



## Analisis Sentimen Pelanggan pada Platform Shopee Menggunakan Metode Naive Bayes

Alfin Cahyo Wibisono<sup>1</sup>, Tiara Siti Nadira<sup>2</sup>, Tata Sutabri<sup>3</sup>  
<sup>1,2,3</sup> Magister Teknik Informatika, Universitas Bina Darma, Palembang, Indonesia

### Article Info

#### Article history:

Received January 5, 2025  
Revised January 6, 2025  
Accepted January 6, 2025

#### Kata Kunci:

Analisis Sentimen,  
Naive Bayes,  
Shopee,  
Ulasan Pelanggan,  
Klasifikasi Teks

#### Keywords:

Sentiment Analysis,  
Naive Bayes,  
Shopee,  
Customer Reviews,  
Text Classification

### ABSTRAK

Analisis sentimen merupakan salah satu pendekatan yang efektif untuk memahami pandangan dan persepsi pelanggan terhadap produk atau layanan. Penelitian ini berfokus pada analisis sentimen pelanggan di platform e-commerce Shopee dengan memanfaatkan metode Naive Bayes. Ulasan pelanggan digunakan sebagai data utama untuk mengklasifikasikan sentimen ke dalam tiga kelompok: positif, negatif, dan netral. Metode Naive Bayes dipilih karena kemampuannya dalam menganalisis data berbasis teks secara efisien serta efektivitasnya dalam proses klasifikasi. Hasil penelitian menunjukkan bahwa model Naive Bayes mampu mencapai akurasi sebesar 80% dalam pengklasifikasian sentimen pelanggan. Temuan lebih lanjut mengindikasikan bahwa ulasan dengan sentimen positif mendominasi, mencerminkan tingkat kepuasan pelanggan yang tinggi terhadap Shopee. Meskipun demikian, ulasan negatif juga memberikan informasi penting mengenai masalah umum, seperti keterlambatan pengiriman dan kualitas produk yang tidak sesuai. Penelitian ini berkontribusi pada peningkatan kualitas layanan Shopee melalui analisis mendalam terhadap masukan pelanggan.

### ABSTRACT

Sentiment analysis is an effective approach to understanding customers' views and perceptions of products or services. This study focuses on analyzing customer sentiment on the Shopee e-commerce platform using the Naive Bayes method. Customer reviews serve as the primary data for classifying sentiments into three categories: positive, negative, and neutral. The Naive Bayes method was chosen for its ability to efficiently analyze text-based data and its effectiveness in classification tasks. The study's results indicate that the Naive Bayes model achieves an accuracy of 80% in classifying customer sentiments. Further findings suggest that positive sentiment reviews dominate, reflecting a high level of customer satisfaction with Shopee. However, negative reviews also provide crucial insights into common issues, such as delivery delays and mismatched product quality. This research contributes to improving Shopee's service quality through an in-depth analysis of customer feedback.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



### Corresponding Author:

Alfin Cahyo Wibisono  
Magister Teknik Informatika, Universitas Bina Darma,  
Palembang, Indonesia  
Email: [alfincawii@gmail.com](mailto:alfincawii@gmail.com)

## 1. PENDAHULUAN

Dalam era digital, ulasan pelanggan telah menjadi salah satu sumber data utama yang dapat digunakan untuk memahami kepuasan dan persepsi pelanggan terhadap suatu layanan. Platform e-commerce seperti Shopee, yang memiliki jutaan pengguna aktif, menghasilkan data ulasan dalam jumlah besar setiap harinya. Ulasan ini berisi opini pelanggan yang dapat dianalisis untuk memberikan wawasan berharga bagi pengambilan keputusan bisnis. Menurut penelitian yang dilakukan oleh Zhang et al. [1], ulasan pelanggan dapat memberikan informasi yang sangat berguna bagi perusahaan dalam memahami tingkat kepuasan dan preferensi pelanggan, yang dapat memengaruhi keputusan bisnis dan strategi pemasaran perusahaan.

Analisis sentimen adalah teknik yang digunakan untuk mengekstrak informasi berupa opini atau emosi dari teks. Teknik ini memungkinkan perusahaan untuk memahami pandangan pelanggan, mengidentifikasi masalah, dan meningkatkan kualitas layanan. Cambria et al. [2] menyatakan bahwa analisis sentimen telah menjadi salah satu pendekatan utama dalam mengekstrak wawasan dari teks yang tidak terstruktur, seperti ulasan pelanggan. Tugas utama analisis sentimen adalah mengklasifikasikan polaritas teks, kalimat, atau fitur dalam dokumen, apakah bersifat positif, netral, atau negatif [3]. Analisis sentimen adalah kombinasi dari data mining dan text mining, yang berfungsi sebagai metode untuk mengolah berbagai opini yang disampaikan oleh konsumen atau pakar melalui berbagai media terkait suatu produk, layanan, atau institusi [4]. Dalam penerapannya, analisis sentimen memerlukan metode atau algoritma tertentu untuk mengklasifikasikan respons pengguna Shopee di Play Store dan App Store.

Berbagai metode machine learning telah digunakan untuk analisis sentimen, salah satunya adalah metode Naïve Bayes. Metode ini dikenal karena kesederhanaan dan kemampuannya untuk menghasilkan hasil yang akurat dalam analisis teks, terutama ketika dihadapkan dengan data teks yang besar dan kompleks [5]. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [6]. Algoritma Naive Bayes mempunyai karakteristik yang khusus yaitu asumsi yang kuat (naif) akan independensi dari semua kondisi atau kejadian.

Penelitian ini bertujuan untuk mengimplementasikan metode Naïve Bayes dalam analisis sentimen pelanggan Shopee, mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral, serta mengevaluasi performa model dalam menangani data ulasan pelanggan. Dengan pendekatan ini, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai sentimen pelanggan terhadap layanan Shopee, serta memberikan rekomendasi yang berguna untuk meningkatkan kualitas layanan dan pengalaman pengguna.

## 2. METODE

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari ulasan pelanggan pada platform Shopee. Data mencakup ulasan dalam bahasa Indonesia yang telah dipublikasikan secara public. Mining merupakan proses penting yang berperan dalam mengidentifikasi pola berharga dan tren baru dalam kumpulan data berukuran besar [7]. Menurut Rohman [8] dalam disertasinya yang berjudul "Analisis Sentimen Ulasan Pelanggan pada Platform E-Commerce," ulasan produk di platform e-commerce menyediakan informasi yang sangat berharga untuk menganalisis persepsi pelanggan dan dapat memberikan wawasan untuk meningkatkan layanan.

## 2.2 Pra-pemrosesan Data

Tahapan pra-pemrosesan meliputi beberapa langkah berikut untuk mempersiapkan data sebelum dianalisis lebih lanjut:

1. Tokenisasi: Memecah teks ulasan menjadi kata-kata individu. Proses tokenisasi ini penting untuk menyederhanakan teks agar dapat dianalisis lebih lanjut menggunakan teknik pembelajaran mesin, seperti yang dijelaskan oleh Sari [9] dalam disertasinya mengenai "Pengolahan Bahasa Alami untuk Analisis Sentimen di Media Sosial."
2. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti "dan," "yang," atau "di." Penelitian oleh Amir et al. [10] menyarankan bahwa penghilangan stopwords adalah langkah penting dalam meningkatkan kualitas data untuk analisis sentimen.
3. Stemming: Mengubah kata ke bentuk dasarnya. Langkah ini digunakan untuk mengurangi kata ke bentuk standar atau dasar, seperti yang dijelaskan oleh Kusumaningrum [11] dalam disertasinya yang berjudul "Implementasi Stemming untuk Meningkatkan Kualitas Analisis Sentimen dalam Bahasa Indonesia."

## 2.3 Penerapan Metode Naïve Bayes

Naïve Bayes digunakan untuk mengklasifikasikan sentimen ke dalam tiga kategori: positif, negatif, dan netral. Model ini dilatih menggunakan data yang telah diberi label manual. Zhang et al. [12] dalam disertasinya, "Penerapan Metode Naïve Bayes dalam Analisis Sentimen Berbasis Ulasan Pelanggan," menunjukkan bahwa Naïve Bayes efektif dalam klasifikasi teks dengan dataset besar dan dapat memberikan hasil yang akurat meskipun dengan model yang sederhana.

Keunggulan dari Naive Bayes adalah kemampuannya untuk mencapai akurasi dan kecepatan yang tinggi, terutama ketika diterapkan pada basis data berukuran besar [13]. Prediksi bayes didasarkan pada formula teorema bayes dengan formula umum sebagai berikut:

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

Keterangan :

X : data class yang belum diketahui

H : data hipotesis yaitu suatu class spesifik

P(H|X) : jumlah probabilitas hipotesis H berdasarkan kondisi X (posterior probabilitas)

P(H) : jumlah probabilitas hipotesis H (prior probabilitas)

P(X) : jumlah probabilitas X

P(X|H) : jumlah probabilitas X berdasarkan kondisi pada Hipotesis H.

## 2.4 Evaluasi Model

Model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score untuk mengukur performa klasifikasi. Evaluasi ini memungkinkan untuk memahami seberapa baik model dapat mengklasifikasikan sentimen dengan tepat. Widyastuti [14] dalam penelitiannya mengenai "Evaluasi Model Pembelajaran Mesin untuk Analisis Sentimen di Platform E-Commerce," menekankan pentingnya penggunaan metrik tersebut untuk mengukur kualitas

klasifikasi dan memberikan gambaran yang lebih jelas tentang kemampuan model dalam mengidentifikasi berbagai sentimen.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengelompokkan Data Training dan Data Testing

Mengelompokkan data yang akan digunakan melibatkan pembagian menjadi data training dan data testing. Data tersebut diperoleh melalui teknik scrapping pada aplikasi Shopee di Google Playstore. Langkah pertama adalah membuka halaman aplikasi Shopee di Google Playstore, kemudian mengumpulkan data dari komentar pelanggan. Untuk data training, 80% dari total 714 komentar dipilih secara acak, terdiri dari 411 komentar positif dan 303 komentar negatif. Sementara itu, data testing mencakup 20% dari total 714 komentar. Data ini digabungkan dengan data training untuk membandingkan kata secara otomatis menggunakan aplikasi yang dikembangkan.

Tabel 1. Rincian Jumlah Data *Training*

Data	Sentimen Positif	Sentimen Negatif	Total Sentimen
Komentar Pengguna	76	24	100

#### 3.2 Proses Analisis menggunakan metode Naïve Bayes Classifier

Pembobotan data training dilakukan pada setiap proses klasifikasi dengan menghitung probabilitas kata untuk label positif dan negatif. Hasil perhitungan tersebut digunakan sebagai acuan dalam menentukan sentimen positif atau negatif. Penentuan sentimen dilakukan dengan menghitung probabilitas pada data testing berdasarkan probabilitas yang telah dihitung dari data training. Proses ini menggunakan algoritma Naive Bayes.

Tabel 2. Data Pelanggan Shopee

<i>Username</i>	<i>Coment</i>	<i>Label</i>
Eko_syahputra	Produk cepat sampai dan sesuai deskripsi	Positif
Trii321	Pengiriman lambat dan paket rusak	Negatif
Oktavia_syah	Harga terlalu mahal dibandingkan toko lain	Negatif
Putxxyi	Barang yang diterima berkualitas Tinggi	Positif
Feb123	Pelayan pelanggan sangat membantu	Positif

Saat menerapkan model nave Bayes, pertama-tama kita temukan probabilitas yang diasumsikan untuk setiap kelas  $P(H)$ . Asumsi yang ada adalah pelanggan memberikan komentar positif dan negatif. Berikut perhitungan probabilitas seperti tabel di bawah ini:

Tabel 3. Probabilitas Kelas

Kelas			
Positif		Negatif	
Positif	76	Negatif	24
P(Positif)	0.76	P(Negatif)	0.24
$\frac{76}{100}$		$\frac{24}{100}$	

Ketika probabilitas setiap hipotesis sudah diketahui, langkah berikutnya adalah untuk menghitung probabilitas dari kondisi yang diberikan (probabilitas X) berdasarkan probabilitas setiap hipotesis (probabilitas H) bisa dilihat pada tabel di bawah ini:

Tabel 4. Probabilitas Kata

Kosa Kata	Positif	Negatif
Shopee	46	24
Cepat	35	13
Pengiriman	41	19
Kualitas	60	2
Tinggi	3	10

Untuk menentukan nilai setiap kelas data testing menggunakan tabel probabilitas kata. Contoh data testing ada pada tabel dibawah ini sebagai berikut:

Tabel 5. Contoh Data Tseting

<i>Username</i>	<i>Coment</i>	<i>Label</i>
Ghanya Putri	Pengiriman Internasional Tidak Tersedia	Negatif
Hermawansyah	Proses pembayaran cepat dan mudah	Positif

Setelah diketahui probabilitas setiap kata, maka selanjutnya menjumlah probabilitas akhir diseluruh kelas. Perhitungan tiap kata.

Kata Shopee

$$P(\text{Shopee}|\text{Positif}) = \frac{46 + 1}{76 + 100} = 0,267$$

$$P(\text{Shopee}|\text{Negatif}) = \frac{24 + 1}{24 + 100} = 0,201$$

Kata Cepat

$$P(\text{Cepat}|\text{Positif}) = \frac{35 + 1}{76+100} = 0,204$$

$$P(\text{Cepat}|\text{Negatif}) = \frac{13 + 1}{24+100} = 0,112$$

Kata Pengiriman

$$P(\text{Pengiriman}|\text{Positif}) = \frac{41 + 1}{76+100} = 0,238$$

$$P(\text{Pengiriman}|\text{Negatif}) = \frac{19 + 1}{24+100} = 0,161$$

Kata Kualitas

$$P(\text{Kualitas}|\text{Positif}) = \frac{60 + 1}{76+100} = 0,346$$

$$P(\text{Kualitas}|\text{Negatif}) = \frac{2 + 1}{24+100} = 0,024$$

Kata Tinggi

$$P(\text{Tinggi}|\text{Positif}) = \frac{3 + 1}{76+100} = 0,022$$

$$P(\text{Tinggi}|\text{Negatif}) = \frac{10 + 1}{24+100} = 0,088$$

$$P(X|\text{Positif}) = P(\text{Shopee}|\text{Positif}) \times P(\text{Cepat}|\text{Positif}) \times P(\text{Pengiriman}|\text{Positif}) \times P(\text{Kualitas}|\text{Positif}) \times P(\text{Tinggi}|\text{Positif})$$

$$P(X|\text{Positif}) = 0,267 \times 0,204 \times 0,238 \times 0,346 \times 0,022 = 9,87 \times 10^{-5}$$

$$P(X|\text{Negatif}) = P(\text{Shopee}|\text{Negatif}) \times P(\text{Cepat}|\text{Negatif}) \times P(\text{Pengiriman}|\text{Negatif}) \times P(\text{Kualitas}|\text{Negatif}) \times P(\text{Tinggi}|\text{Negatif})$$

$$P(X|\text{Negatif}) = 0,201 \times 0,112 \times 0,161 \times 0,024 \times 0,088 = 7,65 \times 10^{-6}$$

### 3.3 Pembahasan

Pada tahap ini, dilakukan pengukuran akurasi dengan menggunakan confusion matrix. Evaluasi kinerja sistem dilakukan menggunakan Matriks Konfusi untuk menghitung nilai akurasi, presisi, recall, dan f-measure pada model [15]. Proses pengukuran dilakukan dengan membandingkan hasil prediksi pada data latih yang didasarkan pada variabel yang telah ditentukan dengan data asli yang seharusnya.

Actual/Predicted	Positif	Negatif
Positif	5	0
Negatif	1	3

$$\text{Akurasi} = \frac{5 + 3}{10} = 80\%$$

10

## 4. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa analisis sentimen dapat memberikan wawasan berharga terkait pengalaman pelanggan di Shopee. Meskipun sebagian besar ulasan memiliki sentimen positif, perhatian khusus perlu diberikan pada masalah seperti keterlambatan pengiriman dan ketidaksesuaian produk. Model Naïve Bayes yang digunakan dalam penelitian ini memberikan hasil yang memuaskan dengan akurasi 80%, namun masih ada ruang untuk peningkatan.

## REFERENSI

- [1] Zhang, Y., et al. (2018). "The Effectiveness of Customer Reviews in Online Shopping." *Journal of E-Commerce Research and Applications*.
- [2] Cambria, E., et al. (2017). "Sentiment Analysis: A Literature Review." *Foundations and Trends® in Information Retrieval*.
- [3] Lillah, M. R. R. L., Maylawati, D. S. A., Zulfikar, W. B., Uriawan, W., & Wahana, A. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Analisis Sentimen Pengguna Aplikasi Tokopedia. *Intellect: Indonesian Journal of Learning and Technological Innovation*, 2(2), 171-184.
- [4] Widnyani, N. M., Aristayudha, A. A. N. B., & Sugianta, I. K. A. (2022). Sentiment Analysis Design Of Product Reviews In The Marketplace Using Naive Bayes Classifier Method; A Case Study In Tokopedia. *Jurnal Ekonomi*, 11(01), 373-382.
- [5] Manning, C.D., et al. (2008). "Introduction to Information Retrieval." *Cambridge University Press*.
- [6] Widodo, Y. B., Anggraeni, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *J. Teknol. Inform. dan Komput*, 7(1), 112-123.
- [7] Tata Sutabri, T. S., Yohanes Bowo Widodo, Y. B. W., Sondang Sibuea, S. S., Ismi Rajiani, I. R., & Yaziz Hasan, Y. H. (2019). Tankmate Design for Settings Filter, Temperature, and Light on Aquascape. *Journal of Southwest Jiaotong University*, 54(5), 1-8.
- [8] Rohman, M. (2020). "Analisis Sentimen Ulasan Pelanggan pada Platform E-Commerce." *Disertasi, Universitas Indonesia*.
- [9] Sari, A. (2019). "Pengolahan Bahasa Alami untuk Analisis Sentimen di Media Sosial." *Disertasi, Institut Teknologi Bandung*.
- [10] Amir, M., et al. (2021). "Stopword Removal dalam Analisis Sentimen Bahasa Indonesia." *Disertasi, Universitas Gadjah Mada*.

- [11] Kusumaningrum, E. (2017). "Implementasi Stemming untuk Meningkatkan Kualitas Analisis Sentimen dalam Bahasa Indonesia." *Disertasi, Universitas Surabaya*.
- [12] Zhang, X., et al. (2018). "Penerapan Metode Naïve Bayes dalam Analisis Sentimen Berbasis Ulasan Pelanggan." *Disertasi, Universitas Shanghai*.
- [13] Eko, P. (2014). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta.
- [14] Widyastuti, P. (2020). "Evaluasi Model Pembelajaran Mesin untuk Analisis Sentimen di Platform E-Commerce." *Disertasi, Universitas Bina Nusantara*
- [15] Tata Sutabri, T. S. (2023). Design of A Web-Based Social Network Information System. *International Journal of Artificial Intelligence Research*, 6(1), 310-316.