

Analisis Sentimen Ulasan Produk Ponsel Pada E-commerce Menggunakan Algoritma Naïve Bayes

^{1*}Fadzilah Prayoganing Gusti, ²Danar Putra Pamungkas, ³Patmi Kasih

¹²³ Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹akuyoga63@gmail.com, ²danar@unpkediri.ac.id, ³fatkasih@gmail.com

Penulis Korespondens : Fadzilah Prayoganing Gusti

Abstrak— Pertumbuhan e-commerce di Indonesia telah mendorong peningkatan pembelian produk secara daring, termasuk ponsel. Namun, kurangnya transparansi kualitas produk menjadi tantangan bagi konsumen. Untuk mengatasi hal ini, penelitian ini menerapkan analisis sentimen terhadap ulasan pelanggan menggunakan algoritma *Naïve Bayes*. Data ulasan diambil dari platform Shopee dan melalui beberapa tahapan preprocessing seperti case folding, cleansing, tokenisasi, stopword removal, dan stemming. Sentimen ditentukan menggunakan *TextBlob*, lalu data dilatih dan diuji menggunakan *Naïve Bayes*. Dari 365 ulasan yang dianalisis, model menghasilkan akurasi sebesar 81%. Hasil ini menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan sentimen ulasan secara efektif, memberikan manfaat bagi konsumen dalam pengambilan keputusan serta bagi penjual dalam meningkatkan layanan.

Kata Kunci— e-commerce, analisis sentimen, *Naïve Bayes*, ponsel.

Abstract— The growth of e-commerce in Indonesia has led to an increase in online purchases, particularly of mobile phones. However, the lack of product quality transparency poses a challenge for consumers. To address this, this study applies sentiment analysis on customer reviews using the *Naïve Bayes* algorithm. Review data was collected from the Shopee platform and underwent several preprocessing stages including case folding, cleansing, tokenization, stopword removal, and stemming. Sentiment labels were assigned using *TextBlob*, followed by training and testing the data with *Naïve Bayes*. From 365 analyzed reviews, the model achieved an accuracy of 81%. These results indicate that the *Naïve Bayes* algorithm can effectively classify sentiment in reviews, benefiting consumers in making informed decisions and helping sellers enhance service quality.

Keywords— e-commerce, sentiment analysis, *Naïve Bayes*, mobile phones

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Pertumbuhan e-commerce di Indonesia meningkat pesat dalam beberapa tahun terakhir, ditandai dengan semakin populernya platform belanja online seperti Shopee. Shopee adalah aplikasi dan situs web belanja online yang menyediakan berbagai macam kategori produk seperti elektronik, pakaian, kebutuhan rumah tangga dan masih banyak lagi [1]. Produk elektronik, khususnya ponsel, menjadi salah satu kategori yang paling banyak dibeli secara daring. Konsumen memilih membeli ponsel secara online karena kemudahan akses, banyaknya pilihan produk, harga yang kompetitif, dan promosi menarik. Namun, di balik kemudahan tersebut, terdapat berbagai permasalahan yang kerap dihadapi pembeli.

Salah satu masalah utama adalah minimnya transparansi kualitas produk sebelum pembelian. Tidak seperti toko fisik, konsumen tidak dapat mencoba atau memeriksa langsung kondisi ponsel sebelum membeli. Oleh karena itu, ulasan pelanggan (review) menjadi sumber

informasi yang sangat penting untuk mengetahui pengalaman pengguna sebelumnya terkait produk maupun penjual. Namun demikian, ulasan yang tersedia sangat beragam, tidak terstruktur, dan sering kali menggunakan bahasa subjektif, sehingga menyulitkan calon pembeli untuk memahami sentimen umum dari pelanggan sebelumnya.

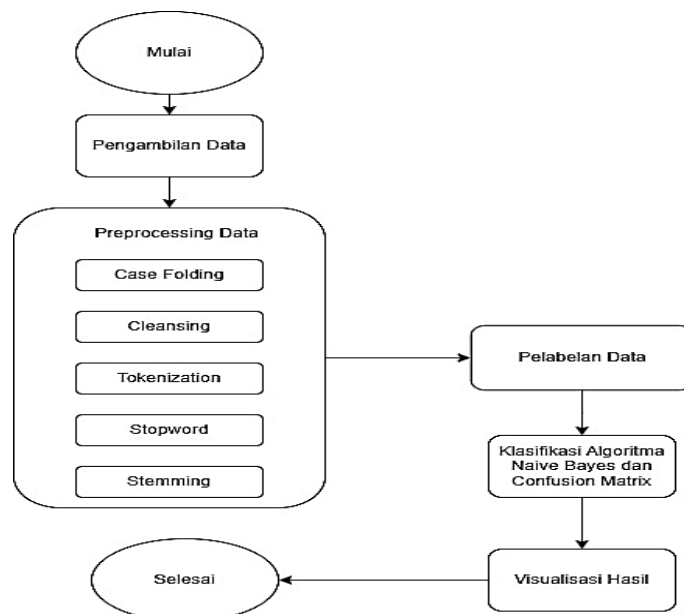
Untuk mengatasi permasalahan tersebut, diperlukan sebuah sistem yang mampu menganalisis sentimen pelanggan secara otomatis dari data ulasan yang tersedia. Analisis sentimen merupakan proses yang dilakukan untuk menyediakan informasi yang terdapat pada dataset tidak terstruktur. Proses ini merupakan proses komputasi yang dilakukan dengan cara memahami, mengekstraksi, dan mengolah data yang berbentuk tekstual secara otomatis sehingga mendapatkan informasi yang terdapat dalam opini atau perilaku seseorang [2]. Dengan mengetahui kecenderungan sentimen apakah positif, negatif, atau netral. Konsumen dapat memperoleh gambaran objektif terhadap kualitas produk, sedangkan penjual dan penyedia platform dapat meningkatkan layanan mereka berdasarkan feedback pelanggan.

Naïve Bayes merupakan salah satu algoritma yang mampu melakukan pengklasifikasian dengan cepat. *Naïve Bayes* juga merupakan salah satu algoritma yang sangat efisien dan efektif bahkan saat digunakan untuk menganalisis data berskala besar yang dikenal mampu melakukan klasifikasi teks dengan baik, terutama dalam konteks analisis sentimen [3].

Penelitian ini bertujuan untuk menerapkan algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen ulasan pelanggan terhadap pembelian ponsel di Shopee, sehingga dapat membantu mengatasi tantangan dalam pengambilan keputusan konsumen serta meningkatkan pengalaman berbelanja secara daring.

II. METODE

Penelitian ini dilakukan menggunakan pendekatan metodologi Waterfall karena seluruh proses dikembangkan secara berurutan dan terstruktur, mulai dari analisis kebutuhan hingga implementasi sistem. Setiap tahapan diselesaikan terlebih dahulu sebelum melanjutkan ke tahap berikutnya. Langkah-langkah pengembangan sistem yang digambarkan sebagai berikut:



Gambar 1. Alur Penelitian

Berikut penjelasan dari gambar tersebut :

1. Pengambilan Data

Data yang digunakan adalah ulasan pelanggan yang melakukan pembelian ponsel di toko Goldencellcityshopee pada e-commerce . Proses pengumpulan data dilakukan melalui teknik *web scraping* menggunakan bahasa pemrograman *Python* dengan bantuan *library* Selenium.

2. Preprocessing Data

Proses *preprocessing* dilakukan agar data yang digunakan bersih dari *noise*, memiliki dimensi yang lebih kecil, serta lebih terstruktur, sehingga dapat diolah [4], ada beberapa tahapan yaitu:

a) Case Folding

Mengubah seluruh teks menjadi huruf kecil untuk konsistensi, bertujuan guna menyamakan seluruh huruf terhadap komentar sebagai huruf kecil [5]

b) Cleansing

Cleansing merupakan proses yang dilakukan untuk membuang karakter dan tanda baca yang tidak diperlukan seperti titik, koma, tanda seru, tanda tanya, termasuk menghilangkan URL dan HTML, menghapus emoji, mention atau hashtag, menghapus kata-kata slang, dan menghilangkan symbol atau huruf yang tidak relevan [6].

c) Tokenization

Tokenization adalah proses pemisahan atau pembagian paragraf menjadi kalimat dan kalimat menjadi kata-kata [7]

d) Stopword

Menghapus kata-kata umum yang tidak memiliki makna penting (seperti "dan", "yang", "di", dll). Proses *stopword removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses *sentiment* [8].

e) Stemming

Stemming yaitu suatu proses mengembalikan menjadi kata dasar dengan cara membuang awalan, akhiran atau sisipan, (misalnya : membeli menjadi beli) [9].

3. Pelabelan Data

Pelabelan dengan *Textblob*, *Textblob* adalah analisis tingkat kalimat, dibutuhkan dataset sebagai input kemudian membagi review menjadi kalimat. Cara umum untuk menentukan polaritas untuk seluruh kumpulan data adalah dengan menghitung jumlah kalimat/ulasan positif dan negatif dan memutuskan apakah tanggapannya positif dan negatif berdasarkan jumlah total ulasan positif dan negatif [10].

4. Pembagian Dataset

Pembagian dataset dilakukan dengan 80% data training dan 20% data testing.

5. Naïve Bayes

Naïve Bayes digunakan untuk mencari nilai probabilitas tertinggi dalam teknik klasifikasi data uji pada kategori yang paling tepat. Persamaan probabilitas yang mendasari adalah :

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Keterangan :

X = Data class yang belum diketahui

H = Hipotesis data class
 $P(H | X)$ = Probabilitas hipotesis H pada kondisi X
 $P(H)$ = Probabilitas hipotesis H
 $P(X | H)$ = Probabilitas X pada hipotesis H
 $P(X)$ = Probabilitas X

6. Confusion Matrix

Hasil prediksi dievaluasi menggunakan confusion matrix Accuracy, Precision, recall, dan F1-Score.

III. HASIL DAN PEMBAHASAN

Berdasarkan pengembangan sistem yang dilakukan adalah dari perancangan sistem yang telah dibuat, maka dapat disajikan pembahasan penelitian sebagai berikut:

a. Dataset Ulasan

Pengumpulan dataset menggunakan *Python* dengan *library* Selenium, pengumpulan data tersebut berjumlah 365 data ulasan pelanggan ponsel di toko Goldencellcityshopee kemudian di export kedalam bentuk csv atau xlsx. Berikut contoh data scrapping yang telah dilakukan pada Tabel 1.

Table 1 Contoh Data Scrapping Ulasan

No	username	rating	timestamp	Ulasan
1	b*****p	5	04/02/2023 8:20:00	Mantap jiwa Seller ramah poll Barang no minus
2	a*****a	5	2023-06-17 12:47	Barang dan pelayanan terbaik, Mulus sesuai
3	demapobae	5	2023-05-24 12:11	Like new, bener" seperti baru, Barang mulus

b. Preprocessing

Setelah dilakukan proses pengambilan dataset berupa teks ulasan produk dari platform daring langkah selanjutnya adalah *preprocessing* atau prapemrosesan data. Tahapan *preprocessing* ini bertujuan untuk membersihkan dan menyederhanakan teks agar siap digunakan dalam proses analisis lanjutan. Adapun tahapan *preprocessing* yang telah dilakukan pada Tabel 2.

Table 2 Preprocessing Data

Review	Case Folding	Cleansing	Tokenaisasi	Stopword	Stemming
Barang bagus, original, fullset, sesuai deskripsi.	barang bagus, original, fullset, sesuai deskripsi.	barang bagus original fullset sesuai deskripsi fitur	['barang', 'bagus', 'original', 'fullset', 'sesuai', 'deskripsi', 'fitur']	barang bagus original fullset sesuai deskripsi fitur	barang bagus original fullset sesuai deskripsi fitur
Fitur sangat baik, harga sangat sepadan.	fitur sangat baik, harga sangat sepadan.	sangat baik harga sangat padan	'deskripsi', 'fitur', 'sangat', 'baik', 'harga', 'sepadan']	sangat baik harga sangat sepadan	sangat baik harga sangat padan

Barang mulus, hanya lecet kecil. Pemakaian 10 hari lancar. Pixel 6A recommended	barang mulus lecet kecil pakai hari lancar pixel a recommended	barang mulus lecet kecil pakai hari lancar pixel a recommended	['barang', 'mulus', 'lecet', 'kecil', 'pakai', 'hari', 'lancar', 'pixel', 'a', 'recommended']	barang mulus lecet kecil pakai hari lancar pixel a recommended	barang mulus lecet kecil pakai hari lancar pixel a recommended
Jangan bertransaksi di toko ini, yang dikirim hp rusak tidak bertanggung jawab!	jangan bertransaksi di toko ini, yang dikirim hp rusak tidak bertanggung jawab!	jangan transaksi toko yang kirim hp rusak tanggung jawab	['jangan', 'bertransaksi', 'di', 'toko', 'ini', 'yang', 'dikirim', 'hp', 'rusak', 'tidak', 'bertanggung', 'jawab']	jangan bertransaksi toko yang dikirim hp rusak bertanggung jawab	jangan transaksi toko yang kirim hp rusak tanggung jawab

c. Pelabelan Data

Setelah data dibersihkan dan distandarisasi, tahap selanjutnya adalah pelabelan sentimen. Pada tahap ini, digunakan pustaka *TextBlob* yang berbasis Python untuk menentukan polaritas setiap ulasan. *TextBlob* melakukan analisis sentimen dengan mengukur nilai polaritas (positif-negatif) dan subjectivity (objektivitas-subjektivitas) dari sebuah teks. Berikut hasil pelabelan pada Tabel 3.

Table 3 Pelabelan Data

Index	Preprocessing	Score	Sentimen
0	barang bagus original fullset sesuai deskripsi fitur sangat baik harga sangat padan	1	Positif
1	barang mulus lecet kecil pakai hari lancar pixel a recommended	1	Positif
2	jangan transaksi toko yang kirim hp rusak tanggung jawab	-1	Negatif

d. Pembagian Data

Langkah selanjutnya adalah pembagian dataset (*data splitting*), yang bertujuan untuk memisahkan data menjadi dua bagian. Dalam penelitian ini, digunakan rasio 80% data untuk pelatihan dan 20% untuk pengujian. Dengan jumlah total data sebanyak 365 ulasan, maka pembagiannya 292 data (80%) digunakan sebagai training 73 data (20%) digunakan sebagai testing.

e. Klasifikasi *Naïve Bayes* dan Evaluasi Hasil

Setelah model dilatih menggunakan data latih, tahap selanjutnya adalah evaluasi hasil untuk mengetahui seberapa baik model dalam melakukan prediksi sentimen pada data uji. Berikut adalah hasil dari pegujian terdapat pada Gambar 3.1.



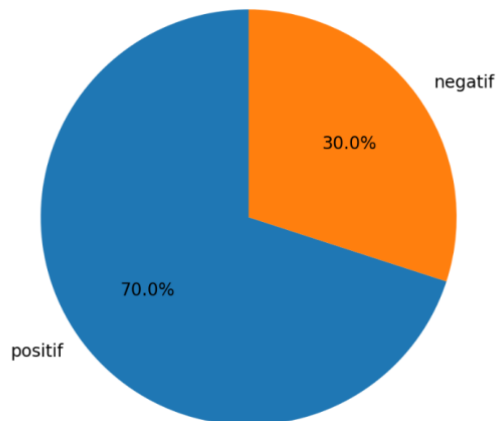
Classification Report:

	precision	recall	f1-score	support
negatif	1.00	0.18	0.30	17
positif	0.80	1.00	0.89	57
accuracy			0.81	74
macro avg	0.90	0.59	0.60	74
weighted avg	0.85	0.81	0.75	74

Gambar 2. Confusion Matrix

Berdasarkan hasil evaluasi model yang ditampilkan dalam classification report, diketahui bahwa model memiliki akurasi sebesar 81%. Hal ini menunjukkan bahwa secara keseluruhan, model mampu mengklasifikasikan data dengan cukup baik.

f. Visualisasi Hasil



Gambar 3. Visualisasi Hasil

Diagram lingkaran menunjukkan bahwa 70% ulasan pelanggan tergolong positif, sementara 30% bersentimen negatif. Hal ini menunjukkan mayoritas pelanggan merasa puas terhadap produk ponsel yang dibeli secara daring, meskipun masih terdapat sebagian yang mengalami ketidakpuasan. Temuan ini mengindikasikan bahwa layanan secara umum sudah baik, namun tetap perlu peningkatan di beberapa aspek.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Naïve Bayes* dapat secara efektif digunakan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap produk ponsel di platform e-commerce Shopee. Dengan proses *preprocessing* yang sistematis dan pelabelan menggunakan *TextBlob*, model yang dibangun mampu mencapai akurasi sebesar 81% dalam membedakan sentimen positif dan negatif. Hasil ini membuktikan bahwa pendekatan analisis sentimen berbasis *Naïve Bayes* tidak hanya mampu membantu konsumen dalam mengambil keputusan pembelian yang lebih baik, tetapi juga dapat memberikan masukan berharga bagi penjual dan platform e-commerce untuk meningkatkan kualitas layanan mereka.

DAFTAR PUSTAKA

- [1] P. Organisasi, “Pengembangan Organisasi di Perusahaan Digital shopee,” no. April, 2025.
- [2] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [3] A. Hanafiah, A. H. Nasution, Y. Arta, R. Wandri, H. O. Nasution, and J. Mardafora, “Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 1, pp. 230–236, 2023, doi: 10.31539/intecom.v6i1.5845.
- [4] F. S. Jumeilah, “Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.
- [5] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, “Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 207–215, 2023, doi: 10.37859/coscitech.v4i1.4790.
- [6] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, “Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification,” *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
- [7] A. F. Anees, A. Shaikh, A. Shaikh, and S. Shaikh, “Survey Paper on Sentiment Analysis : Techniques and Challenges,” *EasyChair*, pp. 2516–2314, 2020, [Online]. Available: https://login.easychair.org/publications/preprint_download/Sc2h
- [8] R. Azhar and M. F. Wijayanto, “Analisis Sentimen di Twitter: Mengungkap Persepsi dan Emosi Publik Seputar Konflik Palestina-Israel,” *Stain. (Seminar Nas. ...)*, vol. 3, pp. 118–124, 2024, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/4132>
- [9] M. S. Anwar, I. M. I. Subroto, and S. Mulyono, “Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal Dan Stemming Berbasis Android,” *Konf. Ilm. Mhs. Unissula 2*, pp. 58–70, 2019.
- [10] F. Fazrin, O. N. Pratiwi, and R. Andreswari, “Perbandingan Algoritma K-Nearest Neighbor dan Logistic Regression pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan Vader dan Textblob,” *J. e-Proceeding Eng.*, vol. 10, no. 2, pp. 1596–1604, 2023.