

Perbandingan Kinerja Naive Bayes dan Decision Tree untuk Analisis Sentimen Ulasan Produk Online di Shopee dan Tokopedia

Rolinus Tomi Nagur¹, Diyan Alamsyah², Malfi Garry Tauran³, Yamin Nuryamin⁴, Ade Priyatna⁵

^{1,2,3} Mahasiswa, Teknologi Informasi Universitas Bina Sarana Informatika,

^{4,5} Dosen, Teknologi Informasi Universitas Bina Sarana Informatika.

¹rolinusnagur2000@gmail.com, ²diyanalamsyah03@gmail.com, ³tauranmalfi@gmail.com, ⁴yamin.yny.@bsi.ac.id,

⁵ade.aeq@bsi.ac.id

Article Info

Article history:

Received November 22, 2025

Accepted December 02, 2025

Published January 1, 2026

Kata Kunci:

Analisis Sentimen

Naïve Bayes

Decision Tree

Text Preprocessing,

Ulasan Produk E-commerce

ABSTRAK

Penelitian ini membandingkan kinerja algoritma Naïve Bayes dan Decision Tree untuk analisis sentimen ulasan produk pada platform e-commerce Shopee dan Tokopedia. Dataset yang digunakan berjumlah 46.086 ulasan yang telah melalui tahap pembersihan data dan preprocessing teks, meliputi penghapusan data kosong dan duplikat, case folding, pembersihan karakter, tokenisasi, serta penghapusan stopword. Label sentimen diturunkan dari rating bintang dan dibagi menjadi tiga kelas: positif, netral, dan negatif. Ciri teks direpresentasikan menggunakan TF-IDF, sedangkan evaluasi kinerja model dilakukan dengan metrik akurasi, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa Naïve Bayes menghasilkan akurasi sekitar 70,85%, lebih tinggi dibandingkan Decision Tree dengan akurasi 67,41%. Kedua model cukup baik dalam mengklasifikasikan sentimen positif, namun masih lemah pada kelas netral karena ketidakseimbangan data dan ekspresi bahasa yang ambigu. Temuan ini menunjukkan bahwa Naïve Bayes lebih sesuai digunakan sebagai dasar pengembangan sistem analisis sentimen ulasan produk e-commerce. Penelitian ini memberikan referensi bagi pelaku industri dalam memanfaatkan ulasan.



Corresponding Author:

Rolinus Tomi Nagur

Teknologi Informasi Universitas Bina Sarana Informatika

Email: *rolinusnagur2000@gmail.com

1. PENDAHULUAN

E-commerce adalah suatu kegiatan atau proses transaksi jual beli barang atau jasa secara online melalui platform seperti Shopee dan Tokopedia. Hal ini bertujuan untuk mempermudah proses jual beli barang agar lebih efektif. Perkembangan e-commerce juga tidak terlepas dari peran penting kemajuan teknologi dan informasi yang sangat krusial, terlebih di negara-negara berkembang seperti Indonesia (Manajemen, Yuli, & Aisah, 2025). Pengaruh perkembangan e-commerce memberikan dampak yang sangat besar di Indonesia. Menurut hasil riset yang dilakukan oleh YouGov pada tahun 2025, Indonesia berada di peringkat kedua setelah Tiongkok sebagai negara yang memanfaatkan platform e-commerce untuk dijadikan alternatif dalam melakukan belanja online. Pertumbuhan e-commerce di Indonesia berkembang pesat. Tercatat pada tahun 2025 pengguna e-commerce mencapai 72,9 juta orang. Angka ini menunjukkan kenaikan yang signifikan jika dibandingkan tahun sebelumnya yang berada di angka 65,65 juta orang pada tahun 2024 (Khaira Ummah Junaedi Putri, 2025). Hal ini juga terjadi secara khusus pada platform besar seperti Shopee dan Tokopedia. Ulasan produk sangat penting dan memberi

pengaruh terhadap produsen maupun konsumen, karena ulasan menjadi sumber informasi tentang kualitas suatu produk (Hasugian, Fakhriya, & Zukhoiriyah, 2023). Hal ini pula yang menyebabkan lonjakan yang signifikan dari konsumen terhadap ulasan produk e-commerce dari kedua platform ini, yang selalu dilengkapi dengan penilaian berupa rating dengan skala 1–5. Dengan demikian, dapat digambarkan pemahaman pelanggan terhadap kualitas sebuah produk dan juga pelayanannya. Akan tetapi, kapasitas ulasan yang sangat besar dan sifat teks yang tidak sistematis, seperti penggunaan bahasa nonbaku, singkatan, dan noise, menyebabkan ulasan tersebut sulit dianalisis secara manual.

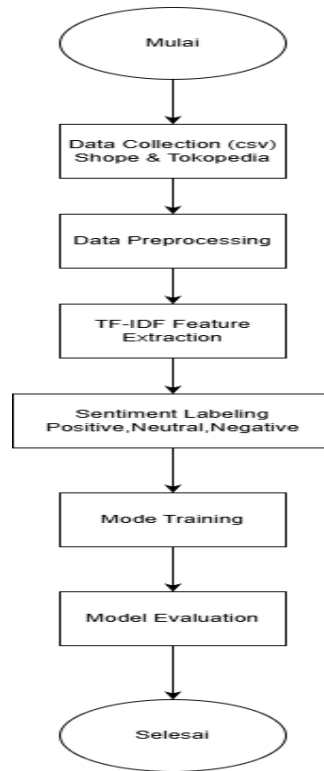
Analisis sentimen adalah salah satu cabang dari pemrosesan bahasa alami (natural language processing) yang digunakan untuk melacak suasana hati masyarakat mengenai suatu produk atau isu yang beredar (Prasetyo, Waskita, & Taryo, 2025). Tugas dasar dalam analisis sentimen adalah mengklasifikasikan polaritas dokumen, kalimat, atau teks opini (Kristanto, Winarno, & Nasiri, 2023). Tujuan analisis sentimen adalah untuk mengenali persepsi atau pandangan pengguna terhadap kualitas produk dan pelayanan dari kedua platform (Ma'rufudin & Yudhistira, 2025). Secara garis besar, terdapat beberapa tahapan yang dilakukan dalam menganalisis sentimen, yaitu pengumpulan data, preprocessing, ekstraksi fitur, klasifikasi, dan evaluasi. Proses pengumpulan data ulasan dilakukan dengan teknik scraping menggunakan bahasa pemrograman Python pada platform Google Colab yang bersumber dari Google Play Store dan App Store (Ubaidillah Fahmi, Anjani Arifiyanti, & Luhur Indayanti Sugata, 2025). Setelah pengumpulan data dilakukan, preprocessing digunakan untuk menghilangkan informasi yang tidak diperlukan sekaligus bertujuan mengurangi dimensi input ke dalam model (Aufa & Qoiriah, 2023). Tahapan ini meliputi case folding, tokenisasi, penghapusan stopword, dan stemming (Ardiani, Sujaini, & Tursina, 2020). Selanjutnya dilakukan tahap ekstraksi fitur yang bertujuan mengubah teks menjadi bentuk numerik, kemudian dilanjutkan dengan proses klasifikasi menggunakan algoritma yang telah ditentukan, dan yang terakhir adalah evaluasi model untuk menilai kinerja metode yang digunakan (Awaludin, Yasin, et al., 2024).

Analisis sentimen yang dilakukan dengan menggunakan algoritma Naïve Bayes dan Decision Tree sangat dibutuhkan untuk mengklasifikasikan ulasan secara akurat ke dalam kelompok sentimen positif, netral, dan negatif (Awaludin, Nuryadi, et al., 2024). Agar dapat memastikan akurasi yang optimal, penelitian ini mengimplementasikan tahapan text preprocessing yang ketat (Awaludin & Gani, 2024). Text preprocessing merupakan tahap penting dalam pengolahan data teks yang bertujuan untuk membersihkan serta menyiapkan data agar dapat dianalisis secara efektif oleh algoritma pembelajaran mesin (Agustina & Herliana, 2025). Selanjutnya dilakukan labeling sentimen dengan mengonversi rating menjadi kategori sentimen individual (negatif, netral, positif). Penelitian ini secara khusus membandingkan dua algoritma klasifikasi teks populer, yaitu Naïve Bayes dan Decision Tree.

Algoritma Naïve Bayes merupakan metode klasifikasi yang berdasar pada Teorema Bayes serta memiliki ciri khas berupa asumsi yang sangat kuat terhadap kemandirian antar fitur (Hananto, Erfiana, Putri, Putri, & Kurniawan, 2023). Sementara itu, algoritma Decision Tree mampu menangani data dengan atribut numerik maupun kategorikal, serta memberikan visualisasi pohon keputusan yang mudah dipahami (Nuryamin & Risyda, 2025). Perbandingan kedua algoritma ini dilakukan untuk mengetahui model mana yang paling akurat dan efektif dalam analisis sentimen ulasan produk online pada dataset yang digunakan. Hasil perbandingan ini akan menjadi dasar penentuan model terbaik untuk implementasi sistem klasifikasi sentimen ulasan di lingkungan e-commerce. Berdasarkan penelitian sebelumnya yang membandingkan dua algoritma, yaitu Naïve Bayes dan Decision Tree C4.5, tentang analisis sentimen produk Es Teh Indonesia di media sosial Twitter, kedua metode tersebut memperoleh akurasi masing-masing 71,96% untuk Decision Tree C4.5 dan 66,11% untuk Naïve Bayes (Rachmat Hidayat, 2024). Sedangkan pada penelitian yang dilakukan oleh Muhamad Iqbal dan rekan-rekannya dengan judul *Perbandingan Metode Decision Tree dan Naive Bayes pada Tingkat Penjualan Minuman Kopi di Kopi Pawon Nusantara*, diperoleh hasil bahwa Decision Tree mencapai akurasi sebesar 97,83%, sementara Naive Bayes memiliki akurasi sebesar 94,35% (Iqbal, Miskiyah, Sham, Anwar, & Fuad, 2024). Jika dilihat dari perbedaan akurasi pada kedua algoritma tersebut berdasarkan dua penelitian sebelumnya, maka penelitian ini akan membandingkan kembali kinerja Naïve Bayes dan Decision Tree dengan menggunakan dataset dari platform Kaggle serta Google Colab sebagai platform untuk mengolah data, untuk menentukan metode yang memiliki kinerja dan akurasi yang tinggi dalam menganalisis ulasan produk dari platform e-commerce Shopee dan Tokopedia.

2. METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen yang bertujuan untuk mengukur dan membandingkan performa dua algoritma klasifikasi, yaitu Naïve Bayes dan Decision Tree, dalam menganalisis sentimen ulasan produk. Berikut merupakan alur penelitian yang dilakukan..



Gambar 1 : Alur penelitian

2.1 Ruang Lingkup Penelitian

Penelitian berfokus pada analisis sentimen ulasan pelanggan dari platform e-commerce Shopee dan Tokopedia. Sentimen yang dianalisis terdiri atas tiga kategori, yaitu positif, netral, dan negatif. Data yang digunakan berupa teks ulasan beserta rating pengguna sebagai dasar penentuan label sentimen.

2.2 Pengumpulan Data

Dataset diperoleh dari platform Kaggle dalam format CSV, yang berisi 68.346 entri ulasan produk. Atribut utama dalam dataset terdiri dari:

- review_id – identitas unik ulasan
- review – teks ulasan pengguna
- rating – penilaian numerik terhadap produk

Dataset menggunakan campuran bahasa Indonesia dan Inggris serta mengandung karakter khusus, sehingga membutuhkan proses pembersihan sebelum digunakan. Dataset ini dapat diakses dan diunduh melalui:

<https://www.kaggle.com/datasets/silviamargareta/dataset-review>

A	B	C	D
review_id	review	rating	
0	slow delivery		1
1	Dateng goods do not conform pesanan??		1
2	PSN k its 20 other DTG		1
3	I am expected that it have a frame and painted in a canva		1
4	The product quality is not good.		1
5	goods came really not the same as shown in the descripti		1
6	Q order came quickly but im so very disappointed! Larger		1
7	Aprox 1 bjunyaa		1
8	Send a brown color to black, very sick.		1
9	Did not receive item Will not order again		1
10	Open items have dumbfounded used to look over n year		1
11	The product quality is not good. There is a flaw		1
12	Good product but right ga ga given promises change		1
13	product defects		1
14	Of 5 orders ... motif that BNR's just ... hope next ??? coul		1
15	Item shipped late		1
16	Sandals too little, just like sz 35.36 only. Shop listing bu		1
17	Quality is ok but the size chart is wrong it stated 122 cm b		1
18	Not to order very kecwaa very jero jero		1
19	D beautiful packaging so nagcrack nong shipping those ite		1
20	Her stuff is nice but it really black color at red ... My right r		1

Gambar 2 : Tampilan Dataset

2.3 Preprocessing Data

Preprocessing data merupakan salah satu hal yang sangat penting dalam analisis *data mining* untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat diolah (Awaludin & Gani, 2024). pengolahan data dilakukan dalam beberapa tahap, sebagai berikut:

- Pembersihan Data Awal
 1. menghapus data duplikat yang terdapat pada dataset.
 2. menghapus entri yang memiliki nilai kosong pada kolom ulasan.
- *Preprocessing Text*
Langkah ini bertujuan untuk mengubah teks mentah ke dalam format yang lebih terstruktur agar dapat diolah oleh model. Tahapan-tahapannya adalah:
 1. *Lowercasing* (mengubah teks menjadi huruf kecil).
 2. Penghapusan angka, tanda baca, emoji, dan simbol.
 3. Tokenisasi.
 4. *Stopword removal*.
 5. *Stemming* atau normalisasi teks.

Hasil *preprocessing* berupa teks bersih (*cleaned review*) yang siap dikonversi menjadi fitur numerik.

2.4 TF-IDF Extraction

TF-IDF adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami (Septiani & Isabela, 2022). Setiap kata pada teks diubah menjadi bobot numerik menggunakan metode TF-IDF untuk merepresentasikan tingkat kepentingannya dalam suatu ulasan.

2.5 Pelabelan Sentimen

Label sentimen ditentukan berdasarkan nilai rating dengan ketentuan:

1. Rating $\leq 2 \rightarrow$ Negatif
2. Rating = 3 \rightarrow Netral
3. Rating $\geq 4 \rightarrow$ Positif

2.6 Mode Traning

Pada tahap ini kedua algoritma Naïve Bayes dan Decision Tree melalui proses pelatihan.

```

# Tahap 5: Split data
X_train, X_test, y_train, y_test = train_test_split(
    df['cleaned'], df['sentiment'], test_size=0.2, random_state=42
)

# Tahap 6: TF-IDF Vectorizer
vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

# --- Model Naive Bayes ---
nb_model = MultinomialNB()
nb_model.fit(X_train_tfidf, y_train)
nb_pred = nb_model.predict(X_test_tfidf)

# --- Model Decision Tree ---
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train_tfidf, y_train)
dt_pred = dt_model.predict(X_test_tfidf)

```

Gambar 3. Tahap model training

Sebelum model dilatih, dataset dibagi menjadi dua bagian menggunakan metode *train-test split*, yaitu 80% data untuk proses pelatihan dan 20% data untuk proses pengujian. Tujuan pembagian ini adalah agar model dapat dievaluasi menggunakan data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi menjadi lebih objektif. Selanjutnya, data teks yang telah melalui tahap *preprocessing* dikonversi menjadi bentuk numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*).

2.7 Model Evaluation

Evaluasi performa model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Perbandingan kedua model dilakukan untuk menentukan algoritma yang memberikan performa terbaik dalam klasifikasi sentimen.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pada penelitian ini digunakan dataset hasil ulasan produk e-commerce Shopee dan Tokopedia yang diperoleh dari platform Kaggle, dengan jumlah data awal 68.346 entri yang terdiri atas kolom *review_id*, *review*, dan *rating*.

```

...   review_id      review  rating
0      0      slow delivery      1
1      1  Dateng goods do not conform pesanan??      1
2      2      PSN k its 20 other DTG      1
3      3  I am expected that it have a frame and painted...      1
4      4  The product quality is not good.      1
Index(['review_id', 'review', 'rating'], dtype='object')
Jumlah data sebelum pembersihan: 68346

```

Gambar 4: Jumlah Data Awal sebelum Preprocessing

Setelah dilakukan pemeriksaan kualitas data, ditemukan 305 ulasan kosong pada kolom *review*, sehingga jumlah data valid menjadi sekitar 68.041 entri. Untuk keperluan pelatihan dan pengujian model, digunakan *subset* penelitian berjumlah 46.086 ulasan, sesuai prosedur yang telah dijelaskan pada Metode Penelitian. Dataset ini dipilih karena bersifat terbuka, sering digunakan dalam analisis sentimen e-commerce, serta memiliki variasi ekspresi bahasa yang cukup kaya untuk proses klasifikasi berbasis teks.

```

... Jumlah data kosong per kolom:
review_id      0
review        305
rating         0
cleaned        0
dtype: int64
Total data kosong: 305

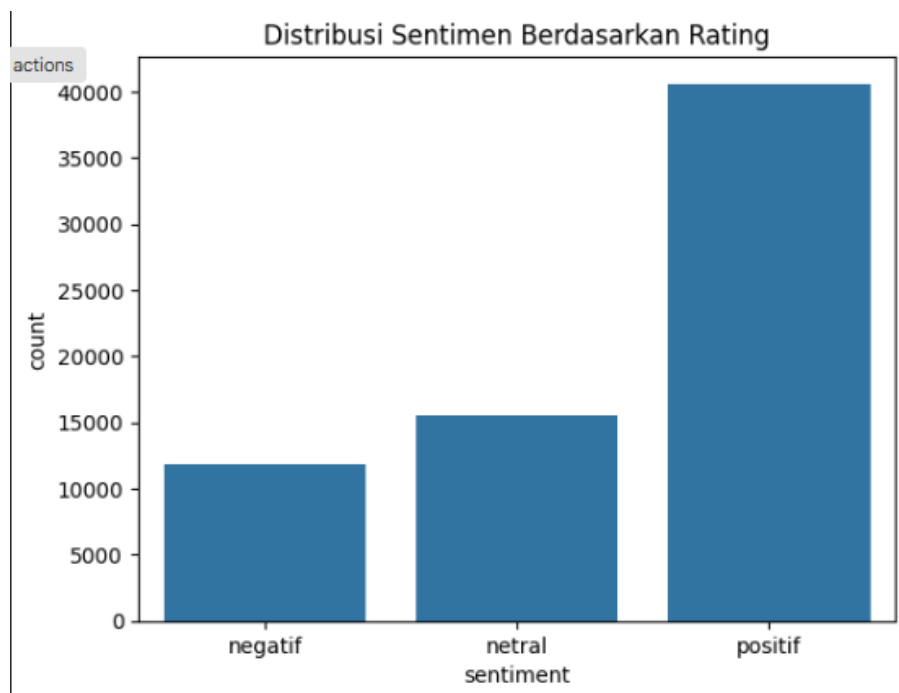
```

Gambar 5: Hasil pembersihan Dan Pemeriksaan Data Kosong

3.2. Praproses

Analisis eksplorasi awal dilakukan untuk memahami karakteristik data ulasan. Distribusi kelas sentimen yang diperoleh dari rating menunjukkan kondisi yang tidak seimbang: sentimen positif mendominasi (± 40 ribu), disusul netral (± 15 ribu), dan negatif (± 12 ribu). Ketidakseimbangan ini menjadi perhatian karena dapat memengaruhi kemampuan model dalam mengenali sentimen minoritas, terutama kelas netral.

Tahap TF-IDF menghasilkan matriks fitur berukuran $N \times |V|$, dengan $N = 46.086$ dan $|V|$ berupa jumlah token unik setelah *preprocessing* (berkisar puluhan ribu). Analisis bobot menunjukkan bahwa kata-kata adjektiva terkait kualitas dan layanan (“bagus”, “cepat”, “murah”) memiliki bobot tinggi, sedangkan kata-kata umum memperoleh bobot rendah.



Gambar 6: Tabel Labeling Sentimen

3.3. Hasil Algoritma

Setelah melakukan dan melewati praproses dan pelabelan sentimen, selanjutnya masuk ke tahap perbandingan akurasi dan juga kinerja dari kedua algoritma, yaitu Naïve Bayes dan Decision Tree.

A. Decision Tree

Setelah diuji pada data pengujian, model Decision Tree menunjukkan hasil akurasi sebesar 67,41%, seperti yang terlihat pada gambar berikut:

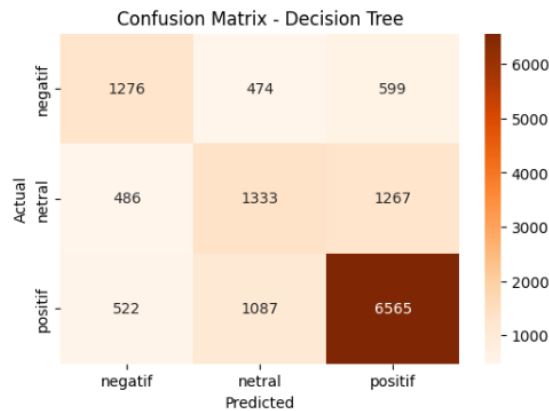
```

=== Decision Tree ===
Akurasi: 0.6741127195238446
precision    recall  f1-score   support

negatif      0.56    0.54    0.55     2349
netral       0.46    0.43    0.45     3086
positif      0.78    0.80    0.79     8174

accuracy          0.67    13609
macro avg         0.60    13609
weighted avg      0.67    13609

```

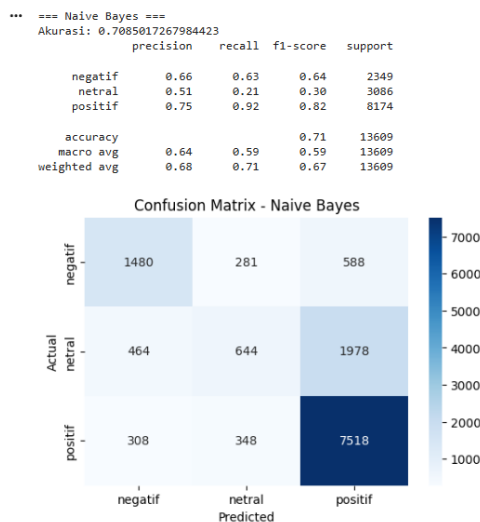


Gambar 7:Hasil Algoritma Decision Tree

Angka ini menunjukkan performa yang cukup baik untuk mengidentifikasi sentimen positif sebagaimana terlihat dari tingginya nilai *precision* dan *recall* pada kelas mayoritas tersebut. Namun, kemampuan dalam mengenali sentimen netral masih kurang optimal, dengan *F1-score* hanya 0,46, karena kelas ini lebih sedikit dan memiliki ekspresi linguistik yang cenderung ambigu. Ketergantungan Decision Tree pada pemilihan fitur tunggal pada setiap *node* menyebabkan sensitivitas tinggi terhadap variasi kecil pada teks.

B. Naïve Bayes

Model Naïve Bayes menunjukkan performa lebih baik dibanding Decision Tree dengan akurasi 70,85%. Khusus pada kelas positif, model mencapai *F1-score* sebesar 0,83, menjadikannya sangat akurat dalam mengenali ekspresi ulasan bernada pujian.



Gambar 8:Hasil Algoritma Naïve Bayes

Meskipun demikian, performa pada kelas netral masih rendah ($\approx 0,21$) karena model sangat bergantung pada asumsi independensi fitur, sedangkan ekspresi netral cenderung tidak memiliki kata

penanda yang kuat. Meski begitu, secara keseluruhan Naïve Bayes lebih stabil dan konsisten karena bekerja efektif pada data berdimensi tinggi seperti teks.

4.KESIMPULAN

Penelitian ini membandingkan kinerja algoritma Naïve Bayes dan Decision Tree dalam melakukan klasifikasi sentimen ulasan produk e-commerce. Seluruh proses, mulai dari pembersihan data, ekstraksi fitur TF-IDF, hingga evaluasi model, dilakukan menggunakan Python. Hasil pengujian menunjukkan bahwa Naïve Bayes memberikan performa yang lebih baik dibandingkan Decision Tree, dengan akurasi berada pada kisaran 70,98%–71,50%, sedangkan Decision Tree memperoleh akurasi sekitar 68,35%–69,70%. Dengan demikian, Naïve Bayes menjadi algoritma yang paling optimal diterapkan pada dataset ini.

DAFTAR PUSTAKA

- Agustina, V., & Herliana, A. (2025). Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing. *Jurnal Media Informatika*, 6(3), 2182–2194. Retrieved from <https://ejournal.sisfokomtek.org/index.php/jumin/article/view/6301>
- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 8(2), 183. <https://doi.org/10.26418/justin.v8i2.36776>
- Aufa, M. J., & Qoiriah, A. (2023). Analisis Sentimen Pengguna Platform Belajar Online Coursera menggunakan Random Forest dengan Metode Ekstraksi Fitur Word2vec. *Journal of Informatics and Computer Science (JINACS)*, 04, 244–255. <https://doi.org/10.26740/jinacs.v4n02.p244-255>
- Awaludin, M., & Gani, A. (2024). Pemanfaatan kecerdasan buatan pada algoritma k-means klustering dan sentiment analysis terhadap strategi promosi yang sukses untuk penerimaan mahasiswa baru. *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, 11(1), 1–6.
- Awaludin, M., & Gani, A. (2024). Pemanfaatan kecerdasan buatan pada algoritma k-means klustering dan sentiment analysis terhadap strategi promosi yang sukses untuk penerimaan mahasiswa baru. *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, 11(1), 1–6.
- Awaludin, M., Nuryadi, H., & Pribadi, G. N. (2024). *Sistem Otomatisasi Laporan untuk Optimalisasi Pelaporan Data Penelitian dan Pengabdian kepada Masyarakat di Universitas Dirgantara Marsekal Suryadarma*. 9675, 1–7.
- Awaludin, M., Yasin, V., & Risyda, F. (2024). The Influence of Artificial Intelligence Technology, Infrastructure and Human Resource Competence on Internet Access Networks. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 9(2), 111–120. <https://doi.org/10.25139/inform.v9i2.8109>
- Hananto, A. D., Erfiana, A. M., Putri, B. L. P., Putri, P. D., & Kurniawan, F. (2023). Algoritma Machine Learning Naïve Bayes pada Analisis Sentimen Kesepakatan Polri dan GNPF-MUI pada Aksi Bela Islam III “212.” *SINTA Journal (Science, Technology, and Agricultural)*, 4(2), 151–160. <https://doi.org/10.37638/sinta.4.2.151-160>
- Hasugian, A. H., Fakhriza, M., & Zukhoiriyah, D. (2023). Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 6(1), 98. <https://doi.org/10.53513/jsk.v6i1.7400>
- Iqbal, M., Miskiyah, S., Sham, S. L., Anwar, S., & Fuad, M. H. (2024). Perbandingan Metode Decision Tree dan Naive Bayes Pada Tingkat Penjualan Minuman Kopi di Kopi Pawon Nusantara. *Jurnal INSAN Journal of Information System Management Innovation*, 4(1), 27–34. <https://doi.org/10.31294/jinsan.v4i1.3682>
- Khaira Ummah Junaedi Putri. (2025). [Update] Data e-commerce Indonesia: panduan lengkap. Retrieved November 24, 2025, from <https://id.techinasia.com/data-ecommerce-indonesia-panduan-lengkap#:~:text=Rp60%2C75 juta.,Tren e-commerce di masa depan,total GMV e-commerce Indonesia>.
- Kristanto, F., Winarno, W. W., & Nasiri, A. (2023). Perbandingan Algoritme Naïve Bayes dan Decision Tree Pada Analisis Sentimen Data Komentar Siswa Pada Aplikasi Digital Teacher Assessment. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 12(2), 538. <https://doi.org/10.35889/jutisi.v12i2.1193>
- Ma'rufudin, M., & Yudhistira, A. (2025). Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 845–857. <https://doi.org/10.52436/1.jpti.717>
- Manajemen, J., Yuli, M., & Aisah, S. (2025). Perkembangan dan Tren E-Commerce di Indonesia, 3(September).
- Nuryamin, Y., & Risyda, F. (2025). JSI (Jurnal Sistem Informasi) Universitas Dirgantara Marsekal Suryadarma Analisis Prediksi Penyakit Diabetes Menggunakan Metode Decision Tree C4.5 Dan Naive Bayes, 5, 234–242. Retrieved from <https://www.kaggle.com/code/melikedilekci/diabetes-dataset-for-beginners>
- Prasetyo, T., Waskita, A. A., & Taryo, T. (2025). Analisis Sentimen Pengguna Seputar Kendaraan Listrik di

- Twitter dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, VIII(2), 108–117.
- Rachmat Hidayat, R. H. (2024). Perbandingan Metode Naïve Bayes Dan Decision Tree C4.5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 7(2), 88–98. <https://doi.org/10.47970/siskom-kb.v7i2.607>
- Septiani, D., & Isabela, I. (2022). Term Frequency Inverse Document Frequency (Tf-Idf) Analysis in Information Retrieval in Text Documents. *Jurnal Sistem Dan Teknologi Informasi Indonesia(SINTESIA)*, 1(2), 81–88.
- Ubaidillah Fahmi, R., Anjani Arifiyanti, A., & Luhur Indayanti Sugata, T. (2025). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Midi Kriing Menggunakan Support Vector Machine (Svm). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4831–4839. <https://doi.org/10.36040/jati.v9i3.13783>