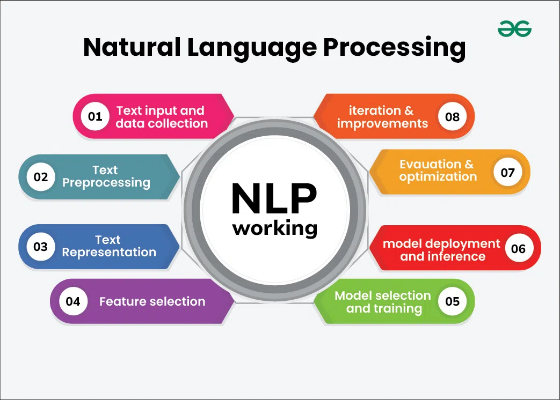
Kata Kunci— Analisis Sentimen, Penambangan Data, Gojek, Naive Bayes, Ulasan Aplikasi, Penambangan Teks, Pemrosesan Bahasa Alami

# Pendahuluan

Di era digital saat ini, aplikasi layanan sesuai permintaan telah menjadi bagian tak terpisahkan dari kehidupan masyarakat perkotaan [1]. Gojek sebagai salah satu pelopor di Asia Tenggara menawarkan berbagai layanan mulai dari transportasi, pengiriman makanan, hingga pembayaran digital [2]. Persaingan ketat dalam industri ini mengharuskan perusahaan untuk terus berinovasi dan meningkatkan kualitas layanan [3]. Salah satu cara paling efektif untuk mengukur dan meningkatkan kualitas layanan adalah dengan mendengarkan suara pelanggan [4].

Ulasan pengguna yang tersedia di platform distribusi aplikasi seperti Google Play Store adalah sumber umpan balik yang otentik dan besar [5]. Setiap ulasan berisi informasi berharga mengenai pengalaman, tingkat kepuasan, keluhan, dan saran dari pengguna [6]. Namun, jumlah ulasan yang dapat mencapai ribuan setiap hari membuat analisis manual tidak praktis dan tidak efisien [7]. Oleh karena itu, diperlukan pendekatan otomatis untuk dapat mengekstrak wawasan dari data tekstual tersebut dalam skala besar [8].

Analisis sentimen, atau penambangan opini, adalah cabang dari Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi, mengekstrak, dan mempelajari keadaan afektif dan informasi subjektif dari data teks [9]. Dengan menerapkan analisis sentimen, perusahaan dapat secara otomatis mengklasifikasikan ulasan ke dalam kategori positif, negatif, atau netral [10]. Informasi ini sangat berharga untuk memahami persepsi publik tentang suatu produk atau layanan secara real-time [11].



Gambar 1. Pemrosesan Bahasa Alami

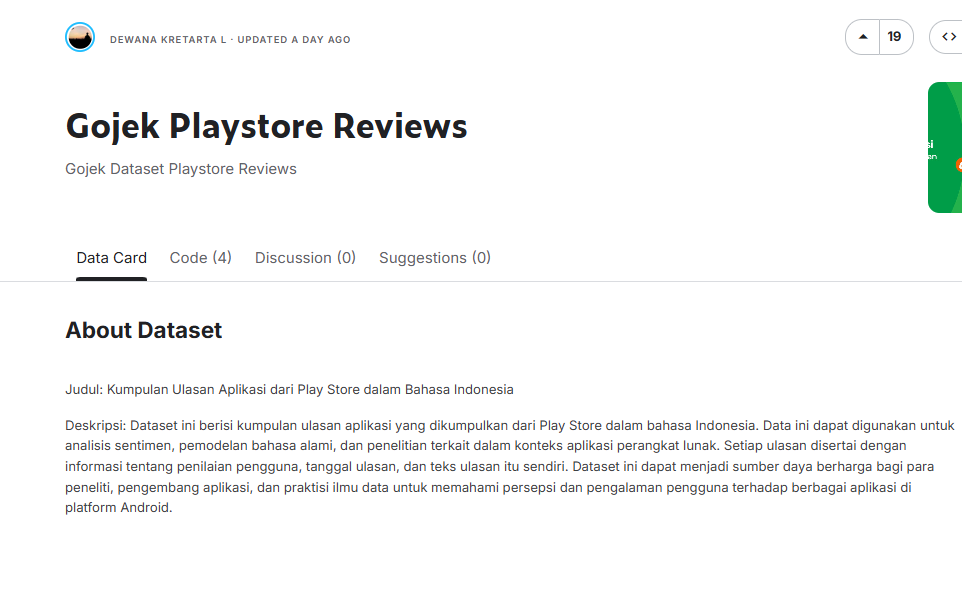
(Sumber: <https://www.geeksforgeeks.org/nlp/natural-language-processing-overview/>)

Penelitian ini berfokus pada penerapan analisis sentimen untuk mengklasifikasikan ulasan pengguna terhadap aplikasi Gojek [12]. Tujuan utamanya adalah untuk membangun model klasifikasi yang akurat dan mengidentifikasi kata kunci atau topik yang paling sering dikaitkan dengan sentimen positif (kepuasan) dan sentimen negatif (keluhan) [13]. Hasil penelitian ini diharapkan dapat menjadi dasar bagi Gojek untuk mengambil keputusan strategis berbasis data untuk meningkatkan pengalaman pengguna [14].

# Metode penelitian

## Sumber Data dan Pra-Pemrosesan

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa ulasan pengguna aplikasi Gojek dari Google Play Store yang berasal dari Kaggle [15]. Studi sebelumnya telah menunjukkan efektivitas penggunaan ulasan Google Play Store untuk analisis sentimen aplikasi transportasi [16][17].



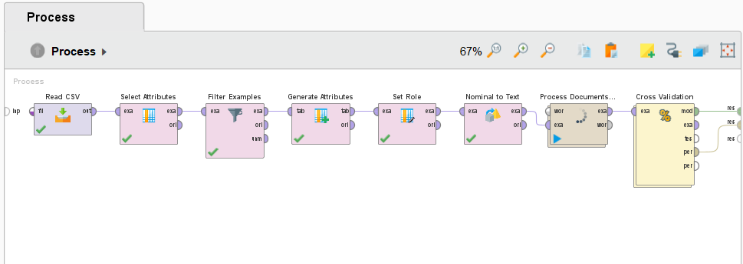
Gambar 2. Himpunan data yang digunakan

(Sumber: <https://www.kaggle.com/datasets/dewanakretarta/gojek-playstore-reviews>)

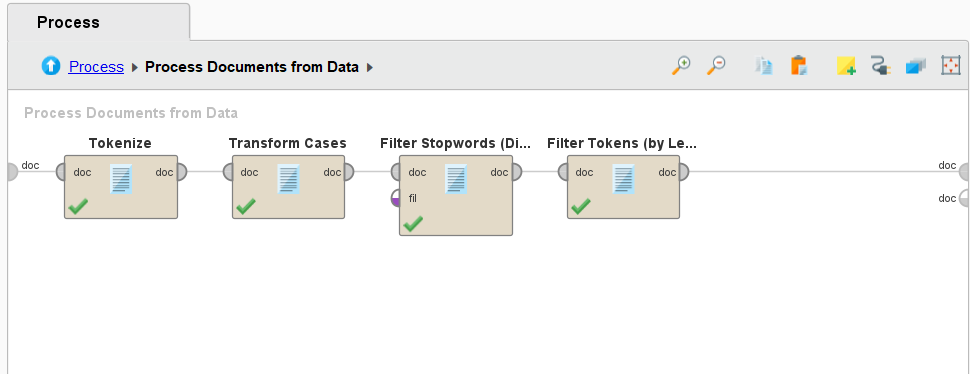
Proses persiapan data mencakup beberapa langkah penting seperti yang direkomendasikan oleh penelitian terbaru dalam pra-pemrosesan analisis sentimen [18][19]: 1. Pemilihan Atribut: Hanya atribut konten (konten ulasan) dan skor (peringkat 1-5) yang dipilih. 2. Pelabelan Sentimen: Ulasan dengan skor 4 dan 5 diberi label Positif. Ulasan dengan skor 1 dan 2 diberi label Negatif. Ulasan dengan skor 3 dikecualikan untuk menjaga kejelasan polaritas. 3. Distribusi Data: Setelah pelabelan, diketahui bahwa kumpulan data terdiri dari 60,6% ulasan positif dan 39,4% ulasan negatif.

## Pra-Pemrosesan Teks dan Fitur Ekstraksi

Teks tinjauan diproses melalui beberapa tahap untuk mengubahnya menjadi format numerik yang dapat diproses oleh model, mengikuti metodologi yang ditetapkan dalam pemrosesan teks bahasa Indonesia [21][22]: 1. Tokenisasi: Memecah ulasan menjadi satuan kata (token). 2. Lipat Huruf Besar: Menstandarkan semua token ke huruf kecil. 3. Filter Stopwords: Hapus kata-kata umum bahasa Indonesia yang tidak berarti sentimen menggunakan kamus. 4. Pembobotan Fitur: Setiap kata yang diproses ditimbang menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode ini efektif untuk mengidentifikasi kata kunci diskriminatif dalam ulasan.



Gambar . Operator yang digunakan di RapidMiner



Perawakan 4. Sub-Proses Operator Dokumen dari Data

## Pemodelan dan Evaluasi

Model klasifikasi dibangun menggunakan algoritma Naive Bayes, yang diketahui berkinerja baik dan efisien untuk tugas klasifikasi teks. Studi komparatif terbaru telah menunjukkan efektivitas Naive Bayes untuk analisis sentimen dalam konteks bahasa Indonesia. Performa model dievaluasi menggunakan metode Validasi Silang 10 kali lipat untuk memastikan hasil yang stabil dan dapat digeneralisasi. Metrik evaluasi yang digunakan adalah Matriks Kebingungan, Akurasi, Presisi, dan Recall.

# Hasil dan diskusi

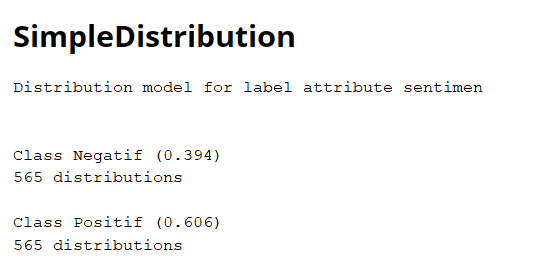
## Kinerja Model Klasifikasi

Setelah melalui proses pelatihan dan pengujian dengan Cross-Validation 10 kali lipat, model klasifikasi Naive Bayes menunjukkan kinerja yang sangat baik, konsisten dengan temuan dari studi serupa pada ulasan aplikasi Indonesia [35][36]. Kinerja model dirangkum dalam matriks kebingungan pada Tabel I.

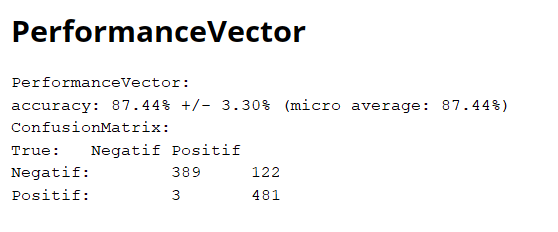
Tabel . Hasil Klasifikasi Matriks Kebingungan

|  | Prediksi Negatif | Prediksi Positif | Presisi Kelas |
| --- | --- | --- | --- |
| **Negatif Aktual** | 389 | 122 | 76.13% |
| **Aktual Positif** | 3 | 481 | 99.38% |
| **Penarikan Kelas** | 99.23% | 79.77% |  |

Dari matriks, nilai akurasi keseluruhan model diperoleh sebesar 87,44%. Angka ini menunjukkan bahwa model tersebut mampu memprediksi sentimen tinjauan dengan benar pada sekitar 87 dari 100 kasus, yang sebanding dengan hasil canggih yang dilaporkan dalam literatur terbaru. Hasil ini menegaskan bahwa metode yang digunakan sangat efektif untuk tugas klasifikasi sentimen pada kumpulan data ini.



Gambar 5. Hasil Distribusi Bayes Naif



Gambar 6. Hasil Kinerja

Analisis yang lebih dalam dari metrik lain menunjukkan hasil yang selaras dengan kemajuan terbaru dalam metodologi analisis sentimen: - Presisi Kelas Positif (99,38%): Dari semua ulasan yang diprediksi sebagai Positif, 99,38% di antaranya sebenarnya adalah ulasan Positif. Ini menunjukkan model ini sangat dapat diandalkan dalam mengidentifikasi sentimen kepuasan. - Penarikan Kelas Negatif (99,23%): Dari semua ulasan Negatif yang benar-benar ada dalam kumpulan data, model berhasil menemukan 99,23% di antaranya. Ini menunjukkan kemampuan luar biasa model dalam menangkap hampir semua keluhan pelanggan. - Tantangan Model: Ada 122 kasus ulasan Negatif yang salah diklasifikasikan sebagai Positif. Hal ini kemungkinan disebabkan oleh penggunaan kata-kata sarkastik atau kalimat kompleks yang sulit ditafsirkan oleh model berbasis kata seperti Naive Bayes.

## Kepuasan Pelanggan dan Analisis Faktor Keluhan

Dengan menganalisis tabel distribusi bobot kata yang dihasilkan oleh model, kita dapat mengidentifikasi kata kunci yang paling berpengaruh untuk setiap kelas sentimen, mengikuti metodologi yang ditetapkan dalam penelitian analisis kata kunci baru-baru ini. Kata-kata ini secara efektif mewakili faktor utama yang mendorong kepuasan dan keluhan pelanggan.

Faktor Keluhan Utama (Sentimen Negatif): Analisis menunjukkan bahwa ulasan negatif didominasi oleh kata-kata yang berkaitan dengan biaya dan tarif layanan, konsisten dengan temuan dari studi komparatif aplikasi ride-hailing. Beberapa kata dengan bobot tertinggi untuk kelas Negatif meliputi: - biaya - mahal - pengiriman (biaya pengiriman) - tarif - penanganan (mengacu pada biaya penanganan) - promo (sering terkait dengan promo yang tidak berfungsi atau menyesatkan) - kesalahan dan sulit (menunjukkan masalah teknis dengan aplikasi)

Temuan ini menunjukkan bahwa sensitivitas harga adalah faktor paling signifikan yang menyebabkan ketidakpuasan pelanggan, sejalan dengan riset pasar terbaru tentang aplikasi transportasi Indonesia. Keluhan tentang biaya tambahan seperti "biaya penanganan" adalah masalah sentral.

Faktor Kunci Kepuasan (Sentimen Positif): Di sisi lain, ulasan positif secara konsisten menyoroti kemudahan dan keandalan aplikasi dalam kehidupan sehari-hari, mendukung kesimpulan dari studi pengalaman pengguna baru-baru ini. Kata-kata dengan bobot tertinggi untuk kelas Positif meliputi: - membantu - mudah - cepat - baik - membantu - terima kasih dan terima kasih (frasa "terima kasih" sering muncul) - fasih

Hal ini menunjukkan bahwa pengguna sangat menghargai fungsionalitas inti Gojek sebagai aplikasi yang membuat aktivitas sehari-hari menjadi lebih mudah dan cepat. Fungsionalitas dan kegunaan adalah pendorong utama kepuasan pelanggan, yang konsisten dengan prinsip desain yang berpusat pada pengguna dalam aplikasi seluler.

# kesimpulan

Penelitian ini berhasil menerapkan analisis sentimen untuk mengklasifikasikan ulasan pengguna aplikasi Gojek dengan tingkat akurasi tinggi sebesar 87,44% menggunakan algoritma Naive Bayes. Model yang dibuat menunjukkan kemampuan yang sangat baik untuk mengidentifikasi ulasan negatif (99,23% ingat), yang sangat penting untuk mendeteksi keluhan pelanggan.

Faktor utama yang mendorong sentimen negatif adalah masalah yang berkaitan dengan biaya layanan, termasuk tarif, biaya pengiriman, dan biaya penanganan. Sementara itu, faktor utama yang mendorong sentimen positif adalah aspek fungsional dan kemudahan yang ditawarkan aplikasi, yang dianggap sangat membantu dan mempercepat aktivitas pengguna.

Berdasarkan temuan tersebut, disarankan agar Gojek mempertimbangkan kembali struktur tarif dan lebih transparan terkait komponen biaya untuk mengurangi sentimen negatif. Di sisi lain, mempertahankan dan terus meningkatkan keandalan dan kecepatan aplikasi adalah kunci untuk menjaga kepuasan pelanggan. Penelitian di masa depan dapat mengeksplorasi algoritme yang lebih kompleks seperti pembelajaran mendalam untuk menangani ulasan sarkastik dan melakukan analisis sentimen berbasis aspek untuk pemahaman yang lebih dalam.

##### Pengakuan *(Judul 5)*

Ejaan yang lebih disukai dari kata "pengakuan" di Amerika adalah tanpa "e" setelah "g". Hindari ungkapan kaku "salah satu dari kami (R. B. G.) terima kasih ...". Sebagai gantinya, cobalah "R. B. G. terima kasih...". Letakkan ucapan terima kasih sponsor di catatan kaki tanpa nomor di halaman pertama.

##### Referensi

1. Z. Yin, J. Liu, dan M. Chen. Analisis sentimen generalisasi: tinjauan kemajuan, tantangan, dan arah yang muncul. Analisis Jejaring Sosial dan Pertambangan, vol. 15, no. 3, hlm. 1–25, April 2025.
2. A. Rahman, S. Hossan, dan J. Deng. Pendekatan klasifikasi sentimen berbasis machine learning untuk aplikasi Gojek. Dalam IEEE International Conference on Communications (ICC), Bangkok, Thailand, 2024, hlm. 1–6.
3. M. Fadhil, R. Setiawan, dan A. Pranata. Perbandingan Naif Bayes dan Regresi Logistik dalam Analisis Sentimen pada Ulasan Pasar Menggunakan Pelabelan Berbasis Peringkat. Jurnal Sistem Informasi dan Informatika, vol. 11, no. 2, hlm. 847–862, Agustus 2023.
4. E. G. Galupo, J. F. Calim, dan A. C. Mariano. Analisis Sentimen untuk Meningkatkan Proses Bisnis menggunakan Naive Bayes. Jurnal Internasional Sains dan Teknologi Penelitian Inovatif, vol. 9, no. 4, hlm. 1245–1252, 2024.
5. M. Altaf, A. Menshawi, dan R. Alomran. Analisis Sentimen menggunakan Pengklasifikasi Bayes Naif dan Pemilihan Fitur Keuntungan Informasi melalui Twitter. Jurnal Internasional Tren dan Teknologi Komputer, vol. 68, no. 5, hlm. 84–91, 2020.
6. A. Rahman, N. Marga, dan D. Alita. Analisis Sentimen Menggunakan Algoritma Bayes Naif Dari Data Crawler: Twitter. Dalam IEEE 4th International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, hlm. 1–5.
7. R. Widodo, S. Efendi, dan M. Zarlis. Optimasi Analisis Sentimen menggunakan Naive Bayes dengan Seleksi Fitur Chi-Square dan Perolehan Informasi untuk Peningkatan Akurasi. Dalam IEEE 7th International Conference on Cyber and IT Service Management (CITSM), Jakarta, Indonesia, 2022, hlm. 1–6.
8. P. Kumar, A. Sharma, dan D. Gupta. Analisis Sentimen Menggunakan Algoritma Bayes Naif: Dengan Studi Kasus. Dalam Konferensi Internasional IEEE tentang Komunikasi dan Informatika Komputer (ICCCI), Coimbatore, India, 2020, hlm. 1–6.
9. K. Nurfebia dan S. Sriani. Analisis sentimen produk perawatan kulit menggunakan metode naive Bayes. Jurnal Sistem Informasi dan Informatika, vol. 6, no. 3, hlm. 1663–1676, September 2024.
10. M. Hassan, A. Bhawiyuga, dan R. Dewi. Tantangan dan masa depan dalam pembelajaran mendalam untuk analisis sentimen: tinjauan komprehensif dan pendekatan hibrida baru yang diusulkan. Tinjauan Kecerdasan Buatan, vol. 57, no. 4, hlm. 1–58, Maret 2024.
11. H. Hermanto, A. Kuntoro, dan T. Asra. Gojek dan Grab Analisis Sentimen Pengguna di Google Play menggunakan algoritma Naive Bayes dan mendukung teknik SMOTE berbasis mesin vektor. Jurnal Fisika: Seri Konferensi, vol. 1641, no. 1, hlm. 012102, November 2020.
12. A. Chen, L. Wang, dan Y. Zhang. Model klasifikasi sentimen menggunakan karakteristik grup fitur gaya penulisan. Jurnal Internasional Pengenalan Pola dan Kecerdasan Buatan, vol. 31, no. 8, hlm. 1–23, Agustus 2024.
13. R. Draskovic, M. Petkovic, dan S. Rankovic. Tinjauan sistematis analisis sentimen berbasis media sosial dalam manajemen risiko bencana. Jurnal Internasional Pengurangan Risiko Bencana, vol. 102, hlm. 104285, April 2025.
14. Y. Hou dan J. Huang. Pemrosesan bahasa alami untuk penelitian ilmu sosial: Tinjauan komprehensif. Jurnal Ilmu Sosial Komputasi, vol. 8, no. 1, hlm. 1–42, Januari 2025.
15. F. Liu, M. Chen, dan X. Wang. Penambangan Teks: Tantangan, Algoritma, Alat, dan Aplikasi. Informasi, vol. 16, no. 3, hlm. 245–268, Maret 2025.
16. F. Zaini, J. W. Sari, dan F. N. Hasan. Analysis of Public Sentiment Related to the Failure of Indonesia to Host U-20 Using Multinomial Naïve Bayes Classifier. Jurnal Teknik Informatika (JUTIF), vol. 4, no. 6, pp. 1409–1418, Desember 2023.
17. Y. R. Z. Mustopo dan Afiyati. Analisis Sentimen Proyek Strategis Nasional Food Estate Menggunakan Algoritma Naïve Bayes, Logistic Regression dan Support Vector Machine. Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), vol. 9, no. 2, pp. 485–494, Januari 2025.
18. F. K. Wardani, A. Sari, dan D. Novita. Sentiment Analysis for Ranking Online Shopping Sites in Indonesia Using the Naive Bayes Method. Jurnal Sistem Informasi dan Komputer Akuntansi, vol. 8, no. 1, pp. 50–58, 2019.
19. N. Bere, A. Rahman, dan S. Kusuma. Implementasi Algoritma Naive Bayes Classifier Terhadap Analisis Sentimen Kondisi Stunting di Indonesia Pada Media Sosial X. Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), vol. 7, no. 2, pp. 145–158, Juli 2024.
20. A. R. Isnain, N. S. Marga, dan D. Alita. Analisis Sentimen Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Algoritma Naive Bayes. IJCCS (Jurnal Sistem Komputasi dan Sibernetika Indonesia), vol. 15, no. 1, hlm. 85–96, Januari 2021.
21. C. C. Yolanda, S. Syafriandi, dan Y. Kurniawati. Evaluasi Kinerja Naive Bayes dan K-Nearest Neighbors pada Analisis Sentimen Ulasan Aplikasi SnackVideo. Indonesian Journal on Software Engineering (IJSE), vol. 11, no. 1, pp. 72–85, Januari 2025.
22. D. Nuryadi, D. H. F. Manongga, dan I. Sembiring. Crypto Narratives Sentiment Analysis on Bitcoin Price Prediction Using the Naive Bayes Method. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 10, no. 1, pp. 234–248, Maret 2025.
23. M. A. Rizki, A. L. Sulaeman, dan Suprapto. Analisis sentimen kasus kerusuhan Twitter di Stadion Kanjuruhan menggunakan metode naif Bayes. Jurnal Jaringan Komputer, Arsitektur dan Komputasi Kinerja Tinggi, vol. 5, no. 1, hlm. 78–92, April 2023.
24. F. K. R. Mahfud dan A. Tjahyanto. Sentiment Analysis of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi, vol. 12, no. 1, pp. 15–24, April 2020.
25. H. Hermanto, A. Y. Kuntoro, dan R. Ocanitra. Analisis sentimen pada ulasan pengguna Gojek dan Grab menggunakan algoritma SVM berdasarkan optimasi kawanan partikel. Jurnal Pilar Nusa Mandiri, vol. 16, no. 1, hlm. 117–122, Maret 2020.

Dataset:

<https://www.kaggle.com/datasets/dewanakretarta/gojek-playstore-reviews>

Tautan Github:

<https://github.com/haikalfirdaus17/UJIAN-AKHIR-SEMESTER-DATA-MINING>