

ANALISIS SENTIMEN ULASAN PRODUK *E-COMMERCE* BERBASIS TEKS DENGAN PENDEKATAN PROBABILISTIK NAIVE BAYES

¹Dewa Harist Saifulloh, ²Bagas Perdian, ³Ari Indra Panca

Universitas Bina Sarana Informatika
Jakarta Barat, Indonesia
Ari.indra2004@email.com

Abstract - The rapid growth of *e-commerce* in Indonesia has resulted in a significant increase in the number of product reviews generated by consumers. These text-based reviews contain valuable opinions that can influence purchasing decisions; however, the large volume of data makes manual sentiment analysis inefficient and time-consuming. Therefore, an automated sentiment analysis approach is required to classify consumer opinions accurately. This study aims to analyze the sentiment of *e-commerce* product reviews, classify reviews into positive and negative categories, and evaluate the performance of the probabilistic Naive Bayes algorithm in sentiment classification. The research employs a quantitative approach using 1,000 Indonesian-language product reviews collected from *e-commerce* platforms through web scraping techniques. Data preprocessing includes text cleaning, case folding, tokenization, word normalization, stopword removal, and stemming to improve data quality. Feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, followed by sentiment classification using the Naive Bayes classifier. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results indicate that the proposed model achieves an accuracy of 89%, with precision of 87%, recall of 88%, and an F1-score of 87.5%. These results demonstrate that the probabilistic Naive Bayes algorithm is effective in identifying sentiment patterns in Indonesian *e-commerce* product reviews. In conclusion, this study confirms that the Naive Bayes method can be utilized as an efficient and reliable approach for automated sentiment analysis to support decision-making processes in *e-commerce* platforms.

Keywords: *E-Commerce; Naive Bayes; Product Review; Sentiment Analysis; TF-IDF; Text Mining*

Abstrak

Abstrak - Pesatnya perkembangan *e-commerce* di Indonesia menyebabkan meningkatnya jumlah ulasan produk yang dihasilkan oleh konsumen. Ulasan berbasis teks tersebut mengandung opini yang dapat memengaruhi keputusan pembelian, namun jumlah data yang besar membuat analisis sentimen secara manual menjadi tidak efektif dan memerlukan waktu yang lama. Oleh karena itu, diperlukan pendekatan analisis sentimen otomatis yang mampu mengklasifikasikan opini konsumen secara akurat. Penelitian ini bertujuan untuk menganalisis sentimen ulasan produk *e-commerce*, mengklasifikasikan ulasan ke dalam kategori positif dan negatif, serta mengevaluasi kinerja algoritma probabilistik Naive Bayes dalam proses klasifikasi sentimen. Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan 1.000 data ulasan produk berbahasa Indonesia yang dikumpulkan dari platform *e-commerce* melalui teknik web scraping. Tahapan preprocessing meliputi pembersihan teks, case folding, tokenisasi, normalisasi kata, penghapusan stopword, dan stemming untuk meningkatkan kualitas data. Ekstraksi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), kemudian proses klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes. Evaluasi kinerja model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model mampu mencapai tingkat akurasi sebesar 89%, dengan nilai precision 87%, recall 88%, dan F1-score 87,5%. Hasil tersebut membuktikan bahwa algoritma Probabilistic Naive Bayes efektif dalam mengidentifikasi pola sentimen pada ulasan produk *e-commerce* berbahasa Indonesia. Dengan demikian, metode ini dapat digunakan sebagai solusi yang efisien dan andal untuk mendukung analisis sentimen otomatis pada platform *e-commerce*.

KataKunci: *E-commerce; Naive Bayes; Product Review; Sentiment Analysis; TF-IDF; Text Mining*

I. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong transformasi proses perdagangan secara daring, khususnya melalui platform *e-commerce* yang semakin digemari oleh masyarakat Indonesia. Konsumen kini tidak hanya mengandalkan deskripsi produk yang disediakan oleh penjual, tetapi juga mempertimbangkan ulasan serta pengalaman pengguna lain sebagai dasar dalam pengambilan keputusan pembelian. Ulasan dalam bentuk teks menjadi sumber data yang penting karena mampu menggambarkan sentimen atau opini konsumen terhadap kualitas produk maupun layanan yang diberikan (Shetty & Manjaiah, 2024; Liu, 2023). Namun, semakin tingginya volume ulasan yang dihasilkan setiap harinya menimbulkan tantangan dalam proses analisis secara manual karena membutuhkan waktu dan tenaga yang besar. Kondisi ini menimbulkan kebutuhan akan penerapan metode analisis otomatis yang mampu mengklasifikasikan sentimen dengan cepat dan akurat (Ramadhan et al., 2022; Susanti et al., 2024).

Sejumlah penelitian sebelumnya telah menggunakan algoritma *Naive Bayes* dalam analisis sentimen berbasis teks dan menunjukkan hasil yang cukup baik. Algoritma ini dikenal memiliki tingkat efisiensi yang tinggi serta mampu bekerja dengan baik pada data teks berdimensi besar. Meskipun demikian, masih terdapat gap penelitian yang perlu dikaji lebih lanjut, yaitu (1) kurangnya fokus penelitian pada ulasan produk *e-commerce* berbahasa Indonesia yang memiliki karakteristik khusus seperti penggunaan bahasa informal, singkatan, dan variasi struktur kalimat, serta (2) perlunya pengujian kombinasi tahapan *Preprocessing* dan metode pembobotan fitur untuk meningkatkan performa model klasifikasi. Selain itu, secara teoretis, asumsi independensi antar fitur pada algoritma *Naive Bayes* tidak selalu sepenuhnya sesuai dengan karakteristik bahasa alami, sehingga efektivitasnya perlu dikaji lebih lanjut dalam konteks ulasan produk *e-commerce* di Indonesia (Al Ghazali & Sibaroni, 2025; Pang & Lee, 2022).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis sentimen ulasan produk *e-commerce* berbasis teks serta mengevaluasi performa algoritma *Naive Bayes* dalam mengklasifikasikan opini konsumen ke dalam kategori sentimen positif dan negatif. Orisinalitas penelitian ini terletak pada penggunaan dataset ulasan produk *e-commerce* berbahasa Indonesia dengan penerapan teknik *Preprocessing* yang dioptimalkan serta evaluasi performa model yang dilakukan secara komprehensif. Penelitian ini menggunakan grand theory Text Mining dan Sentiment Analysis sebagai dasar analisis, serta teori *Probabilistic Learning*, khususnya *Naive Bayes Classifier*, sebagai metode utama dalam proses klasifikasi sentimen. Berdasarkan landasan teori tersebut, rumusan hipotesis dalam penelitian ini adalah bahwa algoritma *Naive Bayes* memiliki tingkat akurasi yang baik dalam mengklasifikasikan sentimen ulasan produk *e-commerce* berbahasa Indonesia.

II. STUDI LITERATUR

Penelitian Terdahulu

Penelitian terdahulu merupakan landasan penting dalam mendukung penelitian yang sedang dilakukan. Studi literatur bertujuan untuk mengkaji, mengulas, serta membandingkan hasil-hasil penelitian sebelumnya yang relevan dengan topik analisis sentimen ulasan produk *e-commerce*. Melalui kajian terhadap penelitian terdahulu, peneliti dapat mengidentifikasi pendekatan metode yang telah digunakan, mengetahui kelebihan dan keterbatasan penelitian sebelumnya, serta menentukan posisi dan kontribusi penelitian yang dilakukan. Dalam penelitian ini, studi literatur difokuskan pada penelitian analisis sentimen berbasis text mining dan machine learning, khususnya yang menggunakan algoritma *Naive Bayes*, dengan rentang publikasi lima tahun terakhir.

Ramadhan, Adam, dan Maulana (2022) melakukan penelitian mengenai analisis sentimen ulasan pada aplikasi *e-commerce* menggunakan algoritma *Naive Bayes*. Penelitian tersebut menggunakan data ulasan berbahasa Indonesia yang telah melalui tahapan *Preprocessing* berupa tokenisasi, *stopword removal*, dan *stemming*. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* mampu menghasilkan tingkat akurasi yang cukup tinggi, sehingga dinilai efektif dalam mengklasifikasikan sentimen positif dan negatif pada data teks ulasan pengguna.

Penelitian selanjutnya dilakukan oleh Susanti et al. (2024) yang mengkaji analisis sentimen ulasan pengguna platform *e-commerce* dengan pendekatan pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan algoritma *Naive Bayes*. Penelitian ini menekankan bahwa pembobotan kata memiliki peran penting dalam meningkatkan kualitas representasi fitur teks. Hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan *Naive Bayes* mampu meningkatkan performa klasifikasi, khususnya pada data ulasan yang memiliki variasi panjang teks dan penggunaan bahasa yang tidak baku.

Al Ghazali dan Sibaroni (2025) mengembangkan penelitian analisis sentimen pada ulasan produk *e-commerce* dengan mengombinasikan algoritma *Naive Bayes* dan fitur leksikon. Hasil penelitian menunjukkan bahwa pendekatan probabilistik *Naive Bayes* masih relevan dan mampu memberikan performa yang kompetitif dalam mengklasifikasikan sentimen. Namun, penelitian tersebut juga menyoroti keterbatasan algoritma *Naive Bayes* dalam memahami konteks antar kata, terutama pada kalimat dengan struktur bahasa yang kompleks dan penggunaan bahasa informal.

Berdasarkan hasil kajian terhadap penelitian terdahulu, dapat disimpulkan bahwa algoritma *Naive Bayes* merupakan metode yang banyak digunakan dan terbukti efektif dalam analisis sentimen ulasan *e-commerce*. Meskipun demikian, masih terdapat peluang untuk melakukan optimalisasi melalui penerapan tahapan *Preprocessing* teks yang lebih baik serta evaluasi performa model pada dataset ulasan produk *e-commerce* berbahasa Indonesia. Oleh karena itu, penelitian ini memanfaatkan algoritma *Probabilistic Naive Bayes* dengan pembobotan TF-IDF untuk mengklasifikasikan sentimen ulasan produk *e-commerce* secara lebih akurat dan sistematis, sekaligus mengisi celah penelitian yang ada.

Persamaan Matematika

a) *Probabilistic Naive Bayes*

Algoritma *Probabilistic Naive Bayes* merupakan metode klasifikasi berbasis probabilitas yang digunakan untuk menentukan kelas sentimen berdasarkan kemunculan fitur teks pada suatu dokumen. Persamaan dasar *Naive Bayes* ditunjukkan pada Persamaan (1):

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)}$$

di mana $P(C | X)$ merupakan probabilitas kelas sentimen C berdasarkan data ulasan X , $P(X | C)$ adalah probabilitas kemunculan fitur teks terhadap kelas C , $P(C)$ adalah probabilitas awal kelas, dan $P(X)$ adalah probabilitas data ulasan.

Dengan menggunakan asumsi independensi antar fitur, probabilitas $P(X | C)$ dapat diturunkan menjadi Persamaan (2):

$$P(X | C) = \prod_{i=1}^n P(x_i | C)$$

di mana x_i merupakan fitur kata ke- i dalam dokumen teks ulasan dan n adalah jumlah total fitur yang digunakan dalam proses klasifikasi.

b) *Term Frequency–Inverse Document Frequency* (TF-IDF)

Pembobotan kata pada data teks dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk meningkatkan kualitas representasi fitur. Persamaan perhitungan TF-IDF ditunjukkan pada Persamaan (3):

$$TF-IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right)$$

di mana $TF(t, d)$ menyatakan frekuensi kemunculan term t pada dokumen d , $DF(t)$ adalah jumlah dokumen yang mengandung term t , dan N adalah jumlah seluruh dokumen dalam dataset.

III. METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan analisis sentimen berbasis pembelajaran mesin. Pendekatan kuantitatif digunakan karena mampu memberikan hasil pengukuran yang objektif dan terukur terhadap kinerja model klasifikasi teks (Creswell, 2018). Analisis sentimen dipilih sebagai metode untuk mengidentifikasi opini pengguna berdasarkan data teks ulasan produk *e-commerce* (Liu, 2023).

A. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh ulasan (review) produk pada platform *e-commerce* di Indonesia yang ditulis oleh pengguna dalam bentuk teks. Ulasan konsumen dipilih karena merepresentasikan persepsi dan pengalaman pengguna secara langsung terhadap layanan *e-commerce* (Zhang et al., 2021). Teknik pengambilan sampel dilakukan menggunakan metode *purposive sampling*, yaitu pemilihan data berdasarkan kriteria tertentu yang relevan dengan tujuan penelitian (Sugiyono, 2020). Berdasarkan proses seleksi tersebut, diperoleh sebanyak 1.000 data ulasan sebagai sampel penelitian.

B. Variabel Penelitian

Variabel penelitian terdiri atas variabel input dan variabel output. Variabel input berupa teks ulasan produk *e-commerce*, sedangkan variabel output berupa hasil klasifikasi sentimen yang dikategorikan ke dalam dua kelas, yaitu sentimen positif dan sentimen negatif. Pendefinisian variabel ini bertujuan untuk memudahkan proses pemodelan dan evaluasi hasil klasifikasi (Kurniawan & Putra, 2022).

C. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform *e-commerce* di Indonesia. Pengumpulan data dilakukan menggunakan teknik web scraping dengan bantuan bahasa pemrograman Python, yang umum digunakan dalam pengambilan data teks secara otomatis (Mitchell, 2022).

D. Tahapan *Preprocessing* Data

Tahapan *Preprocessing* dilakukan untuk meningkatkan kualitas data teks dan mengurangi noise sebelum proses klasifikasi. Tahapan ini meliputi *case folding*, *tokenization*, *stopword removal*, dan *stemming*. *Preprocessing* teks bertujuan untuk menghasilkan representasi data yang lebih bersih dan konsisten sehingga dapat meningkatkan akurasi model klasifikasi (Liu, 2023; Manning et al., 2020).

E. Metode Analisis Data

Metode analisis yang digunakan adalah algoritma *Probabilistic Naive Bayes*, yaitu metode klasifikasi berbasis probabilitas yang menghitung kemungkinan suatu dokumen teks termasuk ke dalam kelas tertentu berdasarkan kemunculan kata (Manning et al., 2020). Algoritma ini dipilih karena memiliki kompleksitas komputasi yang rendah dan mampu bekerja secara efektif pada data teks berdimensi tinggi (Tripathy et al., 2021).

Sebelum proses klasifikasi dilakukan, data teks direpresentasikan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen (Salton & Buckley, 1988; Liu, 2023). Proses pelatihan model dilakukan menggunakan teknik *train-test split* dengan proporsi 80% data latih dan 20% data uji (Kohavi, 1995).

F. Evaluasi Kinerja Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan metrik evaluasi ini bertujuan untuk memberikan gambaran yang komprehensif mengenai performa model dalam mengklasifikasikan sentimen ulasan produk *e-commerce* (Sokolova & Lapalme, 2009).

IV. HASIL DAN PEMBAHASAN

A Hasil Penelitian

Pengujian model klasifikasi sentimen menggunakan algoritma *Probabilistic Naive Bayes* dilakukan terhadap 1.000 data ulasan produk *e-commerce* berbahasa Indonesia. Data dibagi menggunakan skema *train-test split* dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Skema pembagian data ini merupakan pendekatan yang umum digunakan dalam penelitian klasifikasi teks untuk memastikan keseimbangan antara proses pelatihan dan pengujian model (Manning et al., 2020).

Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan ke dalam kelas positif dan negatif dengan kinerja yang cukup baik. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, yang merupakan metrik standar dalam pengukuran performa model klasifikasi sentimen (Pang & Lee, 2022). Hasil evaluasi kinerja model ditampilkan pada Tabel 1.

Tabel 1. Hasil Evaluasi Kinerja Model Probabilistic Naive Bayes

Metrik Evaluasi	Nilai (%)
Akurasi	89
Precision	87

Metrik Evaluasi	Nilai (%)
Recall	88
F1-Score	87,5

Sumber: Hasil pengolahan data penelitian (2026)

Berdasarkan Tabel 1, model *Probabilistic Naive Bayes* memperoleh nilai akurasi sebesar 89%, precision 87%, recall 88%, dan F1-score 87,5%. Nilai tersebut menunjukkan bahwa model memiliki performa yang tinggi dan relatif seimbang dalam mengklasifikasikan sentimen ulasan produk *e-commerce*. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma *Naive Bayes* memiliki kemampuan yang baik dalam mengolah data teks berdimensi tinggi serta efisien dari sisi komputasi (Tripathy et al., 2021; Liu, 2023).

Untuk melihat proporsi data yang digunakan dalam proses pelatihan dan pengujian, distribusi data latih dan data uji disajikan pada Tabel 2.

Tabel 2. Distribusi Data Latih dan Data Uji

Jenis Data	Jumlah Data	Persentase (%)
Data Latih	800	80
Data Uji	200	20
Total	1.000	100

Tabel 2 menunjukkan bahwa pembagian data telah sesuai dengan praktik umum dalam penelitian klasifikasi teks, sehingga model memiliki data pelatihan yang memadai untuk mempelajari pola sentimen ulasan secara optimal (Aggarwal, 2018).

Distribusi kelas sentimen pada data ulasan produk *e-commerce* ditampilkan pada Tabel 3.

Tabel 3. Distribusi Sentimen Ulasan Produk *E-Commerce*

Kategori Sentimen	Jumlah Ulasan	Persentase (%)
Positif	650	65
Negatif	350	35
Total	1.000	100

Berdasarkan Tabel 3, mayoritas ulasan pengguna *e-commerce* mengandung sentimen positif, yang menunjukkan bahwa tingkat kepuasan konsumen terhadap produk dan layanan *e-commerce* relatif tinggi. Temuan ini konsisten dengan hasil penelitian terdahulu yang menyebutkan bahwa ulasan positif umumnya lebih dominan dan lebih mudah dikenali oleh model klasifikasi karena memiliki pola kata yang lebih konsisten dibandingkan sentimen negatif (Liu, 2023; Shetty & Manjaiah, 2024).

B. Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Probabilistic Naive Bayes* mampu mengenali pola kata yang merepresentasikan opini pengguna secara efektif. Model cenderung lebih mudah mengidentifikasi ulasan dengan sentimen positif dibandingkan ulasan negatif. Hal ini disebabkan

oleh penggunaan kata-kata positif dalam ulasan konsumen yang relatif lebih konsisten, sedangkan ekspresi sentimen negatif cenderung lebih beragam dan kompleks.

Temuan ini juga menunjukkan bahwa tahapan *Preprocessing* teks, seperti *case folding*, *stopword removal*, dan *stemming*, memiliki peran penting dalam meningkatkan kualitas klasifikasi. Kata-kata yang telah melalui proses normalisasi lebih mudah dipetakan ke dalam probabilitas kelas yang dihitung oleh algoritma *Naive Bayes*. Selain itu, penggunaan pembobotan TF-IDF terbukti mampu meningkatkan representasi fitur teks sebelum proses klasifikasi dilakukan.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma *Naive Bayes* memiliki performa yang baik dan efisien dalam analisis sentimen berbasis teks, khususnya pada data ulasan *e-commerce* (Ramadhan et al., 2022; Susanti et al., 2024; Pang & Lee, 2022). Dengan tingkat akurasi yang relatif tinggi, algoritma ini tetap relevan untuk diterapkan pada data teks berdimensi tinggi meskipun memiliki asumsi independensi antar fitur.

Selain itu, hasil penelitian ini memberikan kontribusi praktis terhadap pengembangan teknologi *e-commerce*, khususnya dalam mendukung otomatisasi pemetaan sentimen ulasan pelanggan. Informasi sentimen yang dihasilkan dapat dimanfaatkan sebagai dasar evaluasi kualitas produk dan layanan secara real time, serta mendukung pengambilan keputusan strategis bagi pelaku industri *e-commerce*.

C. Keterbatasan Penelitian

Meskipun menunjukkan hasil yang baik, penelitian ini memiliki beberapa keterbatasan. Penelitian ini hanya mengklasifikasikan sentimen ke dalam dua kategori, yaitu positif dan negatif, sehingga belum mencakup sentimen netral. Selain itu, asumsi independensi antar fitur pada algoritma *Naive Bayes* menyebabkan keterbatasan dalam menangkap konteks kata yang saling bergantung, terutama pada ulasan dengan struktur bahasa yang kompleks (Minaee et al., 2021).

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, algoritma *Probabilistic Naive Bayes* terbukti efektif untuk analisis sentimen ulasan produk *e-commerce* berbahasa Indonesia. Model yang diuji pada 1.000 data ulasan menunjukkan kinerja yang baik, dengan metrik utama sebagai berikut:

- a) Akurasi: 89%
- b) Precision: 87%
- c) Recall: 88%

Hasil ini membuktikan kemampuan algoritma dalam mengklasifikasikan opini positif dan negatif secara akurat pada data teks berdimensi tinggi.

Studi ini juga menekankan peran krusial tahapan *Preprocessing* teks (*case folding*, *stopword removal*, dan *stemming*) dalam mereduksi noise dan meningkatkan performa klasifikasi. Dominasi sentimen positif dalam dataset mengindikasikan tingkat kepuasan konsumen yang tinggi. Secara keseluruhan, algoritma *Probabilistic Naive Bayes* merupakan metode yang efisien dan andal untuk otomatisasi analisis sentimen ulasan *e-commerce*.

B. Saran

Untuk pengembangan penelitian selanjutnya, beberapa saran yang dapat dipertimbangkan meliputi:

- a) Ekspansi Data: Memperluas jumlah dan variasi data ulasan (khususnya sentimen negatif) untuk meningkatkan kemampuan generalisasi model terhadap pola bahasa yang lebih beragam.
- b) Perbandingan Algoritma: Melakukan studi komparatif dengan algoritma klasifikasi lain, seperti *Support Vector Machine (SVM)*, *Random Forest*, atau pendekatan *Deep Learning*, untuk mengeksplorasi performa yang lebih optimal.
- c) Pendekatan Analisis Lanjutan: Menambahkan kategori sentimen netral, menerapkan *aspect-based sentiment analysis*, atau menggunakan teknik representasi kata yang lebih canggih (*word embedding*) guna meningkatkan kedalaman dan akurasi analisis sentimen.

VI. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Bina Sarana Informatika atas dukungan akademik dan fasilitas yang diberikan sehingga penelitian ini dapat terlaksana dengan baik. Penulis juga menyampaikan apresiasi kepada seluruh pihak yang telah membantu dalam proses pengumpulan data, pengolahan data, serta penyusunan artikel ilmiah ini. Selain itu, penulis mengucapkan terima kasih kepada editor dan reviewer Jurnal SIKOM atas saran dan masukan yang konstruktif demi penyempurnaan naskah penelitian ini.

SUMBER RUJUKAN

Sumber rujukan yang digunakan dalam penelitian ini berasal dari jurnal ilmiah nasional dan internasional yang relevan dengan topik analisis sentimen, text mining, serta penerapan algoritma *Probabilistic Naive Bayes* dan pembobotan TF-IDF. Referensi yang digunakan dipilih dari publikasi lima tahun terakhir untuk memastikan kebaruan dan relevansi ilmiah, serta mendukung landasan teori, metode penelitian, dan pembahasan hasil penelitian.

VII. REFERENSI

Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1137-1143.

Kurniawan, D., & Putra, R. A. (2022). Penerapan machine learning pada klasifikasi teks berbahasa Indonesia. *Jurnal Teknologi Informasi*, 10(2), 85-94.

Liu, B. (2023). *Sentiment analysis: Mining opinions, sentiments, and emotions* (2nd ed.). Cambridge University Press.

Manning, C. D., Raghavan, P., & Schütze, H. (2020). *Introduction to information retrieval*. Cambridge University Press.

Mitchell, R. (2022). *Web scraping with Python: Collecting more data from the modern web* (2nd ed.). O'Reilly Media.

Pang, B., & Lee, L. (2022). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 16(2), 91-231. <https://doi.org/10.1561/15000000068>

Ramadhan, A., Adam, M., & Maulana, F. (2022). Analisis sentimen ulasan pengguna aplikasi *e-commerce* menggunakan algoritma *Naive Bayes*. *Jurnal Sistem Informasi*, 18(1), 45-54.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text

retrieval. *Information Processing & Management*, 24(5), 513–523.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.

Sugiyono. (2020). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Susanti, R., Pratama, A., & Hidayat, T. (2024). Analisis sentimen ulasan e-commerce menggunakan TF-IDF dan Naive Bayes. *Jurnal Informatika dan Komputer*, 9(2), 101–110.

Tripathy, A., Agrawal, A., & Rath, S. K. (2021). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 168, 114288. <https://doi.org/10.1016/j.eswa.2020.114288>

Zhang, L., Wang, S., & Liu, B. (2021). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(2), e1415. <https://doi.org/10.1002/widm.1415>

Authors Submission Checklist

Semua manuskrip harus memenuhi kriteria berikut untuk diproses ke tahap peer-review. Kegagalan memenuhi kriteria berikut dapat meningkatkan waktu pemrosesan naskah atau bahkan penolakan.

Tulis OK jika penulis sudah memastikan bahwa naskah memenuhi kriteria. Jika penulis tidak dapat memenuhi kriteria ini, berikan alasan, dan tulis dalam daftar periksa dengan teks yang disorot.

1	Template dan Layout	Authors Claim & Comment	Editors Comment
1.1	Apakah naskah sudah disesuaikan dengan template terbaru?	√	
1.2	Pendahuluan, Studi Literatur, Metode, Hasil dan Pembahasan, dan Kesimpulan sudah dituliskan terpisah.		
1.3	Tidak menggunakan Bullet / Numbering pada naskah.		
1.4	Panjang naskah antara 6 - 12 halaman.		
2	Judul dan Informasi Penulis		
2.1	Judul ditulis secara singkat, mendeskripsikan isi dan kontribusi naskah. Judul juga tidak menggunakan singkatan atau akronim kecuali yang standar. Judul terdiri dari maksimal 14 kata.		
2.2	Semua penulis telah ditulis dalam naskah dan juga dalam sistem OJS. Salah satu penulis yang berhubungan melalui sistem OJS ditentukan sebagai penulis korespondensi.		
2.3	Semua penulis harus menyertakan email dalam naskah dan sistem OJS.		
2.4	Informasi afiliasi ditulis dalam 1 baris pada teks, paling sedikit menunjukkan nama lembaga dan negara.		
2.5	Judul naskah ditulis menggunakan 'Capitalize Each Word', kecuali kata penghubung menggunakan huruf kecil.		
2.6	Tempat dan waktu penelitian tidak perlu dicantumkan dalam judul kecuali penelitian dikhususkan untuk waktu dan tempat tertentu.		
3	ABSTRAK dan KATA KUNCI		
3.1	Abstrak terdiri dari 200 hingga maksimal 250 kata.		
3.2	Abstrak disusun secara terstruktur. Terdiri dari Latar Belakang, Tujuan, Metode, Hasil, dan Kesimpulan		
3.3	Jumlah kata kunci minimal 5 kata, dipisahkan dengan tanda titik		

	koma dan bukan merupakan potongan dari judul.		
3.4	Semua kata kunci harus ditulis dalam abstrak.		
3.5	Kata kunci disusun menurut abjad.		
3.6	Kombinasi semua kata kunci mewakili konten, kontribusi, atau tujuan teks.		
3.7	Kata kunci dapat berupa kategori subjek penelitian. Hindari menggunakan kata sifat sebagai kata kunci dalam teks.		
4	PENDAHULUAN		
4.1	Telah menjelaskan konteks / objek penelitian dan fokus mengikuti cakupan judul.		
4.2	Telah menjelaskan mengapa penelitian ini penting.		
4.3	Ini termasuk kesenjangan penelitian yang ada, menjelaskan perbedaan unik antara penelitian ini dibandingkan dengan penelitian serupa sebelumnya.		
4.4	Memiliki tujuan / kontribusi penelitian tertulis secara eksplisit.		
4.5	Panjang bagian ini antara 300 - 700 kata.		
5	STUDI LITERATUR		
5.1	Untuk mendukung penelitian yang sedang dilakukan, peneliti harus menelaah, mengupas dan memberikan beberapa penelitian terdahulu. Dalam hal ini minimal 3 penelitian terdahulu yang mutakhir dengan maksimal 5 tahun ke belakang.		
6	METODE		
6.1	Telah mengutip sumber yang membantu penulis dalam pemilihan metode.		
6.2	Telah menjelaskan bagaimana data diperoleh.		
6.3	Prosedurnya ditulis secara kronologis dan jelas.		
6.4	Penulis telah menulis bagaimana hasil akan diukur, diuji, dan dievaluasi.		
7	HASIL DAN PEMBAHASAN		
7.1	Hasil penelitian ditulis dalam urutan logis, sesuai alur penelitian.		
7.2	Hasil penelitian disajikan dalam bentuk naratif / tekstual, tabel, atau gambar dalam bentuk grafik atau diagram. Hindari menampilkan data mentah.		
7.3	Pembahasan Penjelasan tentang hasil dan analisis mendalam telah ditulis.		
7.4	Perbandingan hasil dengan penelitian lain telah ditulis.		
7.5	Batasan studi atau ancaman terhadap validitas telah ditulis.		
8	KESIMPULAN		
8.1	Kontribusi naskah dan jawaban atas pertanyaan penelitian telah dituangkan secara eksplisit.		
8.2	Kesimpulan sudah dituliskan dalam 1 paragraf utuh		
9	UCAPAN TERIMA KASIH		
9.1	Jika merupakan penelitian yang didanai, informasi tentang lembaga pemberi dana harus diberikan bersama dengan nomor kontrak (jika ada).		
10	REFERENCES		
10.1	Selalu gunakan alat kutipan otomatis untuk mengelola referensi		

	Anda. Pastikan Anda sudah memeriksa ulang metadata kutipan, alat kutipan otomatis biasanya memiliki kesalahan metadata. Gunakan Mendeley untuk membuat Kutipan & Referensi / Bibliografi. Gunakan gaya kutipan Edisi APA 6.		
10.2	Pastikan semua informasi yang terkait dengan referensi ditulis dengan lengkap (mis. Siapa, Dimana, dan Kapan). Periksa dari sumber utama untuk informasi ini. tulis nama jurnal tanpa singkatan. harap sertakan DOI jika tersedia.		
10.3	Mengutip minimal 15 referensi. minimal 80% berasal dari artikel jurnal / konferensi. minimal 80% harus diterbitkan dalam 5 tahun terakhir.		
10.4	Lebih dari 50% referensi harus berasal dari sumber yang kredibel. misalnya Owner, ScienceDirect, IEEE Explorer, ACM Digital Library, Springer Link, Scopus, Clarivate Analytics, dan jurnal terakreditasi nasional Indonesia (SINTA 1/2).		
11	Tabel dan Gambar		
11.1	Sudah menggunakan format yang sesuai dalam menampilkan gambar dan tabel.		
11.2	Gambar atau grafik harus cukup jelas, sehingga tidak bergantung pada teks.		
11.3	Resolusi gambar minimum tidak ditentukan, tetapi harus terlihat ketika dipublikasikan secara online (berwarna) atau dicetak (hitam dan putih).		
11.4	Gambar tidak memiliki masalah hak cipta.		
11.5	Semua gambar dan tabel telah direferensikan dalam naskah.		
11.6	Jumlah maksimum gambar dan tabel masing-masing adalah 10. memberikan alasan kepada editor jika Anda perlu menampilkan lebih dari jumlah yang disarankan.		
12	Languages		
12.1	Tidak ada kesalahan ejaan, kesalahan ketik, dan kesalahan tata bahasa. Anda dapat menggunakan alat bantu seperti Grammarly atau rekomendasi yang disediakan oleh MS Word untuk pengecekan bahasa.		
12.2	Gunakan kalimat langsung dan pendek. Kalimat yang panjang tidak membuat tulisan menjadi lebih profesional; mereka hanya membingungkan pembaca.		

OUTLINING HOW THE ISSUES ARE ADDRESSED

Title of paper:

1. Address all the concerns/recommendations of the reviewers.
2. All amendments made are to be **highlighted in yellow color** in the revised paper.

Reviewer # 1

Final Recommendation (*choose one according to the reviewer's comments and cross unnecessary*) :

- Accepted without modification
- Accepted with minor corrections
- Accepted with major modification
- Rejected

Comments	Addressed (Y/N)	Reply/Action taken

(Please add more rows if needed)

Reviewer # 2



Final Recommendation (*choose one according to the reviewer's comments and cross unnecessary*) :

- Accepted without modification
- Accepted with minor corrections
- Accepted with major modification
- Rejected

Comments	Addressed (Y/N)	Reply/Action taken

Reviewer # 3

Final Recommendation (*choose one according to the reviewer's comments and cross unnecessary*) :

- Accepted without modification
- Accepted with minor corrections
- Accepted with major modification
- Rejected

Comments	Addressed (Y/N)	Reply/Action taken

Meskipun hasil penelitian menunjukkan bahwa algoritma *Probabilistic Naive Bayes* memiliki kinerja yang baik dalam mengklasifikasikan sentimen ulasan produk *e-commerce*, penelitian ini memiliki beberapa keterbatasan. Penelitian ini hanya mengklasifikasikan sentimen ke dalam dua kelas, yaitu positif dan negatif, sehingga belum mencakup sentimen netral. Selain itu, asumsi independensi antar fitur pada algoritma *Naive Bayes* menyebabkan keterbatasan dalam menangkap konteks kata yang saling bergantung, terutama pada ulasan dengan struktur bahasa yang kompleks dan penggunaan bahasa informal.

VII. REFERENSI

- Ramadhan, B. Z., Adam, R. I., & Maulana, I. (2022). Analisis sentimen ulasan pada aplikasi *e-commerce* menggunakan algoritma Naïve Bayes. *Journal of Applied Informatics and Computing*, 6(2), 220–225.
- Susanti, A. R., Wijaya, F., & Lestari, D. (2024). Sentiment analysis of user reviews of *e-commerce* platforms using TF-IDF and Naïve Bayes. *International Journal of E-commerce Studies*, 8(1), 33–41.
- Shetty, A. M., & Manjaiah, D. H. (2024). Analyzing sentiments in *e-commerce*: Techniques, applications, and challenges. *International Journal of Science and Research Archive*, 12(2), 2307–2320.
- Godia, A., & Tiwari, L. K. (2025). Sentiment analysis and classification of product reviews using NLP and machine learning. *Proceedings of ICACCS*.
- Prihartono, W., Nugroho, A., & Setiawan, R. (2025). Analisis sentimen ulasan pengguna aplikasi menggunakan Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(1), 45–52.
- Al Ghazali, N. M., & Sibaroni, Y. (2025). Sentiment classification in *e-commerce* using Naïve Bayes. *Jurnal Ilmu Pengetahuan dan Informatika*, 10(2), 115–123.