

Analisis Sentimen Dalam Tokopedia Terhadap Ulasan Pelanggan Menggunakan Metode Naive Bayes

¹*Rayhan Ferdiansyah, ²Kevin Risky Abadi, ³Erna Daniati

¹⁻³Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail : ¹rayzen415@gmail.com ²kevinrisky18@gmail.com, ³ernadaniati@unpkediri.ac.id

Penulis Korespondens : Erna Daniati, M.Kom

Abstrak— Penelitian ini membahas klasifikasi sentimen ulasan pengguna Tokopedia menggunakan algoritma Naïve Bayes. Semakin banyaknya ulasan konsumen di e-commerce mendorong perlunya sistem otomatis untuk memahami opini pelanggan secara efisien. Dataset diperoleh dari Kaggle dan berisi komentar pengguna terhadap produk masker kesehatan. Proses analisis melibatkan pembersihan teks, tokenisasi, ekstraksi fitur menggunakan TF-IDF, serta klasifikasi sentimen menjadi positif dan negatif. Algoritma Naïve Bayes digunakan karena kemampuannya yang sederhana namun akurat dalam klasifikasi teks. Hasil evaluasi menunjukkan model menghasilkan akurasi tinggi dan mampu mengidentifikasi kecenderungan opini pengguna. Temuan ini diharapkan dapat dimanfaatkan oleh pelaku e-commerce dalam pengambilan keputusan berbasis data pelanggan.

Kata Kunci: Analisis Sentimen, Tokopedia, Naive Bayes, Ulasan Tokopedia, Machine Learning

Abstract— This study focuses on classifying user sentiment in Tokopedia product reviews using the Naïve Bayes algorithm. The growing volume of consumer feedback in e-commerce platforms highlights the need for automated systems to understand public opinion efficiently. The dataset, sourced from Kaggle, contains user reviews of healthcare mask products. The analysis includes text preprocessing, tokenization, feature extraction using TF-IDF, and sentiment classification into positive and negative categories. Naïve Bayes was chosen for its simplicity and effectiveness in text classification. Evaluation results show high accuracy and clear identification of user sentiment trends. The findings are expected to support e-commerce stakeholders in data-driven decision-making.

Keyword: Sentiment Analysis, Tokopedia, Naive Bayes, Tokopedia Reviews, Machine Learning

I. PENDAHULUAN

Pertumbuhan e-commerce di Indonesia, termasuk Tokopedia, telah mendorong peningkatan interaksi digital antara konsumen dan penjual, khususnya melalui fitur ulasan produk. Ulasan tersebut memuat opini pelanggan yang mencerminkan tingkat kepuasan terhadap layanan maupun kualitas produk. Namun, seiring meningkatnya jumlah ulasan, dibutuhkan pendekatan otomatis untuk menganalisis opini konsumen secara efisien.

Analisis sentimen merupakan salah satu pendekatan populer untuk mengklasifikasikan opini pengguna menjadi kategori positif dan negatif. Salah satu metode yang banyak digunakan adalah Naïve Bayes, karena kemampuannya yang sederhana, efisien, dan cukup akurat dalam memproses data teks. Metode ini telah banyak digunakan dalam studi terdahulu dengan berbagai hasil memuaskan

Penelitian oleh Ernawati dan Cahyati menunjukkan bahwa algoritma ini dapat mengklasifikasikan sentimen ulasan masker di Tokopedia dengan akurasi 88% [1] Mulyani juga

mencatat performa serupa pada produk masker Sensi Duckbill [2], sementara Hajaroh menerapkan pendekatan tiga kelas dan memperoleh akurasi sebesar 75,9% [3]. Apriani dan Gustian menunjukkan akurasi 97,13% dalam ulasan aplikasi Tokopedia di Google Play [4], dan Azra melaporkan 86,95% pada ulasan produk teknologi [5].

Meski banyak digunakan, masih terdapat celah yang dapat dijadikan dasar kebaruan. Sebagian besar penelitian terdahulu fokus pada produk ternama atau kategori spesifik. Penelitian ini menggunakan data ulasan produk masker non-merek populer, yang mencerminkan preferensi konsumen yang lebih umum. Selain itu, pelabelan sentimen dilakukan secara otomatis berdasarkan rating, sehingga lebih efisien dibandingkan metode manual. Proses preprocessing juga dibuat minimal agar tetap sederhana dan dapat direproduksi secara luas.

Tujuan utama penelitian ini adalah membangun model klasifikasi sentimen berbasis Naïve Bayes terhadap ulasan Tokopedia. Secara khusus, penelitian ini mencakup pemahaman struktur data, pelabelan otomatis berdasarkan rating, preprocessing teks, ekstraksi fitur menggunakan TF-IDF, pelatihan model, evaluasi performa, dan visualisasi hasil klasifikasi.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi teks berbasis algoritma Naïve Bayes. Tujuan utamanya adalah mengklasifikasikan sentimen ulasan pengguna Tokopedia terhadap produk masker kesehatan ke dalam dua kategori, yaitu positif dan negatif.

Dataset yang digunakan diperoleh dari platform Kaggle dan disusun oleh Zahra Maharani Aulia. Data tersebut terdiri dari lebih dari 503 data ulasan dalam Bahasa Indonesia, lengkap dengan skor rating produk. Ulasan dengan rating ≥ 4 diberi label positif, sedangkan rating < 4 diberi label negatif. Format data berupa file Excel (.xlsx) dengan kolom utama berupa komentar pengguna (review) dan rating (rate).

Pengolahan dan analisis data dilakukan menggunakan Python melalui lingkungan Jupyter Notebook, serta RapidMiner sebagai alat bantu visualisasi dan pembanding. Pustaka Python yang digunakan meliputi: pandas dan numpy untuk manipulasi data; nltk untuk preprocessing teks (tokenisasi, stopword removal); scikit-learn untuk TF-IDF, Naïve Bayes, dan evaluasi model; serta matplotlib dan seaborn untuk visualisasi.

Desain prosedur penelitian terdiri dari tahapan berikut:

1. Import dataset dan pemahaman struktur data
2. Labeling otomatis berdasarkan rating pengguna
3. Pembersihan teks, penghapusan karakter khusus, dan tokenisasi
4. Ekstraksi fitur teks menggunakan TF-IDF
5. Pembagian data latih dan uji dengan rasio 70:30
6. Pelatihan model menggunakan algoritma Multinomial Naïve Bayes
7. Evaluasi model melalui metrik: akurasi, precision, recall, F1-score, dan confusion matrix
8. Visualisasi hasil klasifikasi

Preprocessing teks yang sederhana tetap mampu menghasilkan akurasi yang baik. Hal ini sejalan dengan temuan dalam penelitian sebelumnya yang menunjukkan bahwa proses tokenisasi dan stopword removal memiliki dampak signifikan terhadap kinerja model [6]. TF-IDF digunakan karena efektif dalam merepresentasikan fitur kata dalam klasifikasi teks [7].

Naïve Bayes dipilih karena cocok untuk dataset kategorikal dengan volume sedang hingga besar [8], dan memiliki tingkat efisiensi yang tinggi dalam klasifikasi opini publik. Penelitian oleh Prasetya dan Mustofa juga membuktikan bahwa kombinasi Naïve Bayes dan DBSCAN

menghasilkan sistem pendukung keputusan yang akurat [9], meskipun dalam penelitian ini tidak digunakan metode clustering. Visualisasi hasil seperti word cloud dan grafik batang disarankan untuk memperkuat interpretasi terhadap persebaran kata dan sentimen [10].

Desain dan tahapan yang digunakan memungkinkan replikasi penuh oleh peneliti lain, karena seluruh proses dilakukan dengan pustaka open-source dan dataset publik. Model, kode, dan konfigurasi dapat digunakan ulang untuk penelitian sejenis atau sebagai baseline untuk pendekatan lanjutan.

III. HASIL DAN PEMBAHASAN

3.1 Hasil Labeling dan Distribusi Sentimen

Dataset yang digunakan terdiri dari 503 ulasan pengguna Tokopedia terhadap produk masker kesehatan. Setelah dilakukan proses labeling otomatis berdasarkan nilai rating (≥ 4 sebagai sentimen positif dan < 4 sebagai negatif), diperoleh distribusi sentimen seperti pada Tabel 1 berikut:

Tabel 1. Distribusi Sentimen Berdasarkan Rating

Sentimen	Jumlah	Presentase
Positif	414	82,3%
Negatif	89	17,7%

Temuan ini menunjukkan bahwa mayoritas ulasan bersifat positif, mencerminkan kepuasan pengguna terhadap produk masker yang dibeli melalui Tokopedia. Temuan ini sejalan dengan hasil analisis sentimen Ramadhan., yang menunjukkan bahwa mayoritas ulasan pengguna Shopee terhadap produk fashion juga bersifat positif, dengan F1-score mencapai 95,35% [11].

3.2 Hasil Klasifikasi Menggunakan Naïve Bayes

Model klasifikasi Naïve Bayes dilatih menggunakan 70% data dan diuji menggunakan 30% sisanya. Hasil evaluasi performa model terhadap data uji ditampilkan pada Tabel 2 berikut:

Tabel 2. Evaluasi Kinerja Model

Metrik	Nilai
Akurasi	89,0%
Precision	0.85
Recall	0.86
F1-Score	0.85

Model menunjukkan performa cukup baik dan seimbang dalam mengklasifikasikan kedua kelas. Akurasi yang diperoleh melampaui hasil dari penelitian Hajaroh [3], yang melaporkan akurasi sebesar 75,9% pada klasifikasi tiga kelas. Sementara itu, penelitian oleh Sera dan Putra menunjukkan bahwa preprocessing NLP mampu meningkatkan akurasi klasifikasi hingga 88% [12], mendukung efektivitas preprocessing sederhana yang diterapkan pada penelitian ini.

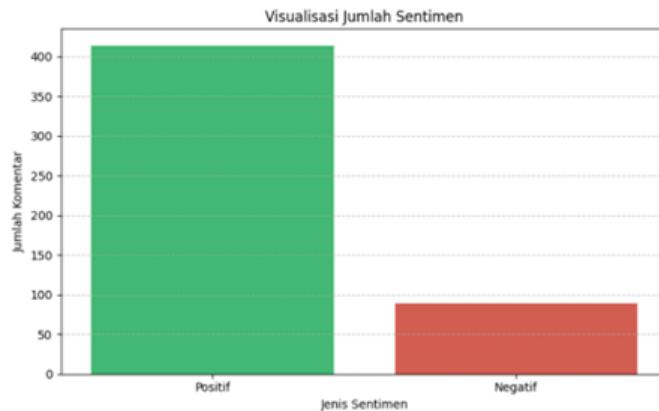
3.3 Visualisasi Distribusi Sentimen

Distribusi sentimen divisualisasikan menggunakan grafik batang dan diagram lingkaran:

- Grafik batang memperlihatkan jumlah ulasan positif jauh lebih banyak dibandingkan negatif.
- Diagram lingkaran menegaskan dominasi sentimen positif dalam dataset.

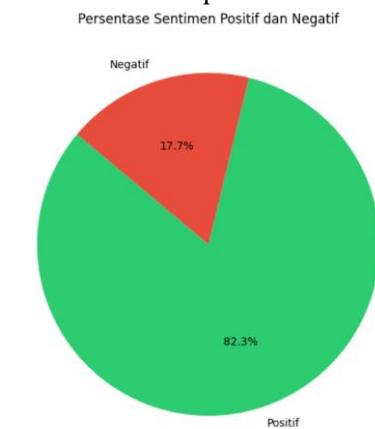
Visualisasi seperti word cloud dan grafik batang juga digunakan dalam studi lain untuk memudahkan pemahaman distribusi opini publik pada ulasan e-commerce [13].

- Grafik batang menunjukkan jumlah ulasan positif jauh lebih tinggi dibandingkan negatif.



Gambar 1. Grafik Jumlah Sentimen

- Diagram pie memperjelas dominasi sentimen positif dalam dataset.



Gambar 2. Diagram Lingkaran Persentase Sentimen

3.4 Pembahasan Temuan

Temuan utama dari penelitian ini adalah sebagai berikut:

- Sentimen positif mendominasi ulasan pengguna Tokopedia terhadap produk masker.
- Model Naïve Bayes menunjukkan performa baik (akurasi 89%) dalam klasifikasi teks dengan preprocessing minimal.
- Pendekatan sederhana menggunakan TF-IDF dan Naïve Bayes tetap mampu menghasilkan hasil kompetitif dibanding metode kompleks.

Hal ini diperkuat oleh penelitian brand skincare lokal review oleh Kaswili., yang menyatakan bahwa Naïve Bayes dapat menjadi baseline klasifikasi sentimen yang andal dengan tingkat akurasi yang kompetitif [15].

3.5 Keterbatasan dan Arah Penelitian Selanjutnya

Beberapa batasan yang perlu dicermati:

- Dataset hanya mencakup satu jenis produk (masker), belum merepresentasikan produk e-commerce lain.

- Labeling berdasarkan rating bisa menimbulkan bias, misalnya rating rendah dengan komentar positif.
- Hanya satu model klasifikasi yang diuji (Naïve Bayes), tanpa bandingan dengan model seperti SVM atau Random Forest.

Selain itu, Naïve Bayes memiliki performa yang bersaing dibandingkan dengan algoritma lainnya [14]

3.6 Implikasi dan Arah Penelitian Selanjutnya

Hasil penelitian ini berpotensi diterapkan dalam sistem evaluasi otomatis untuk membantu pelaku e-commerce memahami sentimen konsumen. Sistem ini dapat diintegrasikan ke dashboard manajemen produk sebagai alat bantu pengambilan keputusan.

Rekomendasi penelitian selanjutnya:

- Melibatkan berbagai jenis produk untuk memperoleh representasi yang lebih luas.
- Membandingkan beberapa algoritma klasifikasi.
- Menerapkan analisis sentimen berbasis konteks dan aspek linguistik seperti emosi dan ironi.

IV. KESIMPULAN

Penelitian ini bertujuan untuk membangun model klasifikasi sentimen terhadap ulasan pengguna Tokopedia menggunakan algoritma Naïve Bayes. Berdasarkan hasil implementasi dan evaluasi, dapat disimpulkan bahwa metode Naïve Bayes efektif dalam mengklasifikasikan opini pengguna ke dalam dua kategori sentimen utama, yaitu positif dan negatif. Dengan pendekatan sederhana melalui teknik TF-IDF dan preprocessing minimal, model mampu mencapai akurasi sebesar 89%, menunjukkan performa yang kompetitif. Hasil ini memperkuat peran algoritma klasifikasi berbasis probabilistik dalam menganalisis data teks pada konteks e-commerce di Indonesia. Penelitian ini memberikan kontribusi terhadap pengembangan sistem klasifikasi opini otomatis yang praktis dan dapat diintegrasikan dalam evaluasi produk digital secara luas. Selain itu, pendekatan ini membuka peluang penerapan lebih lanjut dalam pengolahan data berbasis bahasa alami untuk mendukung pengambilan keputusan berbasis data.

DAFTAR PUSTAKA

- [1] A. Ernawati, A. Ofta Sari, S. Nurhaliza Sofyan, M. Iqbal, and R. Farta Wijaya, “Bulletin of Information Technology (BIT) Implementasi Algoritma Naïve Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan,” vol. 4, no. 4, pp. 533–543, 2023, doi: 10.47065/bit.v3i1.
- [2] E. Dewi, S. Mulyani, A. T. Hidayatuloh, and T. Agustiawan, “SEMINAR NASIONAL CORISINDO Analisis Sentimen Ulasan Produk Pada Top Brand Produk Masker Di Tokopedia Menggunakan Naive Bayes.”
- [3] T. Suprapti and R. Narasati, “IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN ULASAN PRODUK MAKANAN DAN MINUMAN DI TOKOPEDIA,” 2024.
- [4] R. Apriani *et al.*, “ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA,” 2019.

- [5] S. Azra and R. Tahara Shita, “3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta,” 2023.
- [6] E. Daniati and H. Utama, “ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER,” 2023.
- [7] H. Utama, E. Daniati, and A. Masruro, “WEAK SUPERVISION DENGAN PENDEKATAN LABELING FUNCTION UNTUK ANALISIS SENTIMEN PADA TWITTER,” 2024. [Online]. Available: <https://subset.id/index.php/IJCSR>
- [8] A. Tholib, M. Noer Fadli Hidayat, Supriyono, R. Wulanningrum, and E. Daniati, “Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms,” Sep. 2023.
- [9] E. Daniati, *Decision Support Systems to Determining Programme for Students Using DBSCAN And Naive Bayes*. IEEE, 2019.
- [10] T. C. Adisti, E. Daniati, and A. Ristiyawan, “ANALISIS SENTIMEN UJARAN KEBENCIAN PADA TWEET DI TWITTER,” 2025.
- [11] J. Minfo Polgan *et al.*, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” 2022.
- [12] E. Sera, S. Komputer, H. Makassar, and B. Riset dan Inovasi, “Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing Sentiment Analysis of Product Reviews on E-Commerce’s Bukalapak Using Natural Language Processing.” [Online]. Available: <https://www.kaggle.com/>.
- [13] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 233–242, Dec. 2021, doi: 10.29408/edumatic.v5i2.4089.
- [14] E. Hasibuan and E. A. Heriyanto, “View of ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER”.
- [15] K. Sriwenda Putri, R. Setiawan, and A. Pambudi, “ANALISIS SENTIMEN TERHADAP BRAND SKINCARE LOKAL MENGGUNAKAN NAÏVE BAYES CLASSIFIER,” 2023.