

# Analisis Sentimen Ulasan Produk Sneakers Lokal ada Tokopedia Menggunakan Algoritma Naïve Bayes dan Support Vector Machine

I Komang Arya Trisumeikra<sup>1</sup>, Herlawati Herlawati<sup>1,\*</sup>, Agus Hidayat<sup>1</sup>

\* Korespondensi: e-mail: [herlawati@ubharajaya.ac.id](mailto:herlawati@ubharajaya.ac.id)

<sup>1</sup> Informatika, Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya, Jl. Perjuangan No. 81, Marga Mulya, Bekasi Utara, Bekasi, Jawa Barat 17143, Telp/Fax: (021) 88955882; e-mail: [komangaryatri@gmail.com](mailto:komangaryatri@gmail.com), [herlawati@ubharajaya.ac.id](mailto:herlawati@ubharajaya.ac.id), [agus.hidayat@dsn.ubharajaya.ac.id](mailto:agus.hidayat@dsn.ubharajaya.ac.id).

Submitted : 14 September 2025  
Revised : 13 Oktober 2025  
Accepted : 14 November 2025  
Published : 30 November 2025

## Abstract

The development of digital technology is driving transformation in various sectors, including the fashion industry, especially sneakers. Sneakers are now a symbol of modern lifestyle and global trends, with brands such as Nike and Adidas dominating the market. However, high prices are an obstacle for many Indonesian consumers. This opens up opportunities for local brands to offer quality products at affordable prices through e-commerce such as Tokopedia, the second highest traffic platform in Indonesia. The research analyzed sentiment from 1,032 consumer reviews of local sneakers from five stores: NAH Project, Aerostreet, Geoff Max, Ventela, and Brodo. The analysis was conducted using Naïve Bayes and Support Vector Machine (SVM) algorithms. The SVM evaluation results produced the highest accuracy of 98%, compared to Naïve Bayes which reached 96%. This best model is implemented in a web-based application to analyze the sentiment of new reviews, to assess the perceived quality and consumer satisfaction of local sneakers products on Tokopedia.

**Keywords:** Local Sneakers, Naïve Bayes, Sentiment Analysis, Support Vector Machine, Tokopedia

## Abstrak

Perkembangan teknologi digital mendorong transformasi di berbagai sektor, termasuk industri fashion, khususnya produk sneakers. Sneakers kini menjadi simbol gaya hidup modern dan tren global, dengan brand seperti Nike dan Adidas mendominasi pasar. Namun, harga tinggi menjadi kendala bagi banyak konsumen Indonesia. Hal ini membuka peluang bagi brand lokal untuk menawarkan produk berkualitas dengan harga terjangkau melalui e-commerce seperti Tokopedia, platform dengan trafik tertinggi kedua di Indonesia. Penelitian ini menganalisis sentimen dari 1.032 ulasan konsumen terhadap produk sneakers lokal dari lima toko: NAH Project, Aerostreet, Geoff Max, Ventela, dan Brodo. Analisis dilakukan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Hasil evaluasi SVM menghasilkan accuracy tertinggi sebesar 98%, dibandingkan Naïve Bayes yang mencapai 96%. Model terbaik ini diimplementasikan dalam aplikasi berbasis web untuk menganalisis sentimen ulasan baru, guna menilai persepsi kualitas dan kepuasan konsumen terhadap produk sneakers lokal di Tokopedia.

**Kata kunci:** Analisis Sentimen, Naïve Bayes, Sneakers Lokal, Support Vector Machine, Tokopedia

## 1. Pendahuluan

**JSRCS status is accredited by the Directorate General of Research Strengthening and Development No. 225/E/KPT/2022 with Indonesian Scientific Index (SINTA) journal-level of S5, starting from Volume 1 (2) 2020 to Volume 6 (1) 2025**

Perkembangan teknologi digital yang sangat pesat telah mendorong transformasi signifikan di berbagai sektor bisnis, termasuk industri *fashion*. Perusahaan dituntut untuk mampu memanfaatkan teknologi dalam menjalankan strategi bersaing guna mencapai tujuan bisnis (Hajaroh et al., 2024). Dalam industri *fashion*, permintaan pasar terus meningkat seiring dengan tren global yang dinamis dan kebutuhan masyarakat akan penunjang penampilan, yang menjadikan *fashion* sebagai bagian penting dari gaya hidup modern (Yahya, 2023).

*Fashion* tidak hanya soal penampilan, tetapi juga identitas dan ekspresi diri, terutama melalui pemilihan *brand*. Industri sepatu, khususnya *sneakers*, mengalami pertumbuhan pesat karena dianggap *stylish* dan sesuai dengan selera generasi muda. *Brand* global seperti Nike, Adidas, dan Puma mendominasi pasar, menjadi pilihan utama konsumen karena desain dan kualitasnya. Popularitas mereka dipengaruhi oleh pengaruh selebriti dan status sosial yang melekat pada produk mereka (Dharmawan & Handoyo, 2017).

Berdasarkan data dari Statista tahun 2024, Nike memimpin pasar dengan pendapatan lebih dari 35,23 miliar dolar AS, diikuti oleh Adidas (14,56 miliar) dan Puma (4,93 miliar). Tingginya pendapatan *brand* asing tersebut menunjukkan daya tariknya yang kuat di pasar global, termasuk di Indonesia (Statista, 2025). Namun, kebijakan baru terkait ambang batas barang impor turut menyebabkan kenaikan harga sepatu *brand* asing, menjadikannya kurang terjangkau bagi sebagian Masyarakat.

Tingginya harga sepatu *brand* asing membuka peluang bagi pelaku usaha lokal untuk memproduksi *sneakers* dengan kualitas bersaing namun harga lebih terjangkau. Ini menjadi strategi penting untuk menarik perhatian pasar domestik dan mendorong minat masyarakat terhadap produk lokal. *Sneakers* lokal yang kompetitif dari segi harga dan kualitas berpotensi menjadi alternatif utama bagi konsumen Indonesia (Jonathan, 2023).

Perkembangan desain dan konsep *sneakers* setiap tahunnya telah meningkatkan permintaan secara global. Hal ini menciptakan peluang bagi pelaku usaha lokal untuk memasarkan produk melalui platform digital seperti Tokopedia (Safitri et al., 2024). Tokopedia sebagai salah satu *e-commerce* terbesar di Indonesia, berhasil mencatat kunjungan sebanyak 77,6 juta pada Juli 2024, menjadi yang kedua setelah Shopee. Tokopedia menjadi platform strategis dalam memasarkan produk *fashion*, termasuk *sneakers* lokal (Rizaty, 2024).

Tokopedia memfasilitasi kemudahan berbelanja online yang efisien dan aman, serta menyediakan berbagai pilihan produk (Maghfiroh & Iriani, 2021). Berdasarkan data internal, sekitar 80% konsumen digital mempertahankan atau meningkatkan penggunaan layanan digital. Produk *fashion*, khususnya *sneakers*, termasuk kategori yang paling dicari. Ini menunjukkan bahwa pasar lokal memiliki peluang besar untuk bersaing jika didukung oleh strategi pemasaran dan inovasi produk yang tepat (Tim PR Tokopedia, 2025).

Meski peluang besar, pelaku usaha lokal menghadapi tantangan dalam hal persepsi konsumen terhadap kualitas dan popularitas produk lokal dibanding *brand* asing. Persepsi ini dapat diidentifikasi melalui ulasan konsumen di Tokopedia (Dewi et al., 2022). Analisis terhadap

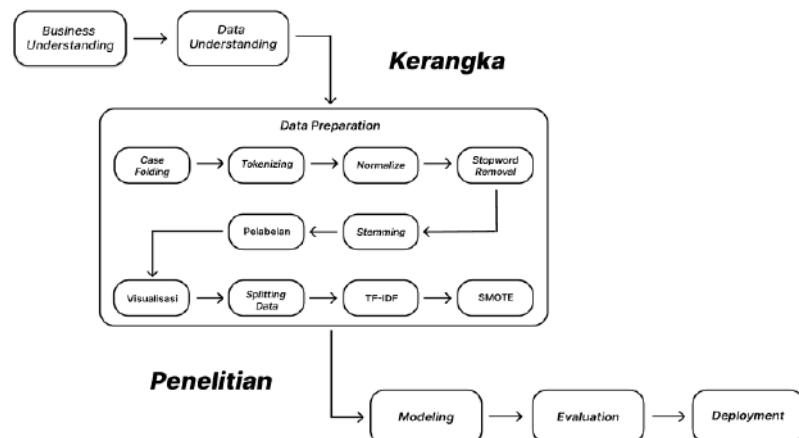
ulasan ini dapat menggambarkan pengalaman dan kepuasan konsumen terhadap produk *sneakers* lokal yang mereka beli dan gunakan.

Beberapa penelitian sebelumnya telah menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk menganalisis sentimen pada berbagai platform. Hasilnya menunjukkan bahwa kedua algoritma ini memiliki kelebihan masing-masing. Pada penelitian sentimen pengguna Lazada, SVM menunjukkan kinerja lebih baik dengan *accuracy* 75%. Namun, pada penelitian komentar Instagram edukatif, *Naïve Bayes* menunjukkan *accuracy* lebih tinggi, yaitu 84% (Setyabudi et al., 2025). Perbandingan ini menunjukkan bahwa efektivitas algoritma dapat bergantung pada karakteristik data yang digunakan.

Dengan begitu, penelitian ini dilakukan dengan menganalisis sentimen ulasan produk *sneakers* lokal di Tokopedia menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. Kedua algoritma dipilih berdasarkan efektivitas dan popularitasnya dalam klasifikasi data teks. *Naïve Bayes* unggul dalam kecepatan perhitungan dengan prinsip probabilistik, sedangkan SVM dikenal mampu menangani data non-linear dengan *accuracy* tinggi. Dengan menggunakan kedua algoritma ini, diharapkan dapat diketahui sentimen konsumen secara objektif terhadap produk *sneakers* lokal.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja sistematis dalam pelaksanaan proses penelitian. Proses diagram alur yang menggambarkan tahapan-tahapan dan diterapkan dalam penelitian ini secara terstruktur dapat dilihat pada Gambar 1.



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Penelitian

### 2.1. Machine Learning

*Machine learning* (pembelajaran mesin) merupakan salah satu metode dalam kecerdasan buatan (*Artificial Intelligence/AI*) yang dirancang untuk meniru kemampuan manusia dalam menyelesaikan berbagai permasalahan. Tujuan utamanya adalah melatih mesin agar mampu mengolah data secara efisien melalui identifikasi pola dalam dataset. Dengan

memanfaatkan algoritma untuk mempelajari hubungan kompleks dalam data, *machine learning* dapat membuat prediksi dan pengambilan keputusan dengan tingkat *accuracy* yang tinggi, tanpa bergantung pada pendekatan berbasis aturan (Wijoyo et al., 2024).

## 2.2. Analisis Sentimen

Analisis sentimen merupakan cabang dalam pengolahan bahasa alami, komputasi linguistik, dan penambangan teks yang berfokus pada pengidentifikasi dan interpretasi sentimen, opini, serta emosi seseorang terhadap suatu objek, seperti produk, layanan, topik, atau individu. Tujuan utamanya adalah mengklasifikasikan teks, baik dalam bentuk dokumen, kalimat, maupun frasa ke dalam kategori sentimen positif, negatif, atau netral. Selain itu, analisis ini juga dapat mengidentifikasi ulasan yang memiliki tingkat perhatian tinggi dari pengguna lain (Puad et al., 2023).

## 2.3. Web Scraping

*Web scraping* merupakan teknik pengambilan data secara otomatis dari situs web untuk disimpan dalam format terstruktur, seperti spreadsheet, basis data, atau file CSV. Teknik ini dapat dilakukan menggunakan berbagai bahasa pemrograman, salah satunya adalah Python, yang populer berkat ketersediaan pustaka-pustaka pendukung seperti *BeautifulSoup*, *Selenium*, dan *Pandas*. Tujuan utama *web scraping* adalah mengubah informasi tidak terstruktur dari situs web menjadi data yang lebih mudah dianalisis dan digunakan (Rizquina & Ratnasari, 2023).

## 2.4. Naïve Bayes

*Naïve Bayes* merupakan algoritma klasifikasi dalam *machine learning* yang didasarkan pada Teorema Bayes dan diperkenalkan oleh Thomas Bayes. Algoritma ini mengadopsi pendekatan probabilistik dengan asumsi bahwa setiap fitur bersifat independen dalam menentukan kelas suatu data, berdasarkan frekuensi kemunculannya dalam data pelatihan. Meskipun asumsi independensi tersebut jarang terpenuhi secara sempurna dalam praktik, *Naïve Bayes* tetap dikenal karena efisiensinya serta kemampuannya menghasilkan kinerja klasifikasi yang baik (Wibisono et al., 2020). Perhitungan dengan menggunakan Teorema Bayes dan distribusi multinomial dapat dijelaskan melalui rumus 1.

$$P(H|e) = \frac{P(e|H) \cdot P(H)}{P(e)} \quad (1)$$

Keterangan:

H : Data dengan kelas yang tidak diketahui (kelas target)

e : Kelas dari data (H) merupakan suatu fitur (kata) tertentu

$P(H|e)$  : Probabilitas kelas (H) diberikan fitur (e) yang diamati (Posterior)

$P(e|H)$  : Kelas yang dimaksud dari fitur (e) jika peluang benar (Likelihood)

$P(H)$  : Kelas tertentu dalam data (H) sebelum mengamati fitur (e) (Prior)

$P(e)$  : Probabilitas pada semua kemungkinan dari fitur (Marginal)

## 2.5. Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah algoritma klasifikasi *linear* dalam pembelajaran mesin yang dikenal karena keunggulan dan *accuracy* tinggi dalam berbagai aplikasi (Hovi et al.,

2022) Tujuan utama algoritma ini adalah menemukan *hyperplane* optimal yang memisahkan dua kelas data dengan margin maksimum. Dalam ruang dua dimensi, *hyperplane* berbentuk garis, sedangkan dalam dimensi yang lebih tinggi tetap disebut *hyperplane*. Pemisahan ini ditentukan oleh *support vector*, yaitu data yang berada paling dekat dengan batas pemisah dan berperan penting dalam menentukan posisi dan margin dari *hyperplane* tersebut (Idris et al., 2023). Pada *Support Vector Machine*, pemisah *hyperplane* dapat dirumuskan menggunakan persamaan berikut:

$$\mathbf{w}_1x_1 + \mathbf{w}_2x_2 + b = 0 \quad (1)$$

Dengan demikian, setiap titik yang berada di sisi atas dari *hyperplane* pemisah akan memenuhi syarat berikut:

$$\mathbf{w}_1x_1 + \mathbf{w}_2x_2 + b > 0 \quad (2)$$

Oleh karena itu, setiap titik yang berada di sisi bawah *hyperplane* pemisah akan memenuhi kondisi berikut:

$$\mathbf{w}_1x_1 + \mathbf{w}_2x_2 + b < 0 \quad (3)$$

### 3. Hasil dan Pembahasan

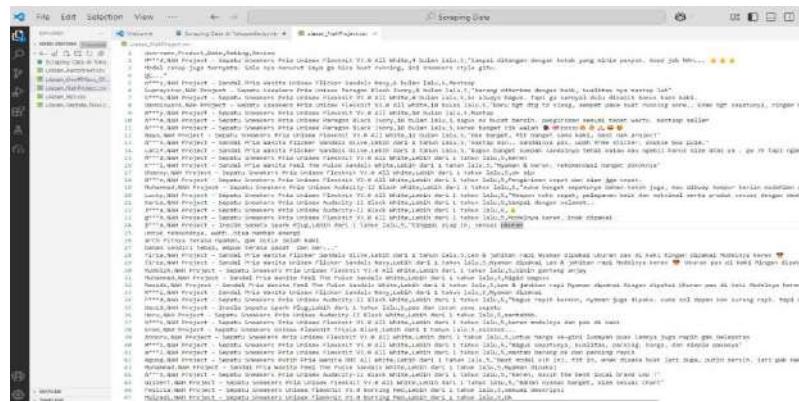
#### 3.1. Business Understanding

Tahap awal, berfokus untuk menganalisis sentimen terhadap ulasan produk *sneakers* lokal di Tokopedia. Tahap ini mencakup identifikasi permasalahan dan pengumpulan data ulasan dari berbagai toko. Ulasan konsumen dianalisis untuk mengklasifikasikan sentimen ke dalam dua kategori, yaitu positif dan negatif, dengan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM). Selain klasifikasi sentimen, penelitian ini juga bertujuan untuk mengevaluasi persepsi dan tingkat popularitas produk di kalangan konsumen. Hasil analisis diharapkan dapat memberikan informasi yang lebih transparan, membantu konsumen dalam memahami ulasan, serta mendukung pengambilan keputusan pembelian secara lebih percaya diri dan berkelanjutan.

#### 3.2. Data Understanding

Tahap kedua, pengumpulan data berupa ulasan konsumen terhadap produk *sneakers* lokal yang diperoleh dari Tokopedia dalam rentang waktu April 2023 hingga April 2025. Total data yang dianalisis berjumlah 2.397 ulasan, yang dikumpulkan dari lima toko daring yang dipilih secara purposif, yaitu NAH Project, Aerostreet, Geoff Max Official, Ventela\_NEW, dan Brodo Footwear\_NEW. Proses pengumpulan data dilakukan melalui teknik *web scraping* dengan menggunakan aplikasi *Visual Studio Code* serta didukung oleh pustaka (*library*) pemrograman Python, yakni *Selenium*, *BeautifulSoup*, dan *Pandas*. Data diperoleh langsung dari halaman web Tokopedia, kemudian disimpan dalam format CSV untuk keperluan analisis lebih lanjut. Setiap entri data terdiri atas lima atribut utama, yaitu *Username*, *Product*, *Date*, *Rating*, dan *Review*, yang mencerminkan identitas pengguna, nama produk, tanggal ulasan, penilaian berupa skor,

serta isi ulasan yang diberikan oleh konsumen. Hasil dari teknik *web scraping* untuk menampilkan dataset dari salah satu toko, yaitu Toko NAH Project dapat dilihat pada Gambar 2.



Sumber: Hasil Penelitian (2025)

Gambar 2. Hasil *Scraping* Ulasan pada Toko NAH Project

Data ulasan *sneakers* lokal dari kelima toko yang telah dikumpulkan melalui tahap *scraping* akan digabungkan menjadi satu dataset. Data hasil penggabungan dataset tersebut disimpan dalam format CSV. Hasil penggabungan dataset dapat dilihat pada Gambar 3.

Dataset berhasil digabungkan dan disimpan sebagai 'Data_Gabungan.csv'					
	Username	Product	Date	Rating	Review
0	Ilham	Geoff Max Official - Dublin Black Gum   Sepatu...	Hari ini	5	ok banget terimakasih semoga awet hadiah untuk...
1	A***d	Athletica - MX Faster Black White   Trail Runn...	Hari ini	3	Overall pas ukuran 42 cuman agak terlalu semp...
2	Ridwan	Geoff Max Official - East Black Gum   Sepatu P...	1 hari lalu	5	Bagus sepatu geoff maxx bukan cuman tipe athle...
3	A***i	Athletica Official Shop - ATH 4 Blitz Black   ...	2 hari lalu	5	pengiriman sangat cepat, untuk yang biasanya s...
4	R***n	Geoff Max - Spirex Black White   Sepatu Pria W...	3 hari lalu	5	Mantap, barangnya sesuai deskripsi dan semoga ...
...	...	...	...	...	...
2392	A***i	BRODO - Sandal Alba Full Black	2 minggu lalu	5	Bagus, cuma kekecilan
2393	i***i	BRODO - Sneakers Inizio Junior Black-White WS	2 minggu lalu	5	Bentuk dan ukuran: Sesuai\nWarna: Sesuai\nCepat
2394	agus	BRODO - Sepatu Signore Low Dark Choco BS	2 minggu lalu	5	barang sesuai size sesuai, mantap
2395	i***f	BRODO - Sneakers Casual Calcio All White	2 minggu lalu	5	Bagus banget emng Brodo salah satu brand lokal...
2396	A***i	BRODO - Sandal Alba Full Black	2 minggu lalu	5	Bagus, cuma kekecilan

Sumber: Hasil Penelitian (2025)

Gambar 3. Hasil Penggabungan Dataset

Selanjutnya adalah setiap toko akan di filter untuk memastikan bahwa produk yang dijual merupakan *sneakers* untuk memastikan sesuai dengan objek yang akan dianalisis. Hasil kode filter dataset dapat dilihat pada Gambar 4.

Produk selain 'sneakers' sudah dihapus. Jumlah data setelah filter: 1229 baris.					
	Username	Product	Date	Rating	Review
0	Ilham	Geoff Max Official - Dublin Black Gum   Sepatu...	Hari ini	5	ok banget terimakasih semoga awet hadiah untuk...
1	Ridwan	Geoff Max Official - East Black Gum   Sepatu P...	1 hari lalu	5	Bagus sepatu geoff maxx bukan cuman tipe athle...
2	R***n	Geoff Max - Spirex Black White   Sepatu Pria W...	3 hari lalu	5	Mantap, barangnya sesuai deskripsi dan semoga ...
3	Oki	Geoff Max Official - Dublin Black Gum   Sepatu...	3 hari lalu	5	modelnya pas, jahitan dan finishingnya rapi. p...
4	Muhammad	Geoff Max Official - Dublin Black White   Sepatu...	3 hari lalu	5	keren kerenn kerenn kerenn kerenn kerenn kerenn
...	...	...	...	...	...
1224	i***f	BRODO - Sneakers Casual Calcio All White	2 minggu lalu	5	Bagus banget emng Brodo salah satu brand lokal...
1225	i***i	BRODO - Sneakers Inizio Junior Black-White WS	2 minggu lalu	5	Bentuk dan ukuran: Sesuai\nWarna: Sesuai\nCepat
1226	i***f	BRODO - Sneakers Casual Calcio All White	2 minggu lalu	5	Bagus banget emng Brodo salah satu brand lokal...
1227	i***i	BRODO - Sneakers Inizio Junior Black-White WS	2 minggu lalu	5	Bentuk dan ukuran: Sesuai\nWarna: Sesuai\nCepat
1228	i***f	BRODO - Sneakers Casual Calcio All White	2 minggu lalu	5	Bagus banget emng Brodo salah satu brand lokal...

Sumber: Hasil Penelitian (2025)

Gambar 4. Hasil Filter Dataset

### **3.3. Data Preparation**

Tahap ketiga, data ulasan yang telah dikumpulkan diolah lebih lanjut untuk keperluan analisis sentimen. Mengingat data mentah masih mengandung kekurangan seperti duplikasi dan nilai kosong, dilakukan proses pembersihan data (*data cleaning*). Dari hasil identifikasi, ditemukan sebanyak 197 entri ulasan yang merupakan duplikasi. Selanjutnya, data duplikat tersebut dihapus untuk memastikan kualitas dan *accuracy* analisis. Hasil pembersihan data dapat dilihat pada Gambar 5.

	Username	Product	Date	Rating	Review
0	Ilham	Geoff Max Official - Dublin Black Gum   Sepatu...	Hari ini	5	ok banget terimakasih semoga awet hadiah untuk...
1	Ridwan	Geoff Max Official - East Black Gum   Sepatu P...	1 hari lalu	5	Bagus sepatu geoff maxx bukan cuman tipe athle...
2	R***n	Geoff Max - Spirex Black White   Sepatu Pria W...	3 hari lalu	5	Mantap, sepatu nya sesuai deskripsi dan semoga...
3	Oki	Geoff Max Official - Dublin Black Gum   Sepatu...	3 hari lalu	5	model nya pas, jahitan nya rapi. pengiriman jg...
4	Muhamad	Geoff Max Official - Dublin Black White   Sepa...	3 hari lalu	1	Sepatu pd bahan nya kualitas nya buruk, gampan...
...	...	...	...	...	...
1027	S***s	BRODO - Sneakers Ace Nova Grey OWS	2 minggu lalu	1	Sepatu yang saya terima sangat mengecewakan. D...
1028	S***o	BRODO - Sneakers Ace Nova Grey OWS	2 minggu lalu	5	Keren min sepatu nya
1029	A***f	BRODO - Sneakers Calcio Full Black	2 minggu lalu	5	respon cepat, barang bagus
1030	i***i	BRODO - Sneakers Inizio Junior Black-White WS	2 minggu lalu	2	Jelek ah bentuk dan ukuran tdk sesuai, kecewa ...
1031	j***f	BRODO - Sneakers Casual Calcio All White	2 minggu lalu	5	Bagus banget emng Brodo salah satu brand lokal...

1032 rows x 5 columns

Sumber: Hasil Penelitian (2025)

Gambar 5. Hasil Penghapusan Duplikasi Dataset

#### a. Case Folding

Tahap ini mencakup proses penyeragaman huruf guna menyeragamkan format teks dalam dataset. Proses ini meliputi konversi seluruh huruf menjadi huruf kecil, serta penghapusan mention, tanda baca, angka, simbol, URL, emoji, dan karakter non-alfanumerik. Hasil penerapan tahap *case folding* ditampilkan pada Tabel 1.

Tabel 1. Hasil Penerapan Tahapan Case Folding

Sebelum	Sesudah
Jujur desain nya biasa aja, kombinasi warna nya pas. <b>Sol</b> tebal, namun sedikit kecewa produk yg saya terima ada bagian yg bocel atau ngelupas. Apalagi dengan model yang buruk dgn bahan seadanya	jujur desain nya biasa aja kombinasi warna nya pas sol tebal namun sedikit kecewa produk yg saya terima ada bagian yg bocel atau ngelupas apalagi dengan model yang buruk dgn bahan seadanya
Terima kasih, pesanan dtng dgn cepat dan barang terbungkus rapi. sepatu nya <b>ORI</b> bagus sekali	terima kasih pesanan dtng dgn cepat dan barang terbungkus rapi sepatu nya ori bagus sekali
model nya bagus sesuai dengan gambar dan ukuran sesuai! 🙌	model nya bagus sesuai dengan gambar dan ukuran sesuai

Sumber: Hasil Penelitian (2025)

#### b. Tokenized

Tahap ini dilakukan dengan memecah teks ulasan menjadi unit terkecil, seperti kata individu, guna mempermudah analisis kata secara terpisah. Sebagai contoh, kalimat dalam ulasan

akan diuraikan menjadi token-token seperti “jujur”, “desain”, “nya”, biasa” dan seterusnya. Proses ini untuk memudahkan dalam mengakses serta menganalisis setiap kata secara terpisah. Hasil penerapan tahap *tokenized* ditampilkan pada Tabel 2.

Tabel 2. Hasil Penerapan Tahap *Tokenizing*

Sebelum	Sesudah
jujur desain nya biasa aja kombinasi warna nya pas sol tebal namun sedikit kecewa produk yg saya terima ada bagian yg bocel atau ngelupas apalagi dengan model yang buruk dgn bahan seadanya	[jujur, desain, nya, biasa, aja, kombinasi, warna, nya, pas, sol, tebal, namun, sedikit, kecewa, produk, yg, saya, terima, ada, bagian, yg, bocel, atau, ngelupas, apalagi, dengan, model, yang, buruk, dgn, bahan, seadanya]
terima kasih pesanan dtng dgn cepat dan barang terbungkus rapi sepatu nya ori bagus sekali	[terima, kasih, pesanan, dtng, dgn, cepat, dan, barang, terbungkus, rapi, sepatu, nya, ori, bagus, sekali]
model nya bagus sesuai dengan gambar dan ukuran sesuai	[model, nya, bagus, sesuai, dengan, gambar, dan, ukuran, sesuai]

Sumber: Hasil Penelitian (2025)

c. *Normalize*

Tahap ini dilakukan dengan mengonversi kata tidak baku, seperti slang dan singkatan, ke dalam bentuk baku sesuai kaidah Kamus Besar Bahasa Indonesia (KBBI). Dalam penelitian ini, kamus kata slang disusun secara manual berdasarkan observasi terhadap data ulasan yang dianalisis. Hasil penerapan proses *normalize* ditampilkan pada Tabel 3.

Tabel 3. Hasil Penerapan Tahap *Normalize*

Sebelum	Sesudah
[jujur, desain, nya, biasa, aja, kombinasi, warna, nya, <b>pas</b> , sol, tebal, namun, sedikit, kecewa, produk, <b>yg</b> , saya, terima, ada, bagian, <b>yg</b> , bocel, atau, ngelupas, apalagi, dengan, model, yang, buruk, <b>dgn</b> , bahan, seadanya]	[jujur, desain, nya, biasa, aja, kombinasi, warna, nya, sesuai, sol, tebal, namun, sedikit, kecewa, produk, yang, saya, terima, ada, bagian, yang, bocel, atau, ngelupas, apalagi, dengan, model, yang, buruk, dengan, bahan, seadanya]
[terima, kasih, pesanan, <b>dtng</b> , <b>dgn</b> , cepat, dan, barang, terbungkus, <b>rapi</b> , sepatu, nya, <b>ori</b> , bagus, sekali]	[terima, kasih, pesanan, datang, dengan, cepat, dan, barang, terbungkus, rapih, sepatu, nya, asli, bagus, sekali]
[model, nya, bagus, sesuai, dengan, gambar, dan, ukuran, sesuai]	[model, nya, bagus, sesuai, dengan, gambar, dan, ukuran, sesuai]

Sumber: Hasil Penelitian (2025)

d. *Stopword Removal*

Tahap ini dilakukan untuk menghilangkan kata-kata yang dianggap kurang bermakna dalam analisis, seperti “yang”, “atau”, “dan”, “dengan”, dan sebagainya. Proses ini dilakukan untuk mengurangi dimensi data dan memfokuskan analisis pada kata-kata yang lebih informatif dalam penentuan sentimen. Hasil dari tahap *stopword removal* ditampilkan pada Tabel 4.

Tabel 4. Hasil Penerapan Tahap Stopword Removal

Sebelum	Sesudah
[jujur, desain, <b>nya</b> , <b>biasa</b> , aja, kombinasi, warna, sesuai, sol, tebal, <b>namun</b> , <b>sedikit</b> , kecewa, produk, <b>yang</b> , <b>saya</b> , terima, <b>ada</b> , <b>bagian</b> , <b>yang</b> , bocel, atau, ngelupas, <b>apalagi</b> , <b>dengan</b> , model, <b>yang</b> , buruk, <b>dengan</b> , bahan, seadanya]	[jujur, desain, aja, kombinasi, warna, sesuai, <b>nya</b> , sesuai, sol, tebal, kecewa, produk, terima, bocel, ngelupas, model, <b>yang</b> , <b>saya</b> , terima, <b>ada</b> , <b>bagian</b> , <b>yang</b> , <b>nya</b> , <b>biasa</b> , aja, kombinasi, warna, sesuai, sol, tebal, kecewa, produk, terima, bocel, buruk, bahan, seadanya]
[terima, kasih, pesanan, <b>datang</b> , <b>dengan</b> , cepat, barang, terbungkus, rapih, sepatu, <b>nya</b> , asli, bagus, <b>sekali</b> ]	[terima, kasih, pesanan, cepat, barang, terbungkus, rapih, sepatu, asli, bagus]
[model, <b>nya</b> , bagus, sesuai, <b>dengan</b> , gambar, ukuran, sesuai]	[model, bagus, sesuai, gambar, ukuran, sesuai]

Sumber: Hasil Penelitian (2025)

#### e. Stemming

Tahap ini dilakukan dengan mengubah kata berimbuhan ke dalam bentuk dasarnya. Proses ini bertujuan mengurangi variasi kata dalam teks agar analisis dapat dilakukan secara lebih konsisten dan efisien. Hasil penerapan tahap stemming ditampilkan pada Tabel 5.

Tabel 5. Hasil Penerapan Tahap Stemming

Sebelum	Sesudah
[jujur, desain, aja, kombinasi, warna, sesuai, sol, tebal, kecewa, produk, terima, bocel, ngelupas, model, buruk, bahan, <b>seadanya</b> ]	[jujur desain aja kombinasi warna sesuai sol tebal kecewa produk terima bocel ngelupas model buruk bahan ada]
[terima, kasih, pesanan, cepat, barang, terbungkus, rapih, sepatu, asli, bagus]	[terima kasih pesan cepat barang bungkus rapih sepatu asli bagus]
[model, bagus, sesuai, gambar, ukuran, sesuai]	[model bagus sesuai gambar ukur sesuai]

Sumber: Hasil Penelitian (2025)

#### f. Pelabelan

Pada tahap ini, dilakukan pelabelan sentimen secara otomatis dengan menggabungkan pendekatan berbasis kamus *lexicon-based* berbahasa Indonesia dan informasi dari *rating* pengguna

Text_Stemming	Rating	Sentiment_Label
bagus banget terima kasih moga awet hadiah orang tua gaul dikit	5	positif
begus sepatu geoff maxx tidak type athletica ringan empuk model geoff maxx sesuai bagus ngeputu n...	5	positif
mantap sepatu sesuai deskripsi imoga awet	5	positif
model sesuai janji rrim capat kerenn banget	5	positif
sepatu bahan kualitas buruk gampang rusak tidak deli boli	1	negatif
keren ukur sesuai urus imoga awet	5	positif
keren mantap sepatu keren cepat	5	positif
bagus sole suka	5	positif
ukur sesuai kualitas bagus harga ramah telima kasih geoffmaxx	5	positif
model keran sesuai gambar proses cepat	5	positif
barang stok sol sepatu luntur upper kecewa tidak rekomendasii chat admin informasi konfirmasi tim..	1	positif
alhamdulillah cepat respon cepat produk begus bangel	5	positif
jujur desain aja kombinasi warna sesuai isi tidak kecewa produk telima bocel ngelupas modal buru..	2	negatif
lumayan barang kecewa bahan kelupa buka	2	negatif
sepatu begus keran begus sesuai banget krim cepat kurir ramah suka geoff max	5	positif
beli sepatu moga awet ban	5	positif
begus sepatu suka	5	positif
kemas dua kali kardus produk sesuai ukur kali profil lebar aran pakai ukur tidak sempit krim ce...	5	positif
kerim cepat respon admin cepat kemas rapih barang begus tidak cacat	5	positif
ya lumayan kecewa sepatu model jalak krim telat hasil kurir tidak ramah	3	negatif

Sumber: Hasil Penelitian (2025)

Gambar 6. Hasil Penerapan Tahap Pelabelan

. Proses Gambar 6 menetapkan label positif atau negatif terhadap teks ulasan berdasarkan hasil analisis serta jumlah bintang yang diberikan. Secara khusus, ulasan dengan *rating* 1 hingga 3 dikategorikan sebagai sentimen negatif, sedangkan *rating* 4 dan 5 dianggap sebagai sentimen positif. Pelabelan dilakukan setelah proses *stemming* untuk memastikan keseragaman dan kebersihan data. Tahap ini penting dalam menyiapkan data yang valid dan representatif sebelum masuk ke proses klasifikasi sentimen. Hasil pelabelan dapat dilihat pada Gambar 6.

Hasil pelabelan sentimen menggunakan pendekatan *lexicon-based* yang disesuaikan dengan rating menunjukkan bahwa sebanyak 872 ulasan termasuk dalam kategori sentimen positif, sedangkan 160 ulasan tergolong sentimen negatif. Distribusi jumlah pada masing-masing kelas sentimen ditampilkan dalam Tabel 6.

Tabel 6. Hasil Jumlah Label Sentimen

Label	Sentimen
Positif	872
Negatif	160

Sumber: Hasil Penelitian (2025)

#### **g. Visualisasi**

Pada tahap ini, dilakukan visualisasi data teks menggunakan teknik *word cloud* untuk menggambarkan frekuensi kemunculan kata dalam ulasan. Metode ini berfungsi sebagai alat eksploratif untuk mengenali kata-kata yang berasosiasi dengan sentimen positif maupun negatif. Hasil perbandingan *word cloud* pada sentimen positif dan negatif dapat dilihat pada Gambar 7.



Sumber: Hasil Penelitian (2025)

Gambar 7. Word Cloud Sentimen Positif (a) dan Word Cloud Sentimen Negatif (b)

#### *h. Splitting Data*

Pada tahap ini, dilakukan pembagian dataset menjadi dua subset utama, yaitu data latih sebesar 80% (825 data) dan data uji sebesar 20% (207 data). Pembagian ini bertujuan untuk melatih model pada data latih dan mengevaluasi kinerjanya secara objektif menggunakan data uji yang belum dikenali sebelumnya. Proses ini dilakukan secara proporsional guna menjaga representasi data, sehingga evaluasi kinerja model dapat mencerminkan kemampuan generalisasi secara lebih akurat. Hasil mengenai distribusi jumlah data pada masing-masing subset dapat dilihat pada Tabel 7.

Tabel 7. Distribusi Jumlah Data Latih dan Data Uji

Distribusi Label	Data Latih	Data Uji
Positif	697	175
Negatif	128	32
Jumlah	825	207

Sumber: Hasil Penelitian (2025)

i. **TF-IDF**

Tahap ini menerapkan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk memberikan bobot numerik pada setiap kata hasil ekstraksi dari data teks. TF mencerminkan frekuensi kemunculan kata dalam suatu dokumen, sementara IDF menunjukkan tingkat keunikan kata di seluruh dokumen. Representasi ini digunakan sebagai input dalam analisis lanjutan. Setelah transformasi, jumlah dokumen pada data latih dan data uji masing-masing adalah 825 dan 207, dengan total fitur yang dihasilkan sebanyak 4.061 pada kedua subset. Rincian hasil transformasi ditampilkan pada Tabel 8.

Tabel 8. Jumlah Data Transformasi Data Latih dan Data Uji

Data Transformasi	Data Latih	Data Uji
Dokumen	825	207
Fitur	4072	4072

Sumber: Hasil Penelitian (2025)

j. **SMOTE**

Pada tahap ini, untuk mengatasi ketidakseimbangan kelas, dilakukan *oversampling* dengan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Teknik ini menghasilkan data sintetis pada kelas minoritas guna menyeimbangkan distribusi antar kelas. *Oversampling* diterapkan setelah transformasi TF-IDF, mengingat SMOTE memerlukan input dalam bentuk numerik. Perbandingan distribusi label sebelum dan sesudah SMOTE ditampilkan pada Tabel 9.

Tabel 9. Distribusi Label Sebelum dan Sesudah SMOTE

Label	Sebelum SMOTE	Sesudah SMOTE
Positif	697	697
Negatif	128	697

Sumber: Hasil Penelitian (2025)

**3.4. Modeling**

Pada tahap ini, dilakukan proses klasifikasi menggunakan dua algoritma pembelajaran mesin, *Naïve Bayes* dan *Support Vector Machine* (SVM) dengan *kernel linear* melalui *library SVC*. *Naïve Bayes* dipilih karena sesuai untuk data teks yang telah direpresentasikan secara numerik melalui TF-IDF, sedangkan SVM digunakan untuk menangani klasifikasi *linear* secara optimal. Kedua model dievaluasi untuk menilai kinerjanya dalam mengklasifikasikan sentimen serta dibandingkan guna menentukan algoritma dengan hasil terbaik pada penelitian ini.

a. *Naïve Bayes*

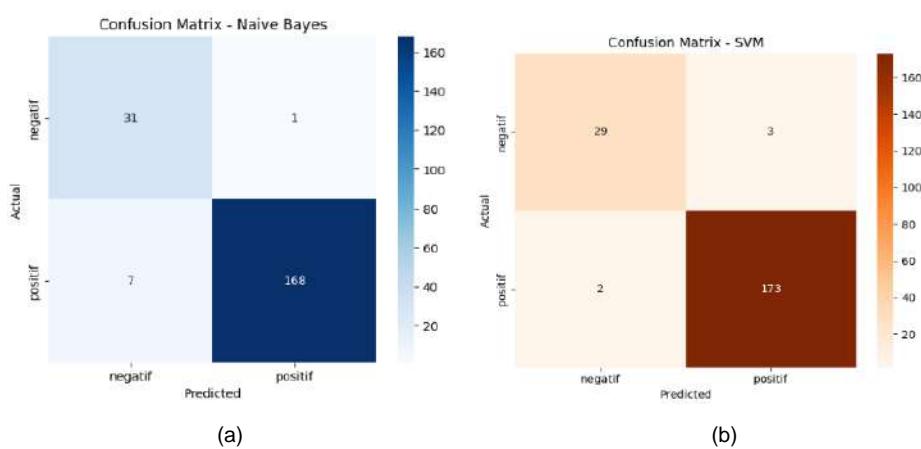
Setelah data diseimbangkan menggunakan metode SMOTE, dilakukan klasifikasi teks menggunakan algoritma *Multinomial Naïve Bayes*. Algoritma ini cocok untuk data dengan representasi numerik seperti TF-IDF, karena mengasumsikan independensi antar fitur terhadap kelas. Hasil pemodelan menunjukkan bahwa model mencapai *accuracy* sebesar 96%, yang mengindikasikan bahwa model memiliki kinerja yang sangat baik dalam mengklasifikasikan ulasan produk *sneakers* lokal di Tokopedia.

b. *Support Vector Machine (SVM)*

Pemodelan dengan algoritma *Support Vector Machine (SVM)* dilakukan menggunakan pendekatan yang sama seperti pada *Naïve Bayes*, yakni setelah proses *oversampling* dengan SMOTE. Implementasi dilakukan melalui library SVC dengan *kernel linear*, yang membentuk *hyperplane linear* untuk memisahkan kelas secara optimal. Pendekatan ini dinilai efektif dalam menangani data berdimensi tinggi, seperti teks yang telah direpresentasikan menggunakan TF-IDF. Hasil pemodelan berhasil memperoleh mencapai *accuracy* sebesar 98%, yang menandakan kinerja yang sangat baik dalam mengklasifikasikan ulasan produk *sneakers* lokal di Tokopedia.

### 3.5. Evaluation

Berdasarkan hasil pemodelan, algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* sama-sama menunjukkan kinerja yang baik. Namun, masing-masing algoritma memiliki keunggulan dan keterbatasan tersendiri yang perlu dipertimbangkan dalam menentukan model yang paling sesuai dengan menggunakan *confusion matrix* dengan skenario data 80:20



Sumber: Hasil Penelitian (2025)

Gambar 8. *Confusion Matrix* Model *Naïve Bayes* (a) dan *Confusion Matrix* Model *Support Vector Machine (SVM)* (b)

Evaluasi kinerja dengan menggunakan *confusion matrix* untuk algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan ulasan produk *sneakers* lokal dari lima toko di Tokopedia ke dalam kelas sentimen positif dan negatif, dapat dilihat pada Gambar 8. Berdasarkan hasil evaluasi kinerja kedua model klasifikasi, baik *Naïve Bayes* maupun *SVM* menunjukkan *Support Vector Machine* kinerja yang cukup baik dalam mengklasifikasikan

sentimen ulasan. Namun, terdapat beberapa perbedaan signifikan yang dapat dijadikan dasar analisis perbandingan. Hasil perbandingan evaluasi berdasarkan *macro average* kedua algoritma dapat dilihat pada Tabel 10.

Tabel 10. Hasil Perbandingan Evaluasi Pada Naïve Bayes dan SVM

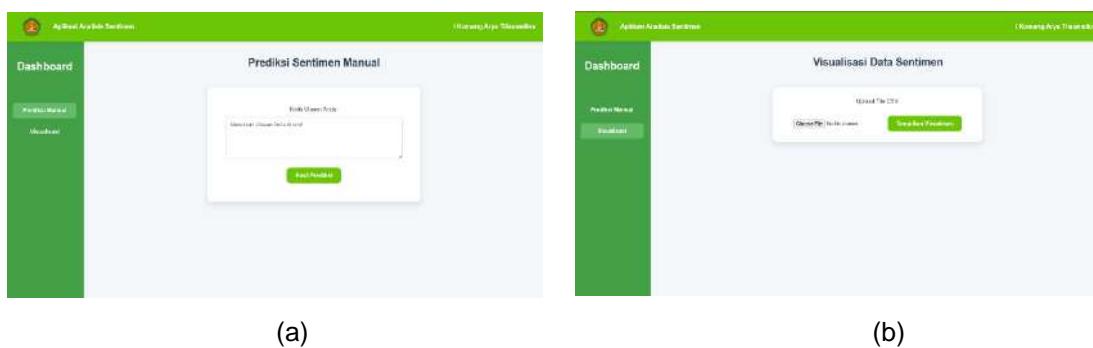
Algoritma	Accuracy	Precision	Recall	F-1 Score
Naïve Bayes	96%	90%	96%	93%
Support Vector Machine	98%	96%	95%	95%

Sumber: Hasil Penelitian (2025)

Berdasarkan Tabel 10, model *Naïve Bayes* mencatat *accuracy* sebesar 96%, dengan nilai rata-rata *macro average precision* sebesar 90%, *recall* sebesar 96%, dan *f1-score* sebesar 93%. Meskipun menunjukkan kinerja yang baik dalam mengidentifikasi ulasan positif, efektivitasnya untuk ulasan negatif masih lebih rendah, sehingga terdapat ketidakseimbangan kinerja antar kelas. Sebaliknya, model *Support Vector Machine* (SVM) memperlihatkan hasil yang lebih stabil dengan *accuracy* mencapai 98%, *precision* sebesar 96%, *recall* sebesar 95%, dan *f1-score* sebesar 95%. Ini menunjukkan bahwa SVM mampu memberikan kinerja yang lebih seragam dan andal pada kedua kelas sentimen. Dengan demikian, berdasarkan nilai rata-rata makro, SVM terbukti unggul secara keseluruhan karena mampu menghasilkan klasifikasi yang lebih seimbang dan akurat antara kelas positif dan negatif.

### 3.6. Deployment

Pada tahap akhir, hasil analisis sentimen tentang ulasan produk *sneakers* lokal di Tokopedia diintegrasikan ke dalam sebuah aplikasi berbasis web. Sistem ini dirancang untuk memberikan akses mudah kepada pengguna dalam melihat dan menganalisis sentimen ulasan secara otomatis, baik untuk eksplorasi data maupun pengambilan keputusan.



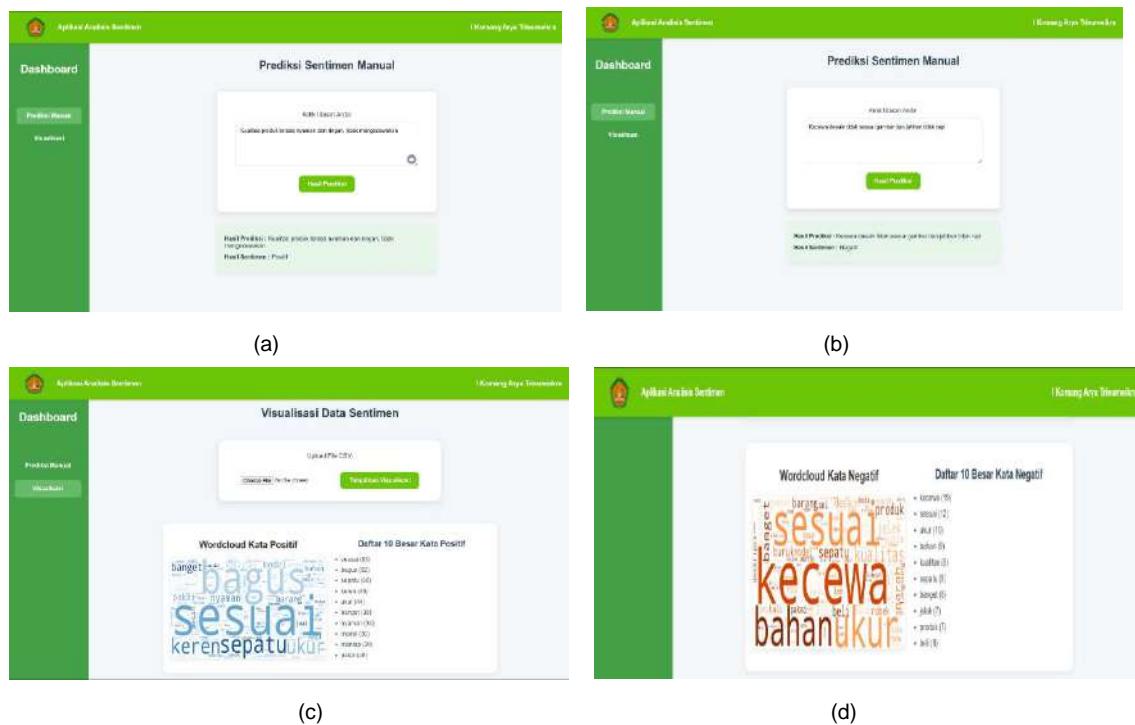
Sumber: Hasil Penelitian (2025)

Gambar 9. Tampilan Antarmuka Pada Aplikasi Analisis Sentimen Berbasis Web Pada Fitur  
Prediksi Manual (A), Dan Fitur Visualisasi (B)

Representasi fitur setiap ulasan menggunakan metode TF-IDF, sehingga representasi ini digunakan sebagai dasar dalam proses prediksi dan pengujian model di sistem. Tampilan antarmuka pada aplikasi analisis sentimen berbasis web dapat dilihat pada Gambar 9.

Pada fitur prediksi manual, pengguna dapat memasukkan kalimat ulasan baru secara langsung terkait produk *sneakers* lokal di Tokopedia. Misalkan ulasan "Kualitas produk terasa nyaman dan ringan, tidak mengecewakan", sistem kemudian menetapkan sentimen sebagai

positif, sebagaimana ditampilkan pada Gambar 10 (a), dan ulasan “Kecewa desain tidak sesuai gambar dan jahitan tidak rapi”, sistem kemudian menetapkan sentimen sebagai negatif, sebagaimana ditampilkan pada Gambar 10 (b). Selain itu, pengguna juga dapat mengunggah kumpulan data uji ke fitur visualisasi, seperti ditampilkan pada Gambar 10 (c) dan (d), yang menghasilkan tampilan *word cloud* serta daftar sepuluh kata besar untuk setiap kelas sentimen.



Sumber: Hasil Penelitian (2025)

Gambar 10. Hasil Fitur Prediksi Manual Sentimen Positif (a), Sentimen Negatif (b), dan Fitur Visualisasi Sentimen Positif (c), dan Sentimen Negatif (d)

#### 4. Kesimpulan

Penelitian ini menganalisis 1.032 ulasan produk *sneakers* lokal dari lima toko Tokopedia (NAH Project, Aerostreet, Geoff Max Official, Ventela\_NEW, Brodo Footwear\_NEW) yang dikumpulkan melalui *web scraping* menggunakan *Selenium*, *BeautifulSoup*, dan *Pandas* dengan kelima atribut utama (nama pengguna, nama produk, tanggal, rating, dan isi ulasan). Setelah *preprocessing* (TF-IDF dan penyeimbangan data menggunakan SMOTE), dua algoritma klasifikasi *Naïve Bayes* dan *SVM* diimplementasikan. Evaluasi menunjukkan bahwa *SVM* unggul dengan *accuracy* 98%, *Support Vector Machine precision* 96%, *recall* 95%, dan *f1-score* 95% (*macro average*), sementara *Naïve Bayes* mencatat *accuracy* 96%, *precision* 90%, *recall* 96%, dan *f1-score* 93%. Hasil ini menegaskan bahwa *SVM* lebih akurat dan seimbang dalam klasifikasi sentimen. Aplikasi berbasis web yang dihasilkan memfasilitasi analisis otomatis terhadap ulasan baru melalui prediksi manual dan visualisasi, sehingga bermanfaat bagi konsumen dalam pengambilan keputusan dan berkontribusi pada pengembangan metode analisis sentimen di ranah *e-commerce*.

## **Daftar Pustaka**

- Dewi, E., Mulyani, S., Hidayatuloh, A. T., & Agustiawan, T. (2022). Analisis Sentimen Ulasan Produk Pada Top Brand Produk Masker Di Tokopedia Menggunakan Naive Bayes. *Seminar Nasional Penelitian Dan Pengabdian Kepada Masyarakat CORISINDO*, 287–292. <https://corisindo.stikom-bali.ac.id/penelitian/index.php/semnas/article/view/58>
- Dharmawan, O. P., & Handoyo, P. (2017). Fenomena Sneakers Impor (Studi Konstruksi Sosial Pemakaian Sneakers Impor Universitas Negeri Surabaya). *Paradigma: Journal of Sociology*, 5(1), 1–8. <https://ejournal.unesa.ac.id/index.php/paradigma/article/view/18327>
- Hajaroh, Suprapti, T., & Narasati, R. (2024). Implementasi Algoritma Naïve Bayes Untuk Analisis Sentimen. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(1), 111–118. <https://doi.org/https://doi.org/10.36040/jati.v8i1.8237>
- Hovi, H. S. W., Id Hadiana, A., & Rakhmat Umbara, F. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, 4(1), 40–45. <https://doi.org/10.36423/index.v4i1.895>
- Idris, I. S. K., Mustofa, Y. A., & Salih, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32–35. <https://doi.org/10.37905/jjeee.v5i1.16830>
- Jonathan, W. (2023). Pengaruh Brand Image Dan Brand Trust Terhadap Keputusan Pembelian Pada Sepatu Sneakers Merek Adidas Dan Nike Di Kota Palembang. *Konsumen & Konsumsi: Jurnal Manajemen*, 2(3), 491–506. <https://doi.org/10.32524/kkjm.v2i3.856>
- Maghfiroh, L., & Iriani, S. S. (2021). Niat Beli Sepatu Merek Lokal oleh Generasi Muda: Pengaruh Consumer Ethnocentrism, Perceived Quality, Perceived Price, dan Perceived Brand Image. *Jurnal Ilmu Manajemen*, 9(2), 617. <https://doi.org/10.26740/jim.v9n2.p617-633>
- Puad, S., Garno, G., & Susilo Yuda Irawan, A. (2023). Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1560–1566. <https://doi.org/10.36040/jati.v7i3.6920>
- Rizaty, M. A. (2024). *Kumpulan Data Seputar Perkembangan E-Commerce di Indonesia pada 2023 dan 2024*. <Https://Dataindonesia.Id/>. <https://dataindonesia.id/ekonomi-digital/detail/laporan-kumpulan-data-seputar-perkembangan-ecommerce-di-indonesia-pada-2023-dan-2024>
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(4), 377–383. <https://doi.org/10.47233/jteksis.v5i4.913>
- Safitri, R., Ali, I., & Rahaningsih, N. (2024). Analisis Sentimen Terhadap Tren Fashion Di Media Sosial Dengan Metode Support Vector Machine (Svm). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1746–1754. <https://doi.org/10.36040/jati.v8i2.9045>
- Setyabudi, S., Aryanny, E., Pembangunan, U., Veteran, N., Timur, J., & Timur, J. (2025).

- Sentiment Analysis of Lazada Marketplace User Ratings With Naïve Bayes And Support Vector Machine Methods Marketplace Lazada Dengan Metode Naïve Bayes. *INOVTEK Polbeng* - *Seri Informatika*, 10(1), 422–433. <https://doi.org/https://doi.org/10.35314/sww8cg21>
- Statista. (2025). *Footwear revenue of Nike, adidas & Puma 2010-2024* | Statista. <Https://Www.Statista.Com>. <https://www.statista.com/statistics/278834/revenue-nike-adidas-puma-footwear-segment/>
- Tim PR Tokopedia. (2025). *Bisa Jadi Ide Jualan Online, Ini Kategori Produk yang Paling Banyak Dicari di Tokopedia!* <Https://Www.Tokopedia.Com>. <https://www.tokopedia.com/blog/bisa-jadi-ide-jualan-online-ini-kategori-produk-yang-paling-banyak-dicari-di-tokopedia-rls/>
- Wibisono, A. D., Dadi Rizkiono, S., & Wantoro, A. (2020). Filtering Spam Email Menggunakan Metode Naive Bayes. *TELEFORTECH : Journal of Telematics and Information Technology*, 1(1), 9–15. <https://doi.org/10.33365/tft.v1i1.685>
- Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'Ban, S. R., Amalia, M., & Febriansyah, R. (2024). Pembelajaran Machine Learning. *OKTAL (Jurnal Ilmu Komputer Dan Science)*, 3(2), 375–380. <https://journal.mediapublikasi.id/index.php/oktal/article/view/2305>
- Yahya, A. K. (2023). Persepsi Masyarakat Terhadap Culture Fashion. *Sintesa*, 2(2), 113–123. <https://doi.org/https://doi.org/10.30996/sintesa.v2i02.8895>