

ANALISIS SENTIMEN TWITTER PEMILIHAN PRESIDEN 2024 MENGUNAKAN NAÏVE BAYES

Rudiyono¹, Rudi Kurniawan², Bani Nurhakim³, Deni Pratama⁴, Riri Narasati⁵

Program Studi Teknik Informatika^{1,2,5}
Program Studi Manajemen Informatika³
Program Studi Komputerisasi Akuntansi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
rudiyono0402@gmail.com

(*) Corresponding Author : rudiyono0402@gmail.com
Published : 30 Maret 2026

Abstract - The 2024 Indonesian Presidential Election became one of the most widely discussed political events on social media, especially Twitter, which functions as a digital public sphere. The large volume of unstructured text requires automated analytical methods capable of identifying and classifying sentiment quickly and accurately. This study aims to analyze public sentiment regarding the 2024 Presidential Election using the Naïve Bayes algorithm and evaluate model performance based on accuracy, precision, and recall. The research follows the Knowledge Discovery in Databases (KDD) stages: data collection, preprocessing, transformation, and sentiment classification using RapidMiner 10.1. The results show that Naïve Bayes performs effectively in classifying positive, negative, and neutral sentiment categories with stable performance metrics. Sentiment distribution indicates diverse public responses toward political issues. The confusion matrix results demonstrate that Naïve Bayes works optimally with short and informal Twitter texts. This study contributes to understanding digital public opinion dynamics and provides analytical insights for policymakers and political observers.

Keywords: Sentiment Analysis, Naïve Bayes, Twitter, 2024 Election, Public Opinion

Abstrak - Pemilihan Presiden Republik Indonesia tahun 2024 menjadi salah satu momen politik yang paling banyak dibicarakan di media sosial, khususnya Twitter. Platform ini berfungsi sebagai ruang publik digital tempat masyarakat menyampaikan opini, kritik, dan dukungan terhadap kandidat. Besarnya jumlah data teks yang tidak terstruktur menuntut adanya metode analisis otomatis yang mampu mengidentifikasi dan mengklasifikasikan sentimen secara cepat dan akurat. Penelitian ini bertujuan untuk menganalisis sentimen opini publik terkait Pemilu Presiden 2024 dengan menggunakan algoritma Naïve Bayes serta mengevaluasi kinerja model berdasarkan nilai akurasi, presisi, dan recall. Metode penelitian mengikuti tahapan Knowledge Discovery in Databases (KDD), meliputi pengumpulan data Twitter, preprocessing (cleansing, tokenizing, stopword removal, stemming), transformasi data, dan proses klasifikasi menggunakan RapidMiner 10.1. Hasil penelitian menunjukkan bahwa model Naïve Bayes mampu mengklasifikasikan tiga kategori sentimen positif, negatif, dan netral secara efektif dengan tingkat akurasi yang stabil. Persebaran sentimen menunjukkan beragamnya opini publik terhadap isu politik selama masa kampanye Pemilu 2024. Hasil evaluasi confusion matrix memperlihatkan bahwa Naïve Bayes bekerja optimal pada teks pendek dan informal khas Twitter. Penelitian ini memberikan kontribusi pada pemahaman dinamika opini publik digital serta dapat menjadi dasar pengambilan keputusan bagi pengamat politik dan pihak berkepentingan dalam memantau persepsi masyarakat terhadap isu nasional.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Twitter, Pemilu 2024, Opini Publik

INTRODUCTION

Media sosial telah menjadi ruang publik digital yang sangat berpengaruh dalam pembentukan opini politik (Adib et al., 2024). Twitter, sebagai salah satu platform dengan pengguna aktif terbesar di Indonesia, menjadi

tempat masyarakat mengekspresikan pendapat tentang Pemilihan Presiden 2024 (Ali et al., 2022). Besarnya volume percakapan politik di Twitter menuntut adanya metode analisis otomatis untuk memahami kecenderungan opini publik (Fernández-Gil & Palau, 2023). Analisis sentimen merupakan salah satu teknik yang banyak

digunakan untuk mengekstraksi emosi dan opini dari data teks (Sari et al., 2022).

Metode ini memungkinkan peneliti memahami sentimen positif, negatif, dan netral secara sistematis dari data dalam jumlah besar (Ma'aly et al., 2024). Berbagai penelitian sebelumnya menunjukkan bahwa Twitter dapat merefleksikan dinamika opini politik masyarakat, terutama menjelang pemilu.

Namun, penelitian terkait analisis sentimen politik berbahasa Indonesia khususnya pada konteks Pemilu Presiden 2024 masih terbatas. Selain itu, belum banyak penelitian yang secara khusus mengevaluasi kinerja algoritma Naïve Bayes pada dataset berbahasa Indonesia bertipe informal seperti tweet (Ashbaugh & Zhang, 2024).

Penelitian ini bertujuan untuk: (1) menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen pada data Twitter terkait Pemilu Presiden 2024; (2) mengetahui persebaran opini publik; serta (3) mengevaluasi performansi model menggunakan metrik akurasi, presisi, dan recall.

Kebaruan penelitian ini terletak pada penggunaan alur KDD lengkap dan evaluasi klasifikasi sentimen berbahasa Indonesia yang bersumber dari percakapan politik aktual Pemilu 2024.

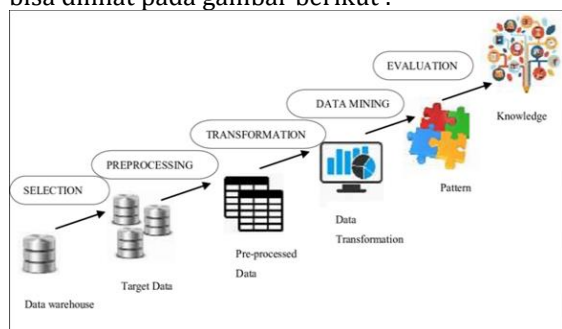
MATERIALS AND METHODS

Jenis Penelitian

Metode penelitian yang digunakan adalah pendekatan kuantitatif. Kuantitatif ini diolah dan dianalisis untuk menarik kesimpulan. Pendekatan ini mengutamakan pengukuran, pengolahan, dan analisis data secara sistematis dan obyektif untuk mendapatkan kesimpulan yang akurat dan dapat dipercaya.

Metode Penelitian

Tahapan dalam penelitian ini menggunakan metode analisa Knowledge Discovery In Database (KDD). Tahapan analisa data bisa dilihat pada gambar berikut :



Gambar 1 Tahapan Knowledge Discovery In Database (KDD)

Pada penelitian ini, penulis menggunakan metode atau tahapan KDD singkatan dari Knowledge Discovery In Database. Berikut merupakan tahapan - tahapan dan proses yang ada pada Knowledge Discovery In Database:

Data Selection

Di awal tahapan terdapat tahap selection, yang mana data selection adalah proses pemilihan subset data dari kumpulan data yang lebih besar berdasarkan kriteria atau parameter tertentu. Ini sering dilakukan untuk fokus pada aspek data tertentu atau untuk mengurangi jumlah data yang perlu dianalisis atau diproses. Pemilihan data mendahului praktik pengumpulan data yang sebenarnya. Data diambil dari Kaggle.com dengan kata kunci data pilpres tahun 2024. dari tanggal 22-23 Juli 2025 sebanyak 1336 twett.

Data Cleaning atau Pre-processing

Pembersihan data untuk proses mengidentifikasi bagian data yang tidak lengkap, salah, benar, tidak akurat atau hilang dan mengoreksi, mengganti atau menghapusnya. Cleansing penelitian ini akan menghapus @, #, karakter URL serta atribut yang hanya menggunakan teks. Operator yang digunakan adalah substitusi.

Transformation

Dalam tahap ini tidak menggunakan transformasi dikarenakan pada proses transformasi tidak ada nya perubahan pada hasil.

Data Mining

Data Mining proses mengubah teks yang tidak terstruktur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru. Pada tahap Riset ini menerapkan Naïve Bayes.

Pattern Evaluation

Pattern Evaluation didefinisikan sebagai mengidentifikasi model yang meningkatkan wawasan berdasarkan ukuran yang diberikan. Temukan tempat menarik untuk setiap model menggunakan ringkasan dan visualisasi agar informasi dapat dipahami oleh pengguna.

Sumber Data

Penelitian ini menggunakan data sekunder yang berasal dari dataset publik mengenai sentimen opini publik terkait Pemilihan Presiden 2024. Dataset tersebut diperoleh dari platform Kaggle.com.

Prosedur Penelitian

Proses penelitian ini meliputi pengumpulan data sekunder dari Twitter, prapemrosesan teks, penerapan algoritma Naïve Bayes untuk klasifikasi sentimen, serta evaluasi menggunakan akurasi, presisi, dan recall. Tahapan dilakukan secara berurutan, mulai dari pemilihan kata kunci, pengambilan data melalui Twitter atau dataset publik, pembersihan data, hingga analisis hasil. Dengan alur tersebut, penelitian diharapkan mampu menghasilkan analisis sentimen yang akurat dan memberikan gambaran menyeluruh tentang pandangan masyarakat terhadap Pemilihan Presiden Indonesia 2024.

Pengumpulan Data

Pengumpulan data dilakukan melalui observasi tidak langsung atau studi dokumentasi digital, yaitu dengan memanfaatkan dataset yang telah dipublikasikan oleh pihak lain. Proses dimulai dengan mengunduh dataset, kemudian dilanjutkan dengan pre-processing dan standarisasi, termasuk penyimpanan ke format Excel untuk memudahkan pemeriksaan, pembersihan data, dan penataan struktur kolom. Data yang telah divalidasi kemudian disimpan kembali dalam format akhir yang siap diproses menggunakan perangkat lunak statistik atau data mining.

Data Cleaning atau Pre-processing

Tahap pengolahan data pada penelitian ini dilakukan melalui rangkaian prapemrosesan yang dirancang untuk memastikan kualitas dan kesiapan dataset sebelum memasuki proses klasifikasi sentimen. Proses dimulai dengan membaca file CSV dari Kaggle yang berisi tweet terkait Pilpres 2024, kemudian dilanjutkan dengan pemilihan atribut penting seperti teks dan label sentimen guna mengurangi gangguan pada model (Putri & Nasruddin, 2023).

Setiap atribut selanjutnya diberikan peran yang sesuai agar struktur data dapat dikenali oleh algoritma pembelajaran mesin (Bria & Witanti, 2023). Data yang duplikat dihapus untuk mencegah bias analisis, sedangkan karakter atau simbol tidak lazim diganti demi menjaga konsistensi dataset (Zhou et al., 2023). Selain itu, nilai hilang diimputasi menggunakan metode yang relevan agar tidak mengganggu kinerja model. Secara keseluruhan, tahapan ini memastikan bahwa dataset berada dalam kondisi optimal untuk dianalisis menggunakan algoritma Naïve Bayes.

Klasifikasi Sentimen

Tahap klasifikasi sentimen dalam penelitian ini merupakan inti dari analisis data, di mana dataset yang telah melalui pra-pemrosesan

diolah menggunakan algoritma Naïve Bayes. Algoritma ini dipilih karena kemampuan yang efektif dalam menangani data teks berukuran besar serta ketahanannya terhadap keterbukaan kelas melalui pendekatan probabilistik yang menghitung peluang setiap data masuk ke kategori tertentu. Validitas penggunaan Naïve Bayes juga diperkuat oleh temuan (Widyassari et al., 2025), yang menunjukkan akurasi tinggi pada sentimen terkait isu politik, serta penelitian terdahulu, yang menegaskan konsistensi kinerjanya pada data Twitter berdasarkan metrik evaluasi seperti akurasi, presisi, dan recall. Dengan demikian, penerapan Naïve Bayes dalam penelitian ini didasarkan pada efisiensinya serta bukti empiris yang mendukung efektivitasnya dalam mengklasifikasikan opini publik

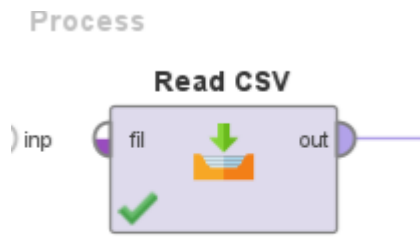
Model Performance

Tahap evaluasi kinerja model merupakan bagian penting dalam analisis sentimen karena menentukan efektivitas algoritma Naïve Bayes dalam mengklasifikasikan tweet terkait Pilpres 2024. Evaluasi dilakukan menggunakan tiga metrik utama Akurasi, Recall, dan Presisi yang secara bersama memberikan gambaran menyeluruh mengenai model yang halus, sensitif, dan reliabilitas. Akurasi digunakan untuk melihat proporsi prediksi yang benar dari seluruh data, sementara recall menilai kemampuan model dalam mendeteksi kelas positif secara tepat, yang sangat penting dalam mengungkap sentimen publik. Presisi mengukur tingkat kebenaran prediksi positif dan berperan penting dalam menjaga validitas analisis, terutama dalam konteks pengambilan keputusan strategi seperti politik. Melalui metrik ketiga ini, peneliti dapat menilai secara komprehensif model kinerja dalam menangani sentimen data terkait Pemilihan Presiden 2024.

RESULTS AND DISCUSSION

Data Selection

Membaca dataset dalam bentuk file excel, menggunakan operator Read Excel seperti tampak pada gambar dibawah ini.



Gambar 2 Parameters Read Excel

Parameter pada operator Read Excel, menggunakan parameter default. Dari hasil pembacaan operator Read Excel didapat informasi sebagai berikut. Tabel 1.

Tabel 1 Statistik dataset

No	Uraian	Keterangan
1	Special Attribute	1
2	Reguler Attribute	4
3	Attribute :	
4	Username	Polymial
5	Sentimen	Polymial
6	Confidence (Positive)	0
7	Confidence (Negativ)	0
8	Confidence (Netral)	0
9	Date	Polymial, Numerik
10	Text	Polymial
11	Length_Text	Numerik

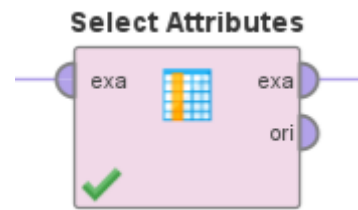
Data Selection ini didapat dari webeset pada Kaggle.com dengan operator search twitter dan kata kunci data tweet twitter Pilpres 2024. Pengambilan data dari tanggal 22 sampai 23 Juli 2025.

Gambar 3 Dataset Pilpres Indonesia 2024

Operator Select Attributes

Operator select attributes ini berfungsi untuk memilih data subset atribut yang akan digunakan dari ExampleSet dan menghapus atribut

lainnya yang tidak digunakan. Tampilan operator Select Attributes dapat dilihat pada gambar 2.



Gambar 4 Operator Select Attributes

Pada operator select attributes dilakukan proses selection terhadap atribut yang akan digunakan dalam penelitian dapat dilihat pada tabel 4.2

Tabel 2 Tabel Atribut Data

No	Atribut	Type
1	Username	Polymial
2	Sentimen	Polymial
3	Date	Polymial, Numerik
4	Text	Polymial
5	Length_Text	Numerik

Semua Atribut ini yang tdigunakan dalam penelitian ini adalah atribut Username, Sentimen, Date, Text, Length_Text. Atribut tersebut tidak mempengaruhi proses data mining yang dijadikan nilai untuk menghitung akurasi dan digunakan di dalam penentu pemilihan Presiden Republik Indonesia Tahun 2024. Hasil yang diperoleh dari proses operator select attributes ini dapat dilihat pada gambar 5.

Gambar 5 Hasil Select Attributes

Data Labeling

Data diberi label secara manual sebagai sentimen positif dan negatif. Label adalah cara yang bagus untuk mengomunikasikan informasi seperti nama lokasi. Pelabelan adalah bagian dari pembelajaran yang diawasi dengan supervised learning

adalah model pembelajaran mesin yang memeriksa data dengan tag atau target dimana evaluasi model akan didasarkan pada target tersebut.

Tabel 3 Label Proses

Tweets	Sentimen
Unggul Telak Dalam Debat Capres Anies Baswedan Memang Paling Layak Menjadi Presiden RI 2024 Pilihan Cerdas Anies Baswedan Anies Muhaimin 2024	Positif
Anies ³ / ₄ melanjutkan perjalanan hingga selamat sampai tujuan usai adanya insiden tersebut Anies Muhaimin 2024	Netral
Pendukung wowo amp asam sulfat tertawa angin tak punya KTP simak dg cermat video lengkap inipak aniesbaswedan yg asal bicara atau wowo yg gagal paham Yu ah kita simak Anies Presiden 2024 AniesMuhaimin2024	Negatif

Klasifikasi Algoritma Naïve Bayes

Proses ini dilakukan untuk mengukur dan mengevaluasi kinerja suatu algoritma. Pada tahap ini akan dilakukan randomisasi atau pengecekan data agar tidak terjadi pengelompokan data.



Gambar 6 Metode Naïve Bayes

Penilaian Kinerja Hasil Klasifikasi Algoritme Naive Bayes:

	true Positive	true Neutral	true Negative	class precision
pred. Positive	940	0	0	100.00%
pred. Neutral	0	186	0	100.00%
pred. Negative	0	0	159	100.00%
class recall	100.00%	100.00%	100.00%	

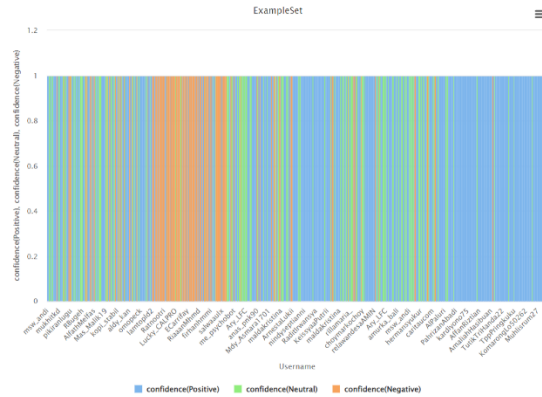
Gambar 2 Performance Naïve Bayes

Berdasarkan gambar 4.16 Hasil Performance tentang analisa sentimen Pemilihan Presiden Indonesia Tahun 2024 pada twitter yang menggunakan metode Naïve Bayes diperoleh hasil akurasi sebesar 100.00% dan micro average 100.00% dalam penentuan sentimen memiliki tingkat precision sebesar 100,00% dan recall sebesar 100.00%, dimana penjelasan tabel sebagai berikut :

Pertama, sebesar 100.00% untuk predicted class negative dengan True Negative (TN) sebanyak 159, True Positive (TP) sebanyak 940 dan True Neutral 186.

Kedua, sebesar 100.00% untuk predicted class positive dengan True Negative (TN) sebanyak 159, True Positive (TP) sebanyak 940.

Ketiga, sebesar 90,60% untuk predicted class netral dengan True Negative (TN) sebanyak 159, True Positive (TP) 940, True Netral (TN) sebanyak 186.



Gambar 3 Bagan Naïve Bayes

Pembahasan hasil analisis kualitatif dan/atau kuantitatif harus dilakukan dengan tujuan untuk memberikan jawaban atas permasalahan yang diteliti dalam makalah. Pada bagian ini, seluruh kegiatan yang telah dilakukan dalam penelitian harus diuraikan secara lengkap. Mulai dari konsep penelitian, perancangan penelitian, hipotesis (bila ada), percobaan, pengamatan data, dan hasil dari data pengamatan yang telah dilakukan.

Pattern Evaluation

Pattern evaluation proses evaluasi pola dimulai dengan mengumpulkan data dari berbagai sumber dan memperhatikan setiap detailnya. Data yang telah dikumpulkan kemudian dianalisis untuk mencari pola dan tren yang menonjol. Setelah itu, pola dan tren tersebut dievaluasi untuk menentukan apakah mereka memiliki nilai yang signifikan atau tidak. Evaluasi dilakukan setelah proses data dilakukan dengan Naïve Bayes. menghasilkan berapa banyak sentimen positive dan negative. Confusion Matrix digunakan dalam evaluasi ini karena berfungsi sebagai alat evaluasi dalam memperkirakan objek yang benar (True) dan salah (False) pada model klasifikasi. Matrix sendiri digunakan dalam membandingkan hasil klasifikasi dengan data yang sebenarnya.

Sumber rumus Confusion Matrix (Sholekha et al., 2022)

Tabel 3 Tabel Confusion Matrix

Predicted Class	Actual Class	
	Class 1	Class 2
Class 1	True Positive (TP)	False Negative (FN)
Class 2	False Positive (FP)	True Negative (TN)

Deskripsi:

TP : True Positive

FN : False Negative

FP : False Positive

TN : True Negative

Rumus yang digunakan untuk menghitung akurasi adalah:

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

Gambar 4 Menghitung accuracy

Rumus yang digunakan untuk menghitung nilai presisi:

$$Precision = \frac{TP}{TP + FP}$$

Gambar 5 Menghitung Nilai Precision

Rumus yang digunakan dalam menghitung nilai Recall adalah:

$$Recall = \frac{TP}{TP + FN}$$

Gambar 6 Menghitung nilai recall

Hasil dari algoritme yang digunakan menunjukkan akurasi dari Naïve Bayes. Menghasilkan accuracy sebesar 100%, precision 100% dan recall 100% pada algoritme Naïve Bayes. Sehingga dapat disimpulkan dari hasil diatas bahwa pada penelitian ini klasifikasi terbaik dalam menghasilkan tingkat akurasi.

CONCLUSION

Melalui metode Naïve Bayes, data-data dari Twitter tersebut diproses untuk menghasilkan informasi tentang sentimen masyarakat terkait pemilihan Presiden Indonesia Tahun 2024. Metode ini didasarkan pada teori probabilitas dan menggunakan klasifikasi untuk menentukan sentimen yang diungkapkan oleh masyarakat dalam setiap tweet. melalui media sosial Twitter yang mengklasifikasikannya menjadi positif dan negatif, hasil akurasi yang didapat pada algoritme Naïve Bayes sebesar Accuracy 100%, precision 100% dan recall 100%.

Peneliti menggunakan metode klasifikasi metode Naïve Bayes karena metode yang tepat digunakan dengan hasil akurasi baik dalam mengklasifikasikan data mining. Dalam studi selanjutnya, lebih banyak set data harus digunakan daripada studi ini untuk meningkatkan informasi tentang Pemilihan Presiden Indonesia Tahun 2024 dan meningkatkan presisi dibandingkan studi sebelumnya. Gunakan metode lain untuk memverifikasi bahwa proses penelitian itu akurat. Pencarian tambahan juga dapat mencari sumber data lain seperti Facebook, Instagram, dan YouTube.

REFERENCE

- Adib, K., Handayani, M. R., Yuniarti, W. D., & Umam, K. (2024). Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM. *SINTECH (Science and Information Technology) Journal*, 7(2), 80–91. <https://doi.org/10.31598/sintechjournal.v7i2.1581>
- Ali, R. H., Pinto, G., Lawrie, E., & Linstead, E. (2022). Analisis sentimen skala besar terhadap twit terkait pemilihan presiden AS 2020. *Jurnal Big Data*, 9, 79. <https://doi.org/https://doi.org/10.1186/s40537-022-00633-z>
- Ashbaugh, L., & Zhang, Y. (2024). A comparative study of sentiment analysis on customer reviews using machine learning and deep learning. *Computers*, 13(12), 340. <https://doi.org/10.3390/computers13120340>
- Bria, N. A. R., & Witanti, A. (2023). Analisis Sentimen Masyarakat Indonesia Menggunakan Algoritma Support Vector Machine Tentang Pilpres 2024. *JATI (Jurnal Mahasiswa Teknik Informatika)*. <https://ejournal.itn.ac.id/index.php/jati/article/view/8184>
- Fernández-Gil, S., & Palau, R. (2023). Wacana media sosial dan pengaruhnya terhadap keputusan pemilih: analisis sentimen dalam twit selama periode pemilu. *Analisis Dan Penambangan Jaringan Sosial*, 13, 46. <https://doi.org/https://doi.org/10.1007/s13278-023-01048-1>
- Ma'aly, A. N., Pramesti, D., Fathurahman, A. D., & Fakhurroja, H. (2024). Eksplorasi analisis sentimen untuk pemilihan presiden Indonesia melalui ulasan daring menggunakan klasifikasi multi-label dengan algoritma pembelajaran mendalam. *Informasi*, 15(11), 705.

- <https://doi.org/https://doi.org/10.3390/info15110705>
- Putri, N., & Nasruddin. (2023). Analisis Sentimen Twitter terhadap Pemilu 2024 Menggunakan Metode Naïve Bayes. *Seminar Nasional Informatika Dan Sains (SENASAINS)*.
- Sari, D. K., Kumorotomo, W., & Kurnia, N. (2022). Struktur penyampaian pesan nasionalisme di Twitter dalam konteks netizen Indonesia. *Analisis Dan Penambangan Jaringan Sosial*, 12, 173.
<https://doi.org/https://doi.org/10.1007/s13278-022-01006-3>
- Sholekha, I., Faqih, A., & Bahtiar, A. (2022). Sentiment Analysis of Public Opinion Covid-19 Vaccine Using Naïve Bayes and Random Forest Methods. *Jurnal Teknik Informatika*, 15(1), 34–43.
<https://doi.org/10.15408/jti.v15i1.24847>
- Widyassari, A. P., Salsabilla, D., & Amrozi, M. A. (2025). Analisis Sentimen Publik di Twitter Terhadap Pelantikan Presiden Prabowo Menggunakan Algoritma Naïve Bayes. *Jurnal Ilmiah NERO*, 10(1), 13–24.
<https://journal.trunojoyo.ac.id/nero/article/view/28701>
- Zhou, F., Wu, Y., & Lin, C. (2023). Imputasi Nilai Hilang untuk Prapemrosesan Data Teks dalam Pembelajaran Mesin. *Transaksi ACM Pada Ilmu Data*, 5(2), 22–39.