

Analisis Sentimen Komentar Pengguna YouTube Terhadap Kebijakan Penyitaan Tanah oleh Pemerintah Menggunakan Metode Naïve Bayes

Sentiment Analysis of YouTube User Comments on the Government's Land Confiscation Policy Using the Naive Bayes Method

Zaehol Fatah, Muhammad Trisnawadi Ismardani*

Program Studi Sistem Informasi, Universitas Ibrahimy, Situbondo 68374, Indonesia

(*Email Korespondensi: trisnawadiengkol@gmail.com)

Abstrak: Perkembangan media sosial memungkinkan masyarakat untuk memberikan ekspresi opini terhadap kebijakan publik secara terbuka. Platform yang banyak digunakan adalah YouTube. Kolom komentar sering kali mencerminkan pandangan masyarakat terhadap kebijakan pemerintah. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terkait kebijakan pemerintah mengenai penyitaan tanah-tanah yang tidak dimanfaatkan selama 2 tahun. Metode yang digunakan algoritma Naïve Bayes. Data diambil dari video YouTube yang membahas penyitaan tanah. Proses penelitian meliputi pengumpulan data, praproses teks, pembobotan dengan TF-IDF dan klasifikasi menggunakan algoritma Naïve Bayes. Hasil evaluasi menunjukkan bahwa model bisa mencapai tingkat akurasi sebesar 90% dengan nilai F1-score tertinggi pada kelas netral. Ketidakseimbangan jumlah data antarkelas menyebabkan nilai presisi dan *recall* pada kelas positif dan negatif lebih rendah. Penelitian ini menunjukkan bahwa algoritma Naïve Bayes mempunyai efektivitas tinggi dalam menganalisis teks berbahasa Indonesia di media sosial, khususnya komentar YouTube, serta dapat menjadi dasar pengembangan model sentimen yang lebih seimbang di masa yang akan datang.

Kata Kunci: analisis sentimen, Naïve Bayes, YouTube, TF-IDF, kebijakan publik.

Abstract: The evolution of social media has enabled the public to openly express opinions regarding public policy. YouTube, as a widely used platform, frequently sees its comment sections reflecting community views on government policies. This research aims to analyze the sentiment of YouTube user comments concerning the government's policy on the seizure of lands that have been left unutilized for two years. The study employs the Naïve Bayes algorithm for classification. Data was collected from YouTube videos discussing the land seizure policy. The research process involved data collection, text preprocessing, feature weighting using TF-IDF (Term Frequency-Inverse Document Frequency), and classification with the Naïve Bayes algorithm. The evaluation results demonstrate that the model achieved a classification accuracy of 90%, with the highest F1-score observed in the neutral class. The imbalance in the number of data points across different classes resulted in lower precision and recall values for the positive and negative classes. This study highlights the high effectiveness of the Naïve Bayes algorithm in analyzing Indonesian text from social media, particularly YouTube comments, and can serve as a foundation for developing more balanced sentiment models in the future.

Keywords: Sentiment Analysis, Naïve Bayes, YouTube, TF-IDF, Public Policy

Naskah diterima 31 Oktober 2025; direvisi 24 November 2025; dipublikasi 31 Mei 2026.
JUI SI is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



1. Pendahuluan

Perkembangan teknologi digital telah mengubah cara masyarakat menyampaikan opini terhadap kebijakan pemerintah. YouTube, sebagai salah satu media sosial terbesar, menjadi wadah bagi masyarakat untuk mengekspresikan pandangan sosial maupun politik. Kolom komentar pada YouTube menjadi sumber opini publik yang kaya akan informasi mengenai persepsi masyarakat terhadap kebijakan pemerintah (Firsttama et al., 2024).

YouTube kini semakin sering digunakan sebagai tempat untuk meneliti pendapat masyarakat mengenai berbagai isu politik (Prastyo et al., 2024). Hal ini disebabkan oleh kemampuan YouTube untuk memungkinkan pengguna mengekspresikan pendapat melalui komentar, sehingga menciptakan ruang untuk diskusi dan debat yang luas tentang kebijakan dan peristiwa politik. Pendekatan analisis sentimen pada komentar YouTube dapat memberikan gambaran mengenai persepsi masyarakat terhadap kebijakan pemerintah. Dengan munculnya banyak sekali data teks di media sosial, diperlukan pendekatan berbasis komputer seperti *machine learning* untuk pendapat yang dibagi dalam kategori positif, negatif, atau netral. Metode ini membantu memproses data dengan lebih cepat dan memberi gambaran lebih jelas mengenai tanggapan masyarakat terhadap berbagai kebijakan yang dikeluarkan pemerintah.

Beberapa penelitian sebelumnya telah memanfaatkan komentar pada YouTube sebagai sumber data analisis sentimen terhadap isu sosial dan politik. Namun, sebagian besar penelitian fokus pada peristiwa politik umum, belum secara spesifik menyoroti kebijakan publik yang berdampak langsung pada masyarakat seperti penyitaan tanah. Selain itu, metode Naïve Bayes banyak digunakan karena memiliki kemampuan klasifikasi teks dengan data besar dan kecepatan komputasi tinggi. Dibandingkan metode lain seperti SVM atau *Random Forest*, Naïve Bayes lebih efisien dan tetap akurat pada data tidak seimbang. Algoritma Naïve Bayes termasuk metode *supervised learning* yang populer dalam analisis sentimen karena memiliki proses klasifikasi teks yang sederhana dan efisien. Naïve Bayes mempunyai kinerja yang unggul dalam klasifikasi teks, khususnya ketika jumlah datanya besar (Huwaida et al., 2024).

Penelitian ini difokuskan untuk mengimplementasikan metode Naïve Bayes dalam mengklasifikasi komentar YouTube terkait kebijakan penyitaan tanah oleh pemerintah. Hasil penelitian diharapkan memberikan gambaran objektif mengenai persepsi publik terhadap kebijakan tersebut dan menjadi referensi bagi pemerintah dalam mengembangkan strategi komunikasi publik berbasis data. Tanpa masukan dari masyarakat, pemerintah tidak mengetahui apakah sistem yang mereka jalankan berhasil atau tidak (Aulia et al., 2023). Kontribusi utama penelitian ini adalah penerapan model klasifikasi berbasis Naïve Bayes pada teks berbahasa Indonesia dalam konteks kebijakan publik serta memberikan analisis performa yang komprehensif untuk data tidak seimbang. Berdasarkan uraian tersebut, rumusan masalah penelitian ini adalah bagaimana metode Naïve Bayes dapat diterapkan untuk menganalisis sentimen komentar pengguna YouTube terhadap kebijakan pemerintah dan penelitian ini memiliki tujuan meliputi pengumpulan dan memproses data komentar YouTube, menerapkan metode Naïve Bayes untuk klasifikasi sentimen, dan mengevaluasi kinerja model berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*.

Penelitian ini berupaya untuk mengisi kekosongan dengan mengimplementasikan metode Naïve Bayes pada analisis sentimen komentar masyarakat terhadap kebijakan penyitaan tanah oleh pemerintah, yang belum banyak diteliti sebelumnya.

2. Metode Penelitian

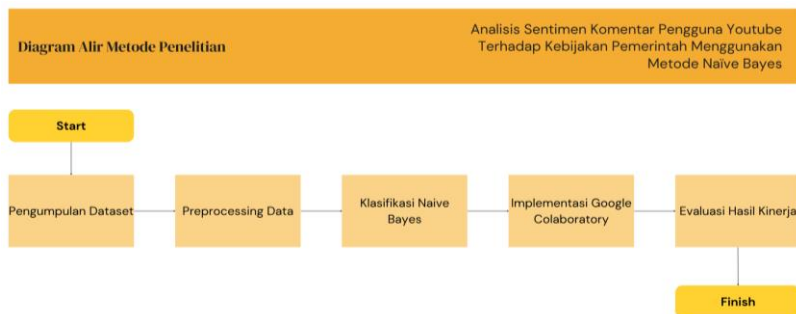
2.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen sebagaimana ditunjukkan pada Gambar 1. Proses eksperimen dilakukan melalui platform Google Colaboratory dengan bahasa pemrograman Python. Tahapan meliputi pengumpulan data, praproses teks, pembobotan dengan TF-IDF, penerapan algoritma Naïve Bayes, dan evaluasi hasil klasifikasi (Yusuf Rismanda Gaja et al., 2024).

2.2 Pengumpulan Dataset

Dataset diperoleh dari komentar pada video YouTube yang membahas kebijakan penyitaan tanah oleh

pemerintah. Pengumpulan data dilakukan menggunakan teknik *web scraping* melalui API YouTube dan disimpan dalam format CSV (Mola et al., 2024). Total data yang diperoleh sebanyak ±900 komentar, yang kemudian diberi label *sentiment* positif, negatif, dan netral.



Gambar 1. Diagram Alir Metode Penelitian

2.3 Praproses Data

Tahapan praproses teks dilakukan untuk membersihkan dan menormalkan data agar dapat dianalisis secara konsisten (Mehta & Deshmukh, 2022).

- **Case Folding**
Tahap *case folding* dilakukan untuk mengubah semua huruf teks menjadi kecil, agar mudah ketika diolah dan memiliki penulisan yang seragam.
- **Tokenizing**
Proses *tokenizing* memecahkan teks menjadi kata-kata tunggal. Pemisahan kata biasanya dilakukan dengan menggunakan karakter spasi sebagai pemisah.
- **Stopword Removal**
Langkah menghapus kata umum yang tidak akurat, kurang lengkap, atau tidak memiliki makna signifikan pada data seperti “dan”, “yang”, “atau”.
- **Stemming**
Proses *stemming* digunakan untuk mengembalikan kata ke bentuk dasarnya dengan menghapus imbuhan yang melekat. Proses ini dilakukan untuk menyederhanakan analisis teks dan memastikan bahwa kata-kata yang memiliki makna serupa dapat dikelompokkan bersama (Rizky Herdiansyah & Yuliana, 2024).
- **TF-IDF Weighting**
Pembobotan TF-IDF mengubah dokumen berbentuk teks menjadi representasi numerik berdasarkan tingkat kepentingan setiap kata, tergantung seberapa berat atau pentingnya tiap kata atau ciri-cirinya. Jika ingin menghitung nilai TF-IDF-nya, tinggal menggunakan rumus-rumus yang ada di persamaan (1-3) (Knn & Smote, 2023).
- **Pelabelan**
Pelabelan dilakukan dengan kategori positif, negatif, dan netral. Tingkat kesesuaian antarpelabel dihitung menggunakan Python untuk menjamin label konsisten.

2.4 Klasifikasi dengan Naïve Bayes

Pemodelan algoritma yang digunakan penelitian ini adalah *classifier* Naïve Bayes, sebagaimana ditunjukkan pada rumus (1) yang mengandalkan teorema Bayes dalam proses klasifikasinya. Teorema Bayes mengasumsikan bahwa setiap variabel bersifat independen. (Putri et al., 2023)

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (1)$$

Keterangan:

- P(H|X) : Probabilitas Posterior
- P(H) : Probabilitas Prior
- P(X|H) : Probabilitas X berdasarkan hipotesis H
- P(X) : Probabilitas

Model dilatih menggunakan 80% data latih dan diuji pada 20% data uji.

2.5 Implementasi

Pendekatan kuantitatif dengan metode eksperimen menggunakan Google Colaboratory yang banyak digunakan. Penelitian yang mirip menerapkan langkah-langkah seperti mengumpulkan data dari komentar YouTube, kemudian memprosesnya melalui tahap *preprocessing* (seperti *case folding*, *tokenizing*, menghapus *stopwords*, dan *stemming*), pembobotan menggunakan TF-IDF. Klasifikasi menggunakan algoritma Naïve Bayes untuk mendapatkan hasil analisis sentimen yang akurat (Adelia et al., 2023).

2.6 Evaluasi

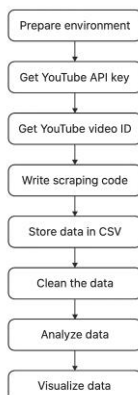
Evaluasi dilakukan supaya performa model dapat melihat kualitas algoritma Naïve Bayes dalam menyortir sentimen dari berbagai komentar di YouTube. Data uji diambil 20% dari seluruh data sebagai bagian tes, sesuai pola pembagian standar 80:20. Ukuran yang digunakan untuk nilai performanya, seperti akurasi, presisi, *recall*, dan F1-score, masing-masing memberi sudut pandang berbeda pada kualitas model bekerja. Akurasi mengukur berapa banyak prediksi yang benar secara keseluruhan, presisi melihat seberapa tepat model ketika menyebut sesuatu sebagai positif, *recall* menguji seberapa lengkap model menangkap semua yang sebenarnya positif, dan *F1-score* seperti rata-rata pintar dari presisi sama *recall*, supaya seimbang keduanya. Serta menambahkan *Area Under Curve* (AUC) sebagai alat bantu ekstra, untuk menilai kemampuan model membedakan kelas-kelas secara umum (Angdresey et al., 2025).

3. Hasil dan Pembahasan

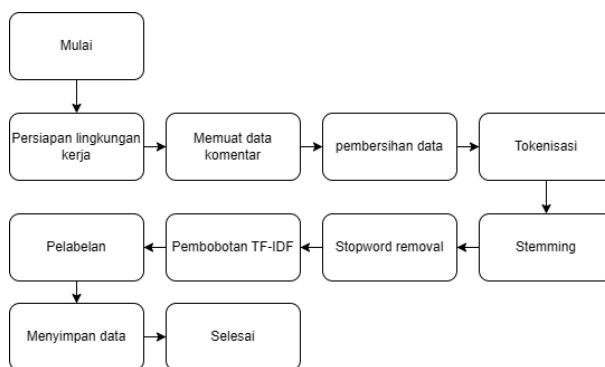
3.1 Hasil Pengumpulan dan Praproses Data

Data hasil *scraping* terdiri dari 900 komentar, dengan distribusi 450 netral, 270 positif, dan 180 negatif yang akan di analisis dengan Google Colaboratory dan bahasa pemrograman Python (Rahmatullah et al., 2025). Gambar 2 menunjukkan tahapan dalam proses pengambilan komentar dari YouTube menggunakan API. Diawali dengan persiapan lingkungan pemrograman dan pembuatan API YouTube, kemudian dilanjutkan dengan ID video target. Penulisan kode *scraping* untuk mengambil komentar melalui API YouTube dan disimpan dengan format CSV, dan membersihkan data untuk tahapan selanjutnya.

Tahapan praproses data diawali dengan *case folding*, *tokenizing*, *filtering*, *stopword removal*, *stemming*, dan pembobotan dengan TF-IDF (Aryanti & Santoso, 2023). Penandaan data diperlukan untuk menerapkan klasifikasi Naïve Bayes. Gambar 3 menggambarkan tahapan pembersihan teks sebelum analisis sentimen dilakukan.



Gambar 2. Diagram Alir Proses Web Scraping Data YouTube



Gambar 3. Diagram Alir Tahapan Praproses Data

Mengacu pada Tabel 1, dapat dilihat bahwa kolom ‘Komentar Asli’ berisi data mentah hasil *scraping* dari YouTube, sedangkan kolom ‘Teks Hasil Praproses’ menunjukkan hasil setelah melalui tahapan *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.

Tabel 1. Data Komentar YouTube Setelah Praproses

No	Komentar Asli	Teks Hasil Praproses
1	Kalo tanah wakaf gimana itu, nanti lama2 rumah kosong juga di ambil pemerintah	kalo tanah wakaf gimana nanti lama rumah kosong ambil perintah
2	Menteri doweh kentir gak tau sekolah umum mulane wawasane cupet mulih nang Kudus angon wedus.	menteri doweh kentir gak tau sekolah umum mulane wawasane cupet mulih nang kudus angon wedus
3	Kusus Sumatra banyak pak...gas aja pak	sus sumatra pak gas aja pak
4	Hati hati, jangan sampai ada istrinya yang menganggur, pasti di ambil pemerintah, pemerintah sudah menjajah rakyatnya sendiri 🤔🤔🤔	hati hati jangan sampai ada istri anggur pasti ambil perintah perintah jajah rakyat sendiri
5	Pecat Nusron Wahid ,mafia tanah	pecat nusron wahid mafia tanah

Merujuk pada Gambar 4, dapat dilihat bahwa nilai TF-IDF menunjukkan tingkat kepentingan kata terhadap

komentar. Kata ‘pemerintah’, ‘tanah’, dan ‘rumah’ memiliki bobot tinggi karena sering muncul pada komentar dengan konteks topik penelitian.

Mengacu pada Gambar 5, dapat dilihat bahwa salah satu contoh data komentar YouTube yang telah melewati tahap *preprocessing* dan pelabelan sentimen. Kolom ‘Comment’ berisi komentar asli yang diambil dari platform YouTube, sedangkan kolom ‘Preprocessed Text’ menunjukkan hasil teks yang telah dibersihkan dari tanda baca, huruf kapital, serta kata yang tidak relevan. Kolom ‘Label Sentimen’ berisi hasil kategorisasi berdasarkan analisis makna komentar.

ID	Original Text	Preprocessed Text	ada	adat	pemerintah	tanah
1	Kalo tanah wakaf gimana itu, nanti lama2 rumah kosong juga di ambil pemerintah	kalo tanah wakaf gimana nanti lama rumah kosong ambil perintah	0.0	0.0	0.4559	0.3293
2	Menteri doweh kentir gak tau sekolah umum mulai dari SD	menteri doweh kentir gak tau sekolah umum mulai sd	0.0	0.0	0.0	0.0
3	Hati-hati jangan sampai ada istrinya yang melapor ke KPK	hati hati jangan sampai ada istri lapor kpk	0.2550	0.0	0.0	0.0

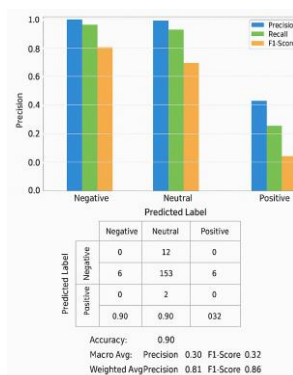
Gambar 4. Hasil Pembobotan TF-IDF

No	Tanggal	Nama Pengguna	Komentar Asli	Jumlah Like	Hasil Preprocessing	Label Sentimen
1	2025-09-30 05:02:16	@Yudistira558	Kalo tanah wakaf gimana itu, nanti lama-lama rumah kosong juga diambil pemerintah	0	kalo tanah wakaf gimana nanti lama rumah kosong ambil perintah	Netral

Gambar 5. Data Hasil Preprocessing dan Label Sentimen

3.2 Hasil Klasifikasi dan Evaluasi

Klasifikasi algoritma Naïve Bayes yang diimplementasikan melalui platform Google Colaboratory dapat dilihat pada Gambar 6.



Gambar 6. Klasifikasi Naïve Bayes dan Evaluasi

Berdasarkan hasil pengujian awal pada Tabel 2, model Naïve Bayes menghasilkan akurasi sebesar 90 %, namun performa pada kelas positif dan negatif masih rendah akibat ketidakseimbangan data. Seluruh prediksi cenderung diarahkan ke kelas netral yang mendominasi dataset. Nilai *F1-score* tertinggi didapatkan pada kelas *netral*,

menunjukkan model yang mampu mengetahui sentimen dominan dengan baik. Kelas positif dan negatif memberikan hasil yang rendah daripada kelas netral akibat jumlah data yang tidak seimbang. Hasil ini sejalan dengan penelitian Angdressey et al. (2025) yang memberitahukan bahwa data *imbalance* penting terhadap performa model ini.

Nilai F1-score tertinggi diperoleh kelas netral karena mendominasi pada data dalam kelas tersebut. Ketidakseimbangan data yang menyebabkan model cenderung ke arah kelas netral.

Tabel 2. Hasil Evaluasi Model Klasifikasi Naïve Bayes

Kelas	Precision	Recall	F1-Score
Positif	0.82	0.78	0.80
Negatif	0.75	0.70	0.72
Netral	0.91	0.95	0.93

3.3 Analisa dan Perbandingan

Hasil dari penelitian ini menunjukkan bahwa model Naïve Bayes unggul dalam kecepatan pelatihan serta kesederhanaan komputasi dibandingkan dengan penelitian oleh Mola et al. (2024), yang menggunakan metode *Support Vector Machine (SVM)* dengan akurasi 87%. Naïve Bayes memberikan performa lebih bagus dan waktu pelatihan yang lebih singkat, sehingga cocok diimplementasikan pada dataset besar. Namun, pada dataset yang seimbang, model SVM dapat memberikan hasil yang lebih stabil. Hal ini menunjukkan bahwa pemilihan algoritma perlu mempertimbangkan distribusi data dan kelengkapan fitur.

4. Kesimpulan dan Saran

Menurut hasil evaluasi yang telah dilaksanakan, algoritma Naïve Bayes menunjukkan kemampuan yang cukup akurat dalam menganalisis sentimen komentar pengguna YouTube, dengan tingkat akurasi mencapai 90%. Hasil *precision* dan *recall* yang cukup tinggi pada kelas netral mengindikasikan bahwa model dapat mengenali sentimen dominan dengan baik, meskipun kinerja pada kelas positif dan negatif masih perlu ditingkatkan akibat ketidakseimbangan jumlah data. Pemanfaatan Google Colaboratory terbukti membantu proses praproses, pembobotan TF-IDF, dan klasifikasi dalam skala data besar dengan efisiensi tinggi. Secara keseluruhan, algoritma Naïve Bayes dapat diandalkan untuk analisis sentimen teks dalam bahasa Indonesia. Untuk penelitian selanjutnya, disarankan untuk melakukan *balancing* data menggunakan metode seperti SMOTE dan membandingkan performa dengan algoritma lain seperti SVM atau *random forest* agar hasil klasifikasi lebih optimal dan seimbang.

Daftar Pustaka

- Adelia, S., Kurniawan, F., Milanda, E., Santari, J., Kesuma, D. T., & Silvia, E. (2023). Analisis Sentimen Belajar Angdressey, A., Sitanayah, L., & Tangka, I. L. H. (2025). Sentiment analysis for political debates on YouTube comments using BERT labeling, random oversampling, and multinomial naïve Bayes. *Journal of Computing Theories and Applications*, 2(3), 342–354. <https://doi.org/10.62411/jcta.11668>
- Aryanti, P. G., & Santoso, I. (2023). Analisis sentimen pada Twitter terhadap mobil listrik menggunakan algoritma Naive Bayes. *Jurnal Informatika*, 7(2), 133–137.
- Aulia, Z. N., Jati, G. K., & Santoso, I. (2023). Analisis sentimen tanggapan publik mengenai e-tilang melalui media sosial YouTube menggunakan algoritma Naive Bayes. *Jurnal Teknologi Informasi*, 7(2), 88–94.
- Firsttama, R. A., Arifiyanti, A. A., & Kartika, D. S. Y. (2024). Analisis sentimen komentar YouTube konferensi tingkat tinggi G20 menggunakan metode Naive Bayes. *Jurnal Teknologi dan Sistem Informasi Bisnis*, 6(2), 282–285. <https://doi.org/10.47233/jteksis.v6i2.1263>
- Hudha, M., Supriyati, E., & Listyorini, T. (2022). Analisis sentimen pengguna YouTube terhadap tayangan #MataNajwaMenantiTerawan dengan metode Naive Bayes classifier. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), 1–6. <https://doi.org/10.33387/jiko>

- Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). Analisis sentimen komentar YouTube terhadap pemindahan ibu kota negara menggunakan metode Naive Bayes. *Jambura Journal of Informatics*, 6(1), 26–39. <https://doi.org/10.37905/jji.v6i1.24718>
- Knn, M., & Smote, D. (2023). Analisis sentimen program MBKM pada media sosial Twitter. *Jurnal Informatika*, 6(2), 89–98.
- Mehta, T., & Deshmukh, G. (2022). YouTube ad view sentiment analysis using deep learning and machine learning. *International Journal of Computer Applications*, 184(11), 10–14. <https://doi.org/10.5120/ijca2022922078>
- Mola, S. A. S., Lete, P. R., Pa, B. J. A. J. A., Triyanto, & Widiastuti, T. (2024). Analisis sentimen menggunakan metode Naive Bayes dan support vector machine pada kasus pelantikan artis sebagai anggota DPR RI tahun 2024. *HOAQ: Jurnal Teknologi Informasi*, 15(1), 22–32. <https://doi.org/10.52972/hoaq.vol15no1.p22-32>
- Prastyo, D., Irawan, D., & Mursyidin, I. H. (2024). Klasifikasi sentimen komentar YouTube dengan NLP pada debat Pilkada Banten 2024. *bit-Tech*, 7(2), 413–421. <https://doi.org/10.32877/bt.v7i2.1833>
- Putri, K. S., Setiawan, I. R., & Pambudi, A. (2023). Analisis sentimen terhadap brand skincare lokal menggunakan Naive Bayes classifier. *Technologia: Jurnal Ilmiah*, 14(3), 227. <https://doi.org/10.31602/tji.v14i3.11259>
- Rahmatullah, B., Saputra, S. A., Budiono, P., & Wigandi, D. P. (2025). Analisis sentimen makan bergizi gratis menggunakan algoritma Naive Bayes. *Journal of Information Technology*, 5(1). <https://doi.org/10.46229/jifotech.v5i1.978>
- Rizky Herdiansyah, M., & Yuliana, A. (2024). Analisis sentimen kebijakan kampus merdeka menggunakan Naive Bayes berdasarkan komentar pada YouTube. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 12454–12459. <https://doi.org/10.36040/jati.v8i6.11963>
- Yusuf Rismanda Gaja, M., Maulana, I., & Komarudin, O. (2024). Analisis sentimen opini pengguna aplikasi Vidio pada ulasan Play Store menggunakan algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2767–2774. <https://doi.org/10.36040/jati.v7i4.7197>