



## ANALISIS SENTIMEN TERHADAP KEYWORD TERTENTU PADA PLATFORM MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAIVE BAYES

David<sup>1</sup>, Agus Ramdhani Nugraha<sup>2</sup>

Program Studi Teknik Informatika, STMIK DCI

Email : [Davidsaja317@gmail.com](mailto:Davidsaja317@gmail.com)<sup>1</sup>, Email : [agus.tsm78@gmail.com](mailto:agus.tsm78@gmail.com)<sup>2</sup>

### ABSTRAK

Perkembangan teknologi informasi telah mendorong peningkatan interaksi publik di media sosial seperti X (sebelumnya Twitter), tempat berbagai opini dibagikan. Analisis sentimen menjadi pendekatan yang tepat untuk memahami persepsi publik, terutama terhadap isu tertentu yang dicari melalui kata kunci. Penelitian ini mengimplementasikan algoritma Naive Bayes yang dipilih karena kesederhanaan, efisiensi, dan kesesuaian untuk klasifikasi teks. Proses yang dilakukan meliputi *crawling* data berdasarkan *keyword*, *preprocessing* teks, ekstraksi fitur dengan *Term Frequency – Inverse Document Frequency* (TF-IDF), penyeimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), pelatihan model Naive Bayes, dan penyajian hasil dalam bentuk tabel serta grafik. Sistem juga dilengkapi dengan deteksi *bot* dan penanganan ketidakseimbangan kelas untuk meningkatkan akurasi. Model berhasil mengklasifikasikan sentimen dengan akurasi sekitar 72%. Meskipun performa untuk sentimen netral masih terbatas akibat data latih netral yang sedikit, sistem ini dapat membantu dalam memahami opini publik dan berpotensi dikembangkan lebih lanjut.

**Kata Kunci:** Analisis Sentimen, Keyword, X

### I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi di era digital saat ini memberikan pengaruh pada berbagai aspek kehidupan manusia. Salah satu teknologi yang semakin banyak digunakan adalah kecerdasan buatan (AI), yang mampu memberikan solusi otomatisasi, analisis data, dan pengambilan keputusan

berdasarkan data yang akurat (Zhang & Lu, 2021; Vasuki et al., 2024). Implementasi AI dapat ditemukan dalam pengelolaan opini publik di media sosial (Vasuki et al., 2024).

Platform media sosial X (sebelumnya Twitter) merupakan salah satu media sosial populer sebagai tempat berbagi pendapat dan mengekspresikan

opini tentang berbagai hal (Sallam et al., 2025). Dalam pencarian teks di media sosial X, kata kunci atau keyword sangat penting untuk menemukan konten yang relevan (Rahayu & Noeraida, 2022).

Analisis sentimen menjadi aspek krusial untuk memahami opini publik yang tersebar luas di platform digital (Wankhade et al., 2022). Dengan analisis sentimen, perusahaan, pemerintah, atau individu dapat mengetahui persepsi pengguna, yang kemudian dapat dimanfaatkan untuk membantu proses pengambilan keputusan (Wankhade et al., 2022). Dalam analisis sentimen, data yang sudah diberi label atau "anotasi" dengan sentimen sangat diperlukan. Namun, proses pelabelan data ini seringkali dilakukan secara manual, memakan waktu, biaya, tenaga, dan rawan kesalahan (Qureshi et al., 2022; Alzubaidi et al., 2023).

Naive Bayes adalah salah satu algoritma data mining yang populer, mudah digunakan, dan sering digunakan dalam klasifikasi teks (Chen et al., 2020; Yew et al., 2021). Algoritma ini memiliki keunggulan dalam memberikan prediksi yang mudah diinterpretasikan dengan sedikit parameter yang perlu diubah, sehingga menjadi pilihan populer untuk berbagai tugas klasifikasi, terutama pada data teks (Yew et al., 2021).

Berdasarkan uraian di atas, penelitian ini berjudul "Analisis Sentimen Terhadap Keyword Tertentu Pada Platform Media Sosial X Menggunakan Algoritma Naive Bayes", yang diharapkan dapat mengklasifikasi sentimen tanpa

harus melakukan anotasi data latih secara manual dengan harapan dapat mengklasifikasi sentimen apapun tanpa harus melakukan anotasi data latih secara manual.

Rumusan masalah dalam penelitian ini adalah bagaimana mengimplementasikan algoritma Naive Bayes untuk menganalisis sentimen terhadap *keyword* tertentu di media sosial X, dan seberapa akurat algoritma tersebut dalam menganalisis sentimen terhadap dataset dengan *keyword* tertentu.

Penelitian ini bermaksud melakukan analisis sentimen dengan Naive Bayes sebagai metode pengklasifikasian sentimen pengguna, dengan tujuan menghasilkan model analisis sentimen yang dapat memahami opini publik berdasarkan dataset dari media sosial X, meningkatkan kemampuan analisis sentimen, dan mengetahui kecenderungan opini (positif, negatif, atau netral) terhadap topik yang direpresentasikan oleh *keyword* tertentu.

## II. LANDASAN TEORI

Suatu pendapat melalui sikap, perasaan, dan pemikiran seseorang dapat disebut sebagai sentimen. Klasifikasi analisis sentimen terbagi menjadi tiga kategori utama: positif, negatif, dan netral, yang dapat diidentifikasi melalui pembelajaran mendalam (Parikh dan Shah, 2021; Adagele dan Gupta, 2023). Analisis sentimen merupakan studi berbasis komputasi untuk memahami opini dan emosi masyarakat terhadap sesuatu. Penelitian di bidang ini telah

berkembang pesat dan banyak dilakukan selama dekade terakhir. Berbagai fitur, algoritma, serta koleksi data yang dimanfaatkan dalam model analisis sentimen telah disusun dan didokumentasikan dengan baik (Lighthart, Catal, dan Tekinerdogan, 2021). Analisis sentimen, yang sering disebut sebagai *emotion Artificial Intelligence* (AI) atau *opinion mining*, merupakan bidang studi yang berfokus pada pemahaman makna sentimen dalam informasi yang tidak terstruktur. Bidang ini telah berkembang pesat karena kemajuan cepat dalam teknik pembelajaran mendalam serta melimpahnya dataset digital yang tersedia. Keadaan emosi dari pemberi opini diambil dari teks yang ada melalui penerapan teknik pembelajaran mesin, dan pemrosesan bahasa alami (Raghunathan, dan Kandasamy, 2023).

Tahapan analisis sentimen :

### 1. Preprocessing Data

Tahap ini penting untuk membersihkan dan memilah data yang tidak relevan (Manik et al., 2021). Proses ini dilakukan melalui fungsi pipeline preprocess text dan menghasilkan bentuk teks yang bersih. Langkah-langkahnya meliputi:

1. Cleaning: Meminimalkan nilai salah atau tidak valid dalam dataset (Pilowsky et al., 2024).
2. Case folding: Mengubah huruf kapital menjadi huruf kecil (Prasetyo et al., 2023).
3. Menghilangkan noise: Menghapus angka, simbol, atau

elemen tidak relevan lainnya (Akbar et al., 2024).

4. Spelling Normalization: Memperbaiki kesalahan pengejaan, singkatan, atau bahasa gaul (Manik et al., 2021).
5. Tokenizing: Memecah kalimat menjadi kata (Manik et al., 2021).
6. Stopword Removal: Menghapus kata yang sering muncul namun kurang bermakna (Taufiqurrahman et al., 2021).
7. Stemming atau Lematisasi: Mengubah kata ke bentuk dasarnya (Akbar et al., 2024).

### 2. Ekstraksi Fitur dengan TF-IDF

Teks diubah menjadi representasi numerik menggunakan metode TF-IDF. Term Frequency (TF) adalah jumlah kemunculan sebuah term dalam suatu dokumen. Inverse Document Frequency (IDF) menunjukkan seberapa jarang sebuah kata muncul di seluruh dokumen dalam koleksi (Manik et al., 2021). Rumus TF-IDF :

$$IDF = \log\left(\frac{D}{df_t}\right)$$

Dimana :

$D$  : Jumlah semua dokumen  
 $df_t$  : Total seluruh dokumen yang mengandung term ( $t$ )

$$W_{t,d} = TF \times IDF$$

Dimana  $W_{t,d}$  adalah bobot dari kata  $t$  pada dokumen  $d$

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Total kata dalam dokumen } d}$$

Media sosial X atau yang dulu dikenal sebagai media sosial twitter

merujuk pada platform berbasis internet seperti blog dan jejaring sosial yang memungkinkan pengguna untuk berpartisipasi, berbagi, serta menciptakan informasi dengan mudah. Media sosial ini mempermudah penggunanya untuk membuat atau mengunggah *tweet* berupa pesan singkat yang dapat dibagikan oleh siapa saja di dalam aplikasi tersebut. Platform ini hanya mendukung konten dalam bentuk teks, gambar, dan video (Nurhapiyah, Steviana, dan Raharja, 2024). Media sosial X sering digunakan sebagai sumber data opini publik dan sentimen untuk analisis, di mana data tersebut dapat dimanfaatkan untuk memahami pandangan masyarakat terhadap suatu topik (Safitri, Umaidah, dan Maulana, 2023).

*Keyword* atau kata kunci adalah kata atau frasa yang dimasukkan oleh pengguna saat melakukan pencarian pada kolom pencarian (Akhlak, 2022). Dengan menggunakan kata kunci seseorang dapat mendapatkan informasi sesuai dengan apa yang diinginkan atau dicari. Disisi lain *keyword* digunakan oleh seorang penulis agar orang lain dapat menemukan tulisannya dengan mudah (Rahayu & Noeraida, 2022)

Naive Bayes adalah metode klasifikasi berbasis statistik yang digunakan untuk memperkirakan peluang suatu data termasuk ke dalam kelas tertentu. Dalam konteks ini, digunakan dataset sebagai dasar untuk menghitung

probabilitas kelas berdasarkan data pelatihan. Naive Bayes

*Classifier* bekerja dengan asumsi bahwa setiap atribut (variabel) dalam data saling berdiri sendiri atau tidak saling memengaruhi (Martantoh & Yunih, 2022). Algoritma klasifikasi Naive Bayes memiliki struktur algoritma yang sederhana sehingga banyak digunakan dalam bidang analisis *big data* karena dapat bekerja dengan cepat dan efisien. Keuntungan lainnya dari penggunaan Naive Bayes adalah ia hanya perlu memperkirakan parameter yang diperlukan berdasarkan sejumlah kecil data pelatihan (Chen, et al. 2021). Proses perhitungannya dimulai dengan mengambil data uji, kemudian dilanjutkan dengan menghitung probabilitas masing-masing kriteria berdasarkan data pelatihan. Setelah itu, dihitung juga peluang tiap fitur dari data uji dengan mengacu pada data pelatihan. Terakhir, hasil tersebut dikalikan dengan nilai  $P$  untuk setiap kelas guna menentukan kelas dengan probabilitas tertinggi (Martantoh & Yunih, 2022).

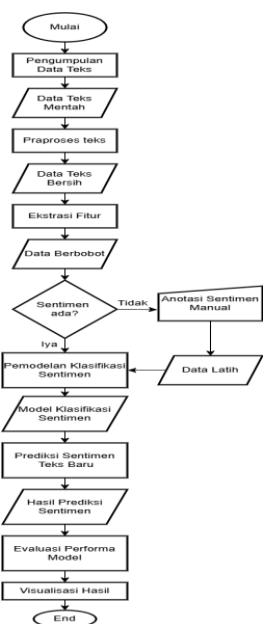
$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

1.  $P(H|X)$ : Probabilitas bahwa hipotesis benar setelah melihat data (*posterior*).
2.  $P(H)$ : Probabilitas hipotesis sebelum melihat data (*prior*).
3.  $P(X|H)$ : Probabilitas data jika hipotesis benar.
4.  $P(X)$ : Probabilitas data secara keseluruhan

### III. ANALISIS MASALAH

Sistem analisis sentimen untuk data dari media sosial X umumnya melibatkan pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, klasifikasi sentimen, dan evaluasi performa model. Meskipun demikian, sistem yang ada menghadapi beberapa tantangan:

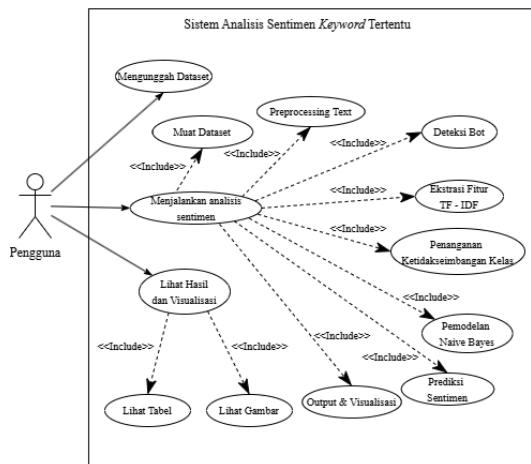
1. Penanganan data masif: Keterbatasan dalam anotasi data latih secara manual yang memakan waktu, biaya, tenaga, dan rawan kesalahan.
2. Kurangnya identifikasi akun bot: Akun *bot* dapat memengaruhi hasil analisis.
3. Tidak ada penanganan ketidakseimbangan kelas: Jika data latih tidak seimbang terhadap kelas tertentu, model bisa bias



Gambar 1. Flowchart Analisis Sentimen Umum

Sistem analisis sentimen yang lama umumnya melakukan tahapan pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, klasifikasi sentimen, dan evaluasi performa model. Analisis menunjukkan keterbatasan dalam penanganan data masif karena anotasi data latih masih manual, yang memakan waktu dan biaya, serta kurangnya identifikasi akun *bot*.

### IV. PERANCANGAN SISTEM



Gambar 2. Use Case Diagram Sistem Analisis Sentimen Keyword Tertentu

Alur utama sistem dimulai dari pemuatan dataset hasil crawling berdasarkan keyword tertentu. Dataset kemudian diproses melalui tahap preprocessing untuk membersihkan teks dari elemen seperti URL, emoji, dan kata-kata tidak bermakna, serta menerapkan lematisasi dan penanganan negasi (Manik et al., 2021). Proses ini menghasilkan teks yang siap dikonversi menjadi representasi numerik menggunakan metode Term Frequency – Inverse Document Frequency (TF-IDF) (Manik et al., 2021). Model Naive Bayes dilatih menggunakan dataset yang

terdapat kelas untuk positif, negatif, dan netral dengan penanganan ketidakseimbangan kelas dilakukan dengan metode Synthetic Minority Oversampling Technique (SMOTE). Sistem juga menerapkan deteksi bot, dengan mengecek duplikasi teks dalam 30 menit sebagai indikasi akun otomatis. Hasil klasifikasi kemudian divisualisasikan dalam bentuk tabel dan grafik, seperti confusion matrix, word cloud, dan distribusi sentimen.



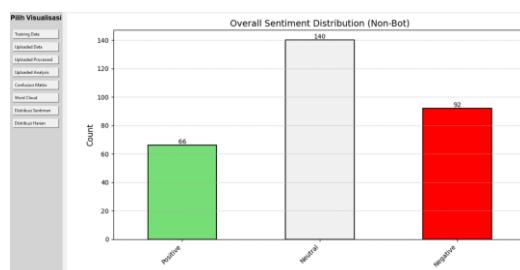
Gambar 3. Flowchart Diagram Sistem Analisis Sentimen Keyword Tertentu

1. Memuat Dataset X: Data mentah diperoleh melalui *crawling* dari media sosial X menggunakan *auth token* dalam format CSV.
2. Praproses Teks: Membersihkan data dari *noise* seperti tautan, tanda baca, emoji, singkatan, dan kesalahan ejaan. Meliputi *cleaning*, *case folding*, menghilangkan *noise*, *spelling normalization*, *tokenizing*, *stopword removal*, dan *stemming* atau lematisasi.
3. Deteksi Bot: Menerapkan mekanisme deteksi *bot* berbasis kemiripan teks dan waktu (*created\_at*). Data duplikat dalam 30 menit akan dikeluarkan.
4. Representasi Fitur dengan TF-IDF: Teks diubah menjadi representasi numerik menggunakan TF-IDF, memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan dan jarang munculnya kata di seluruh koleksi dokumen.
5. Penanganan Ketidakseimbangan Data dengan SMOTE: Menerapkan teknik SMOTE pada data latih untuk menyeimbangkan distribusi kelas sentimen, dengan memilih sampel minoritas dan membuat data sintetis baru.
6. Pelatihan Model Klasifikasi Naive Bayes: Model dilatih menggunakan data latih yang telah diproses, divektorisasi, dan diseimbangkan. Parameter

7. *smoothing (alpha)* 1.0 digunakan untuk menghindari probabilitas nol.
8. Evaluasi Model: Mengukur performa model berdasarkan *accuracy*, *precision*, dan *recall*. Laporan evaluasi disajikan dalam

## V. IMPLEMENTASI

Sistem analisis sentimen ini dirancang untuk mengklasifikasikan opini pengguna dari media sosial X berdasarkan *keyword* tertentu ke dalam kategori sentimen positif, negatif, atau netral. Sistem dikembangkan menggunakan bahasa pemrograman Python dengan antarmuka grafis berbasis *tkinter*, serta menerapkan algoritma Naive Bayes untuk klasifikasi sentimen.

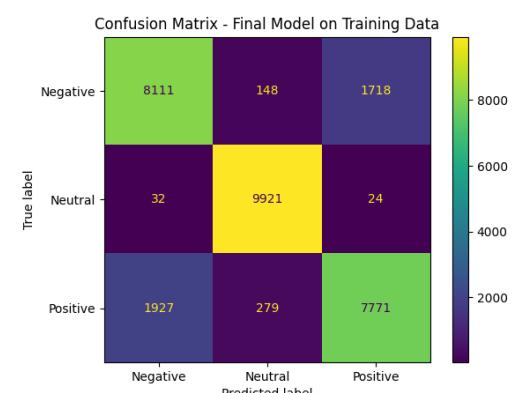


Gambar 4. Tampilan Visualisasi Hasil Performa model dievaluasi dengan mengukur nilai performa berdasarkan *accuracy*, *precision*, dan *recall*. Laporan evaluasi disajikan dalam bentuk *confusion matrix* dan *classification report*

	precision	recall	f1-score	support
Negative	0.74	0.73	0.73	1995
Neutral	0.14	0.57	0.23	28
Positive	0.74	0.72	0.73	1995
accuracy			0.72	4018
macro avg	0.54	0.67	0.56	4018
weighted avg	0.73	0.72	0.73	4018

Gambar 5. Classification Report

model mencapai akurasi keseluruhan sekitar 72% dengan performa yang cukup baik untuk kelas sentimen negatif dan positif (F1-score masing-masing 0.73). Namun, model menunjukkan kesulitan signifikan dalam mengklasifikasikan sentimen netral, yang tercermin dari presisi yang sangat rendah (0.14) dan F1-score yang rendah (0.23), kemungkinan besar disebabkan oleh jumlah sampel (*support*) yang sangat terbatas untuk kelas netral (hanya 28 sampel) dalam set data validasi ini, yang membuat model sulit untuk belajar dan menggeneralisasi pola untuk kelas tersebut secara efektif. Sehingga perlu ditambahkan konfigurasi khusus untuk kelas netral, dimana ketika nilai *confidence* dari kelas positif atau negatif tidak terlalu meyakinkan maka teks yang dianalisis akan diberi kelas netral.



Gambar 6. Confusion Matrix

*Confusion matrix* secara visual merangkum performa model klasifikasi sentimen final pada data latihnya sendiri, yang telah melalui proses penyeimbangan kelas menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Sumbu vertikal (*True label*) merepresentasikan

kelas sentimen aktual dari setiap sampel data, sedangkan sumbu horizontal (*Predicted label*) menunjukkan kelas sentimen yang diprediksi oleh model. Nilai-nilai di sepanjang diagonal utama dari kiri atas ke kanan bawah menunjukkan jumlah prediksi yang benar: model berhasil mengidentifikasi 8111 sampel sebagai negatif, 9921 sampel sebagai netral, dan 7771 sampel sebagai positif, sesuai dengan label aslinya. Angka-angka di luar diagonal utama (*off-diagonal*) merepresentasikan kesalahan klasifikasi. Sebagai contoh, 1718 sampel yang sebenarnya negatif keliru diprediksi sebagai positif, dan 1927 sampel yang sebenarnya positif keliru diprediksi sebagai negatif, mengindikasikan adanya beberapa kesulitan dalam membedakan secara sempurna antara kedua polaritas ini. Untuk kelas netral, model salah mengklasifikasikan 32 sampel netral sebagai negatif dan 24 sampel netral sebagai positif. Sebaliknya, 148 sampel negatif dan 279 sampel positif juga keliru diklasifikasikan sebagai netral. Secara keseluruhan, tingginya angka pada diagonal utama dibandingkan angka *off-diagonal* menunjukkan bahwa model memiliki kecocokan yang sangat baik dengan data latih yang telah diseimbangkan, meskipun perlu dicatat bahwa performa ini pada data latih mungkin tidak sepenuhnya mencerminkan kemampuan generalisasi model pada data baru yang tidak terlihat

## VI. KESIMPULAN

Berdasarkan hasil penelitian dan implementasi sistem analisis sentimen

terhadap keyword tertentu di platform media sosial X, dapat disimpulkan bahwa :

1. Penelitian ini berhasil menghasilkan model analisis sentimen berbasis algoritma Naive Bayes yang mampu mengklasifikasikan opini publik pada media sosial X berdasarkan keyword tertentu. Model dilengkapi dengan tahap preprocessing data, ekstraksi fitur TF-IDF, penyeimbangan kelas dengan SMOTE, dan deteksi bot. Hasil evaluasi menunjukkan akurasi sekitar 72%, dengan performa baik untuk kelas positif dan negatif namun masih perlu ditingkatkan untuk kelas netral. Ini membuktikan bahwa pendekatan Naive Bayes efektif dan efisien digunakan untuk memahami persepsi publik pada platform media sosial berbasis teks.
2. Sistem yang dikembangkan meningkatkan kemampuan analisis sentimen dalam penerapan algoritma Naive Bayes untuk mengidentifikasi sentimen positif, negatif, atau netral untuk semua jenis dataset berdasarkan keyword tertentu. Proses pengolahan data mentah hingga klasifikasi akhir menunjukkan bahwa algoritma Naive Bayes dapat menangkap pola sentimen dengan baik pada data yang sudah melalui tahap preprocessing. Pendekatan ini dapat dijadikan dasar pengembangan sistem analisis sentimen yang lebih luas. Beberapa saran yang dapat diberikan untuk pengembangan sistem di masa mendatang antara lain:

1. Menambahkan *dataset* netral sehingga model lebih mampu mengenali sentimen netral.
2. Mengembangkan sistem agar mampu menganalisis sentimen dari berbagai bahasa beserta data latih dari tiap bahasa.
3. Menerapkan algoritma lain seperti *Support Vector Machine* (SVM) atau *Random Forest* dapat dibandingkan untuk mengevaluasi keunggulan relatif dari algoritma Naive Bayes yang digunakan.
4. Mengembangkan sistem dalam versi *web* agar lebih mudah diakses oleh pengguna.

## VII. DAFTAR PUSTAKA

- Adagele, S. S., Gupta, P. (2023). Aspect based sentiment analysis using various supervised classification techniques: an overview. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(6), 204-210.
- Akbar, Y., Regita, A. N. H., Sugiyono, Wahyudi, T. (2024). Analisa sentimen pada media sosial "x" pencarian keyword chatgpt menggunakan algoritma K-Nearest Neighbors (KNN). *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 5(3), 3291 - 3305.
- Akhhlak, M. L. M., (2022). Otomatisasi Riset Kata Kunci. *Jurnal Transformasi dan Pengembangan IPTEK*, 18(2), 12-25.
- Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaria, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, 10(46), 1-86.
- Amalia, B. S., Umaidah, Y., Mayasari, R. (2021). Analisis sentimen review pelanggan restoran menggunakan algoritma Support Vector Machine dan K-Nearest Neighbor. *SITEKIN: Jurnal Sains, Teknologi dan Industri*, 19(1), 28-34.
- Chen, H., Hu, S., Hu, R., Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 30(2021), 1-12.
- Chen, S., Webb, G. I., Liu, L., Ma, X., 2020, A Novel Selective Naïve Bayes Algorithm, *Knowledge-Based Systems*, ISSN: 1950-7051, Vol. 192, 1-12.
- Kirana, M. C., Perkasa, N. P., Lubis, M. Z. Fani, M. (2019). Visualisasi kualitas penyebaran informasi gempa bumi di Indonesia menggunakan Twitter. *Journal Appl. Informatic Computer*, 3(1), 23-32.
- Lighthart, A., Catal, C., Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 54, 4997-5053.
- Manik, G., Ernawati, I., Nurlaili, I. (2021). Analisis sentimen pada review pengguna e-commerce bidang pangan menggunakan metode Support Vector Machine (studi kasus: review sayurbox dan tanihub pada google play). *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2(2), 64-74.

- Martantoh, Eko., Yanih, N. (2022). Implementasi metode Naive Bayes untuk klasifikasi karakteristik kepribadian siswa di sekolah MTS Darussa'adah menggunakan PHP MySQL. *Jurnal Teknologi Sistem Informasi*, 3(2), 166-175.
- Nurhipipah, A., Steviana, S., & Raharja, R. M. (2024). Strategi kampanye di era digital: dengan memanfaatkan aplikasi X dalam Pemilu 2024. *ALADALAH: Jurnal Politik, Sosial, Hukum Dan Humaniora*, 2(3), 272–281.
- Parikh, S. M. & Shah, M. K. (2021). Classification approach for sentiment analysis using Machine Learning. *Applications of Artificial Neural Networks for Nonlinear Data*, 94-115.
- Pilowsky, J. K., Elliott, R., Roche, M. A. (2024). Data cleaning for clinician researchers: Application and explanation of a data-quality framework. *Australian critical care*, 37, 827-833.
- Prasetyo, S. D., Hilabi, S. S., Nurapriani, F. (2023). Analisis sentimen relokasi ibukota Nusantara menggunakan algoritma Naive Bayes dan KNN. *Komtekinfos*, 10(1), 1-7.
- Putra, I. M. A. A. D., Sunarya, I. M. G., Gunadi, I. G. A., (2024). Perbandingan algoritma Naive Bayes berbasis Feature Selection Gain Ratio dengan Naive Bayes kovensional dalam prediksi komplikasi hipertensi. *JTIM : Jurnal Teknologi Informasi dan Multimedia*, 6(1), 37-49.
- Putri, D. D., Nama, G. F., Sulistiono, W. E., (2022). Analisis sentimen kinerja Dewan Perwakilan Rakyat (DPR) pada Twitter menggunakan metode Naive Bayes classifier. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 10(1), 34-40.
- Qureshi, M. A., Asif, M., Hassan, M. F., Mustafa, G., Ehsan, M. K., Ali, A., Sajid, U. (2022). A Novel Auto-Annotation Technique for Aspect Level Sentiment Analysis. *Computer, Materials & Continua*, 70(3), 4987-5004.
- Raghunathan, N., Kandasamy, S., (2023), Challenges and issues in sentiment analysis: a comprehensive survey, *IEEE Access*, 2023.
- Rahayu, R. N., Noeraida, (2022). Kesesuaian Kata Kunci Jurnal Reinwardtia dengan Tesaurus Agrovoc. *Al-Maktabah*, 21(2), 26-37.
- Romzi, M., Kurniawan, B. (2020). Pembelajaran Pemrograman Python dengan Pendekatan Logika Algoritma. *Jurnal Teknik Informatika Mahakarya*, 3(2):37-44.
- Safitri, T., Umaidah, Y., Maulana, I. (2023). Analisis sentimen pengguna Twitter terhadap BTS menggunakan algoritma Support Vector Machine. *Journal of Applied Informatics and Computing (JAIC)*, 7(1), 34-41.
- Sallam, M., Al-Mahzoum, K., Sallam, M., Mijwil, M. M., 2025, DeepSeek: Is it the End of Generative AI Monopoly or the Mark of the Impending Doomsday?, *Mesopotamian Journal of Big Data*, ISSN: 2958-6453, Vol. 2025. 2025.
- Santoso, J. T., Migunani. (2021). *Desain & Analisis Sistem Berorientasi Obyek dengan UML*. Semarang: Yayasan Prima Agus Teknik.