

## Implementasi *Naïve Bayes* untuk Rekomendasi Pembelian Produk pada Aplikasi E-commerce

Boy Betrand Situngkir<sup>1</sup>, Endson Danielgar Limbong<sup>2</sup>, Very Andreas Pandiangan<sup>3</sup>, Rivaldo Calvin Siagian<sup>4</sup> Yennimar<sup>5</sup>

<sup>1,2,3,4,5</sup>Universitas Prima Indonesia, Jl. Sampul No.3, Medan, 20118, Indonesia.

### Info Artikel

#### Riwayat Artikel:

Received 2025-05-30

Revised 2025-06-29

Accepted 2025-07-15

**Abstract** – Electronic commerce (e-commerce) is a platform that influences buying and selling habits in Indonesia, with data from the Central Statistics Agency 2023 showing 31,753 e-commerce businesses using consumer review data as a determinant of product and service quality. This research aims to develop a sentiment-based product recommendation system using the Naïve Bayes algorithm. The research methodology includes collecting 1,287 data samples obtained from customer reviews using Web Scraper technology on the official MSI Official Store e-commerce platforms in the Tokopedia, Shopee, and Blibli applications. The results of data preprocessing yielded 921 clean data, and the Naïve Bayes Algorithm was applied as a classification model and system implementation in a website application. The data was then divided into 80% for training and 20% for testing. Model evaluation showed an accuracy of 82% for training data and 71% for testing data. These results indicate the effectiveness of the Naïve Bayes algorithm in forming a sentiment-based product recommendation system. This recommendation system helps users make more informed purchasing decisions based on consumer sentiment analysis. This research contributes to the development of intelligent recommendation systems that can improve user decision-making in the digital market.

**Keywords:** Consumer Review; E-Commerce; Naïve Bayes; Recommendation System; Sentiment Analysis.

#### Corresponding Author:

Boy Betrand Situngkir

Email:

boyetrandsitungkir@gmail.com



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

**Abstrak** – Electronic commerce (e-commerce) adalah platform yang mempengaruhi kebiasaan jual-beli di Indonesia, dengan data dari Badan Pusat Statistik 2023 terdapat 31.753 usaha e-commerce dengan data ulasan konsumen sebagai penentu kualitas produk dan layanan. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi produk berbasis sentimen menggunakan algoritma Naïve Bayes. Metodologi penelitian meliputi pengumpulan 1.287 sampel data yang diperoleh dari ulasan pelanggan menggunakan teknologi Web Scraper pada Platform e-commerce resmi MSI Official Store di aplikasi Tokopedia, Shoopee, dan Blibli. Hasil preprocessing data menghasilkan 921 data bersih dan diterapkan Algoritma Naïve Bayes sebagai model klasifikasi dan implementasi sistem dalam aplikasi website. Data kemudian dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi model menunjukkan akurasi sebesar 82% untuk data training dan 71% untuk data testing. Hasil ini menunjukkan efektivitas algoritma Naïve Bayes dalam membentuk sistem rekomendasi produk berbasis sentimen. Sistem rekomedasi ini membantu pengguna dalam membuat keputusan pembelian yang lebih informatif berdasarkan analisis sentimen konsumen. Penelitian ini memberikan kontribusi dalam pengembangan sistem rekomendasi cerdas yang dapat meningkatkan pengambilan keputusan pengguna di pasar digital.

**Kata Kunci:** Analisis Sentimen, E-Commerce, Naïve Bayes, Sistem Rekomendasi, Ulasan Konsumen

## I. PENDAHULUAN

Belanja adalah aktivitas penting bagi masyarakat Indonesia, tidak hanya untuk memenuhi kebutuhan, tetapi juga sebagai gaya hidup dan sumber kepuasan emosional, terutama bagi masyarakat perkotaan yang menjadikannya relaksasi [1]. Perkembangan teknologi digital dan informasi semakin mendorong transformasi cara berbelanja dari yang sebelumnya bersifat konvensional menuju model daring (online). Perkembangan digital telah mengubah cara belanja ke **online**, didorong oleh penetrasi internet dan preferensi masyarakat akan kemudahan dan efisiensi. Berbagai **platform belanja online** menawarkan diskon, *cashback*, dan pengiriman cepat. **Fitur ulasan** konsumen menjadi referensi dan indikator kepuasan [2]. Di sisi lain, proses psikologis dalam otak manusia juga memberikan kontribusi terhadap munculnya rasa senang dan dorongan positif saat berbelanja, terutama ketika mendapatkan promosi atau rekomendasi produk yang dianggap relevan.

Electronic commerce (e-commerce) sebagai bentuk implementasi dari digitalisasi perdagangan telah menjadi salah satu sektor yang mengalami pertumbuhan signifikan di Indonesia. E-commerce tidak hanya sekadar aktivitas jual beli barang dan jasa melalui internet, melainkan mencakup kolaborasi lintas sektor yang memperluas jangkauan pasar, meningkatkan efisiensi logistik, serta memperkuat sistem informasi bisnis [3]. Kemudahan yang ditawarkan dalam proses pembelian, pembayaran, hingga pengantaran menjadikan e-

commerce sebagai pilihan utama masyarakat dalam memenuhi kebutuhannya [4]. Berdasarkan data yang dirilis oleh Badan Pusat Statistik pada tahun 2023, perkembangan usaha e-commerce di Indonesia menunjukkan tren positif. Tercatat terdapat 31.753 usaha e-commerce yang tersebar di 302 kabupaten/kota di 34 provinsi, dengan dominasi pelaku usaha yang memiliki pendapatan di bawah Rp300 juta (82,97%). Sementara itu, hanya sebagian kecil (0,21%) pelaku usaha yang memiliki pendapatan lebih dari Rp50 miliar. Dari sisi pengeluaran, 89,06% usaha e-commerce mencatatkan pengeluaran di bawah Rp300 juta, menunjukkan bahwa e-commerce juga diminati oleh pelaku usaha kecil dan menengah sebagai model bisnis yang menjanjikan. Perubahan ini tentu tidak terlepas dari transformasi perilaku konsumen [5]. Arnold *et al* pada [6] menyatakan bahwa perkembangan teknologi digital dalam industri 4.0 telah mengubah lanskap bisnis secara menyeluruh, termasuk dalam perilaku belanja konsumen yang kini lebih mengandalkan platform daring, sistem pembayaran elektronik, serta informasi yang disediakan oleh ulasan konsumen. Sejalan dengan itu, Herawati dan Fasa menyebutkan bahwa perilaku konsumen merupakan rangkaian proses yang melibatkan pencarian informasi, evaluasi produk, hingga keputusan pembelian. Faktor-faktor seperti kenyamanan platform, kepercayaan terhadap penjual, serta pengaruh ulasan dari pengguna lain sangat memengaruhi pengambilan keputusan [7].

*Naïve Bayes* adalah metode klasifikasi dalam data mining yang efektif untuk mengekstraksi informasi dan memprediksi kelas objek yang tidak diketahui dari basis data besar, dengan memanfaatkan teknik statistik, matematika, dan machine learning [8]. *Naïve Bayes* adalah metode yang berlandaskan Teorema Bayes menggunakan teknik probabilitas dan statistik untuk memprediksi peluang berdasarkan data historis, dan telah terbukti memiliki kinerja tinggi dalam berbagai penelitian [9]. *Naïve Bayes* juga sangat efektif dan cepat dalam mengolah data berdimensi tinggi, seperti mengklasifikasikan ulasan, komentar, atau sentimen lain sebagai positif, negatif atau netral. Penelitian sebelumnya yang dilakukan oleh [10] menggunakan text mining untuk analisis sentimen pengguna Indodana di Play Store. Tahapan yang dilakukan dalam analisis sentimen pelanggan meliputi pengumpulan data, preprocessing, pelabelan data, visualisasi data, TF-IDF, *Naïve Bayes*, dan evaluasi. Hasil yang diperoleh dari penelitian ini dengan *accuracy* sebesar 86% dan pelabelan menggunakan *rating* 87%.

Adapun penelitian dari [11] difokuskan untuk memodelkan perilaku konsumen yang sadar lingkungan dengan *Partial Least Squares Structural Equation Modeling* (PLS-SEM) dan *Naïve Bayes*. Pada penelitian ini, riset dilakukan dengan menerapkan *Theory of Planned Behavior* (TPB) dengan variabel-variabel seperti sikap, norma subjektif, kontrol perilaku, kualitas produk, pengaruh sosial dan budaya. Kemudian model tersebut divalidasi dan menguji hipotesisnya terhadap 364 responden menggunakan PLS-SEM, dengan prediksi model menggunakan *Naïve Bayes*. Hasil penelitian mengungkapkan sebagian konsumen masuk dalam kategori "kurang sadar lingkungan" dengan *training model accuracy* sebesar 94,5% dan *testing model accuracy* sebesar 93%.

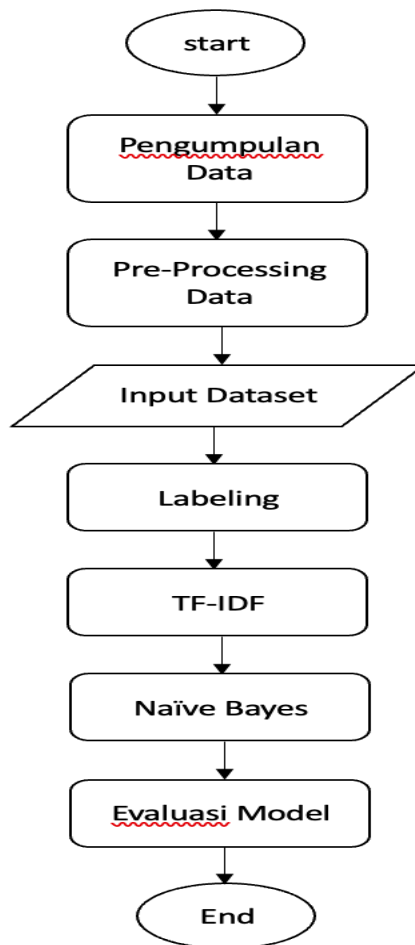
Sementara itu, penelitian yang dilakukan oleh [12] dengan menganalisis sebanyak 1000 ulasan pengguna yang diperoleh menggunakan teknologi *Web Scrapping*. Pada penelitian ini diterapkan *Knowledge Discovery in Databases* (KDD) untuk mengekstraksi pengetahuan yang relevan dan berguna yang diolah dengan sejumlah algoritma. Penelitian dapat mengambil solusi untuk meningkatkan layanan dan meningkatkan kepuasan terhadap aplikasi DANA dengan analisis kelemahan pada layanan berdasarkan sentimen menggunakan algoritma *Naïve Bayes* berdasarkan sentimen ulasan pelanggan. Hasil yang diperoleh pada penelitian ini dihasilkan sebanyak 576 ulasan positif, ulasan netral sebanyak 205, dan ulasan negatif sebanyak 219 ulasan. Metode *Naïve Bayes* menghasilkan *Performance Vector* didapat *accuracy*: 74.60% +/- 3.53% (*micro average* : 74.60%) dan *kappa* : 0.550 +/- 0.061 (*micro average* : 0.550).

Penelitian yang dilakukan [13] dalam analisis review aplikasi Shopee menggunakan *Naïve Bayes*, N-Gram, dan *Information Gain* memperoleh tingkat *accuracy* sebesar 92%, *precision* sebesar 56%, *recall* sebesar 65%, dan *F1-Score* sebesar 60%. Pada penelitian lain oleh [14] dalam analisis opini publik pada otomatisasi AI dalam layanan restoran robotik dengan menggunakan algoritma *Naïve Bayes* dan *Synthetic Minority Over-sampling Technique* (SMOTE) menghasilkan tingkat *accuracy* sebesar 70,11% dan *accuracy* tanpa SMOTE sebesar 48,90%. Penelitian oleh [15] dalam analisis sentimen pengguna terhadap aplikasi dompet digital pada platform media sosial Twitter (x) mendapatkan tingkat *accuracy* sebesar 72% untuk DANA, 88% untuk GoPay, dan 88% untuk LinkAja. Namun, penelitian ini akan menganalisis sentimen ulasan konsumen pada Platform e-commerce resmi MSI Official Store di aplikasi Tokopedia, Shoopee, dan Blibli. Untuk membangun sistem rekomendasi produk berdasarkan ulasan konsumen sehingga pengelola usaha dapat meningkatkan kualitas produk pada aplikasi e-commerce untuk membangun kepercayaan antara konsumen dan pengelola usaha.

## II. METODE

*Naïve Bayes* dipilih dalam penelitian ini karena cukup mudah dalam implementasinya, efisiensi dalam mengolah data teks berukuran besar, serta kemampuannya memberikan hasil yang kompetitif meskipun dibandingkan dengan algoritma yang lebih kompleks. Sehingga sangat cocok untuk klasifikasi teks [8]. Keunggulan dalam hal kecepatan dan efisiensi menjadikannya ideal untuk memproses data dalam jumlah besar, seperti ulasan customer pada Tokopedia, Shopee, Blibli [9].

Sebagaimana ditampilkan oleh flowchart pada gambar 1, proses analisis sentimen pada penelitian ini mencakup enam tahap utama, yaitu: pengumpulan data, pra-pemrosesan, pelabelan berdasarkan kamus sentimen, pembobotan menggunakan TF-IDF, klasifikasi dengan algoritma *Naïve Bayes*, dan evaluasi model. Hasil klasifikasi dikategorikan ke dalam tiga jenis sentimen: positif, netral, dan negatif.



Gambar 1. Flowchart

#### A. Pengumpulan Data

Hasil data yang diperoleh digunakan dalam penelitian ini diperoleh melalui platform e-commerce resmi MSI Official Store pada Tokopedia, Shopee, dan Blibli pada ulasan pembelian customer. Teknik pengambilan data dilakukan menggunakan extension pada google chrome yaitu google scraping [16]. Dataset dari masing-masing e-commerce yang dihasilkan dari scraping data kemudian diunduh dalam format csv yang terdiri dari 500 data pada Tokopedia, 691 data untuk shopee, 99 data untuk Blibli. Dataset berisi ulasan konsumen yang diambil dari awal tahun 2024 – 2025.

#### B. Pre-Processing Data

Pada tahap pre-processing data, data duplikat dihapus dan fitur yang tidak relevan seperti hastag tidak disertakan dalam analisis. Seluruh proses pembersihan data dilakukan menggunakan python dengan bantuan pustaka *Natural Language Tollkit* (NLTK) yang umum digunakan untuk pemrosesan bahasa alami [17].

Langkah-langkah *Pre-processing* meliputi penghapusan retweet, URL, tanda baca, serta konversi hruuf dan emoji. Selanjutnya dilakukan tokenizing untuk memecah kalimat menjadi kata, stopword removal untuk membuang kata-kata yang tidak penting, normalization untuk menyamakan bentuk kata, dan stemming untuk mengembalikan kata ke bentuk dasar. total keseluruhan data yang telah disatukan dari 3 platform yang berbeda berjumlah 921 data. Proses ini bertujuan untuk meningkatkan kualitas data sebelum masuk ke dalam klasifikasi.

### C. Labeling

Klasifikasi data ulasan dilakukan dengan menggunakan kamus sentimen yang dibagi ke dalam dua kategori utama, yaitu sentimen positif dan sentimen negatif. Kata-kata sentimen ini disusun dalam file csv dan tidak menggunakan basis data. Penentuan kategori dilakukan berdasarkan kemunculan kata kunci tertentu dalam teks ulasan [18].

Sebagai contoh, kata “bagus”, “mantap”, dan “cepat” dikategorikan sebagai sentimen positif, sementara kata “jelek”, “tidak sesuai”, dan “lambat” dianggap sebagai sentimen negatif. Selain itu, intensitas makna juga diperhatikan, di mana frasa seperti “sangat puas” memiliki bobot sentimen yang lebih kuat dibanding “puas”. Kata negasi seperti “tidak” juga diperhitungkan karena dapat membalikkan makna sentimen dari kata berikutnya.

### D. TF-IDF Weighting

TF-IDF (*Term Frequency-Inverse Document Frequency*), yang menggabungkan dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Teknik ini digunakan untuk menghitung bobot atau tingkat kepentingan suatu kata dalam sebuah ulasan terhadap keseluruhan kumpulan ulasan konsumen yang dianalisis [19]. Setelah data ulasan dikonversi ke dalam bentuk numerik menggunakan TF-IDF, langkah selanjutnya dalam pembangunan model *machine learning* adalah melakukan pembagian data menjadi dua bagian, yaitu 80% untuk data pelatihan (training) dan 20% untuk data pengujian (testing), sesuai dengan proporsi yang digunakan dalam penelitian ini. Pembobotan TF-IDF ini digunakan sebagai representasi fitur untuk memaksimalkan performa klasifikasi oleh algoritma *Naïve Bayes*. Hasil dari proses pembobotan ini menghasilkan daftar kata-kata penting (fitur) dalam setiap ulasan yang paling berkontribusi terhadap klasifikasi sentimen, baik positif maupun negatif. Fitur-fitur tersebut menjadi dasar dalam pengambilan keputusan sistem rekomendasi.

### E. Naïve Bayes Classification

Eksperimen dilakukan dengan membagi dataset menjadi dua bagian, yaitu 80% data pelatihan (training) dan 20% data pengujian (testing). Tujuan dari proses klasifikasi ini adalah membangun model prediksi berdasarkan pola yang dipelajari dari data pelatihan. Model kemudian diuji menggunakan data pengujian untuk mengevaluasi performanya dalam mengklasifikasikan sentimen pada data yang belum pernah dilihat sebelumnya. Algoritma *Naïve Bayes* yang digunakan dalam penelitian ini adalah jenis *Multinomial Naïve Bayes*, yang secara khusus cocok untuk data teks. Model ini bekerja dengan menghitung kemungkinan (probabilitas) suatu teks termasuk ke dalam kategori tertentu (positif atau negatif), dan memilih kategori dengan probabilitas tertinggi sebagai output klasifikasi [20]. Evaluasi terhadap performa model dilakukan dengan menggunakan *confusion matrix* dan metrik evaluasi lainnya seperti akurasi, precision, recall, dan F1-score, yang bertujuan untuk mengetahui seberapa baik model dapat mengklasifikasikan ulasan produk berdasarkan sentimen pengguna.

### F. Evaluasi Model

Tahap evaluasi model adalah fase krusial yang dilakukan untuk mengukur sejauh mana performa atau kinerja algoritma klasifikasi yang digunakan dalam penelitian dapat memberikan prediksi yang tepat. Evaluasi ini bertujuan untuk memastikan bahwa model yang dikembangkan dapat mengenali pola data secara akurat dan dapat diterapkan dalam pengambilan keputusan yang tepat. Dalam tahap evaluasi, diterapkan sejumlah indikator atau metrik evaluasi yang bersifat kuantitatif. Metrik-metrik ini memungkinkan peneliti untuk menilai tingkat efektivitas dan ketepatan prediksi model, serta memahami kekurangan dan keunggulannya. Metrik yang digunakan dalam penelitian ini mencakup akurasi, presisi, recall, dan F1-score.

## III. HASIL DAN PEMBAHASAN

### A. Labeling Data

Labeling data diperlukan untuk membuat label positif atau negatif ulasan customer. Proses ini ditentukan oleh skor yang diberikan pada setiap komentar yang ada, setiap skor diatas 1 memenuhi syarat untuk label positif, sedangkan skor dibawah 1 akan diberikan label nefatif, yang dimana setiap skor akan ditentukan dari kamus positif dan kamus negatif.

	komentar	rating	sentimen
0	laptop sudah terima respon cepat packing rapih...	5	positif
1	alhamdulillah sampai cuma 2 hari spec sesuai d...	5	negatif
2	jujur ini alam pertama saya beli barang elektr...	5	negatif
3	barang sesuai dengan deskripsi so far performa...	5	positif
4	barang bagus sesuai	5	positif

Gambar 2. Text Labeling

### B. Teks Preprocessing

Kemudian beberapa sampel data yang sudah di cleaning akan dilanjutkan ke tahap teks processing. Langkah-langkah Pre-processing meliputi penghapusan retweet, URL, tanda baca, serta konversi huruf dan emoji. Selanjutnya dilakukan tokenizing untuk memecah kalimat menjadi kata, stopword removal untuk membuang kata-kata yang tidak penting, normalization untuk menyamakan bentuk kata, dan stemming untuk mengembalikan kata ke bentuk dasar, penting untuk diingat setiap proses di atas memiliki peran yang sangat kritis dalam memproses teks dengan efektif yang nantinya akan di implimentasikan ke dalam evaluasi model

### C. Hasil Penfumpulan Data

Beberapa sampel data yang data ditampilkan untuk training data berdasarkan kata kunci positif dan negatif yang paling sering muncul pada dataset yang telah didapatkan ditampilkan dalam tabel berikut :

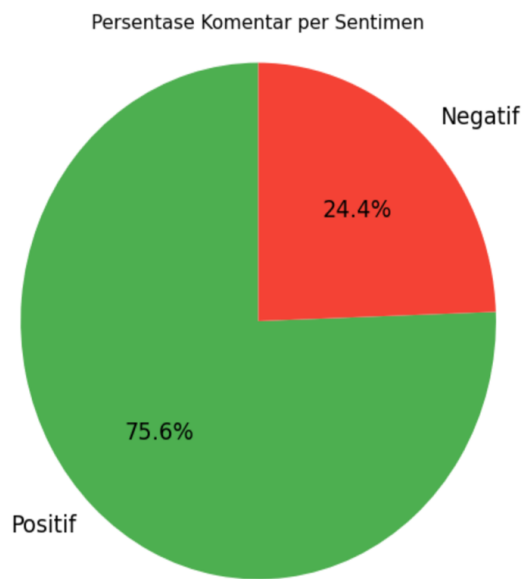
TABEL 1.  
 TABEL DATA TEKS PRE-PROCESSING

No. Review	Review	Skala Review	Kategori
1	Laptop sudah diterima, Respon cepat, packing rapih, barang sesuai dan berfungsi, terima kasih	4	Positif
2	Barang sesuai dengan deskripsi, so far performa mantap, belum ada kendala. semoga awet.	3	Positif
3	Tidak ada packing kayu	-1	Negatif
4	Alhamdulillah sampai cuma 2 hari, spec sesuai dan tidak ada defect selama pengiriman	0	Negatif
5	co packing kayu dengan harapan menyediakan keamanan lebih utk laptop yg dibeli. Alhamdulillah pertanyaan terkait packing kayu ini tidak dijawab menjelang co, juga tidak tampak sepotong kayu pun pada paket laptop yg dibeli. jadi maaf tidak ada fotonya	-1	Negatif
6	Barang sesuai dengan deskripsi, so far performa mantap, belum ada kendala. semoga awet.	3	Positif
7	Barang bagus sesuai	2	Positif
8	Agak kecewa sih dengan kualitas barangnya. saya komplain via chat terkait casing belakang yg longgar dan kurang rapat (dipencet2 masuk kedalam). katanya mau di cek tapi tidak ada follow up nya. menurut saya juga overheating, kipas nyala keras bahkan saat laptop standby di posisi desktop home. sudah coba update wondows dan setting msi center tetap tidak ngaruh Efeknya ke baterai.	-2	Negatif

...	...	...	...
...	...	...	...
921	Saya sudah coba masukan Code Product Key nya, tapi selalu dijawab Kode sudah diklaim oleh orang lain. Penjual tidak merespon komplain saya.	-1	Negatif

*D. Visualisasi Label*

Data yang sudah diolah divisualisasikan dengan pemberian label positif dan negatif. terlihat hasil dari analisis sentiment terbilang cukup baik dengan perolehan sentiment label positif dengan nilai 75,6 % lebih besar dibanding label negatif yang mendapat nilai 24,4%.



Gambar 3. Visualisasi Label

*E. Splitting Data*

Hasil dari pre-processing data yang telah dilakukan, kemudian akan dilakukan splitting data terhadap 921 data komentar yang dibagi menjadi 80:20 yang dimana 80% data training dan data 20% data testing, yang menampilkan jumlah total dari data training dan data testing yang sudah di klasifikasi menjadi sentiment positif dan negatif seperti gambar di bawah ini :

Perbandingan Jumlah Data Training dan Testing per Kelas:

Class	No	Data Training	Data Testing
positif	1	567	129
negatif	2	169	56
Total	3	736	185

Gambar 4. Hasil Splitting Data

*F. Pembangunan Model Machine Learning*

Algoritma yang digunakan untuk pembangunann model machine learning pada penilitan ini menggunakan algoritma *Naïve Bayes*, setelah itu penelitian akan menguji model yang sudah dibangun dengan menggunakan data yang sudah di bagi menjadi 80:20 dan mencari hasil dari nilai akurasi dan classification reportnya, berikut hasil dari penggunaan model machine learning:

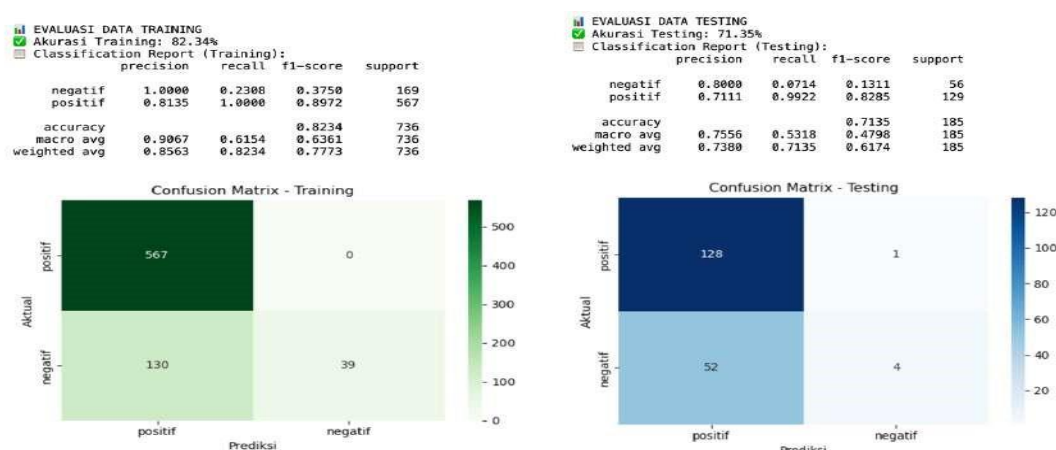
Classification Report:

	precision	recall	f1-score	support
negatif	0.80	0.07	0.13	56
positif	0.71	0.99	0.83	129
accuracy			0.71	185
macro avg	0.76	0.53	0.48	185
weighted avg	0.74	0.71	0.62	185

Accuracy Score: 71.35%

Gambar 5. Accuracy dan Classification Report

Hasil dari penelitian ini menunjukkan akurasi sebesar 71% pada data testing dan 82% pada data training. Jika dibandingkan dengan beberapa studi sebelumnya, seperti penelitian oleh Rifki Kosasih & Anggi [20] yang memperoleh akurasi 80%, serta penelitian Adelia Nopiazinda Br Panjaitan & Ismail Husein [21] yang mencatat akurasi 90%, maka hasil penelitian ini berada pada tingkat yang lebih rendah. Meski demikian, hasil ini masih kompetitif, khususnya mengingat bahwa data yang digunakan berasal dari tiga platform berbeda dan menggunakan pendekatan labeling berbasis kamus sentimen, bukan rating pengguna. Lalu hasil dari perhitungan data testing dan data training nya menggunakan model *machine learning Naïve Bayes* serta menambahkan *confusion matrix* nya menghasilkan akurasi dan *classification report* seperti gambar dibawah ini :



Gambar 6. Hasil Confusion Matrix

Perbedaan akurasi antara data training (82,34%) dan testing (71,35%) disebabkan oleh beberapa faktor, terutama overfitting, yaitu kondisi ketika model terlalu menghafal data training sehingga kurang mampu mengenali pola baru pada data testing. Selain itu, ketidakseimbangan data (*imbalanced data*) antara kelas positif dan negatif juga membuat model cenderung memprediksi kelas mayoritas (positif), yang terlihat dari recall kelas negatif yang sangat rendah.

#### G. Menerapkan Model Machine Learning pada Aplikasi Website

Setelah memperoleh hasil dari data yang telah diproses melalui model machine learning, tahap selanjutnya adalah mengembangkan sebuah website interaktif yang berfungsi sebagai alat bantu konsumen dalam mengambil keputusan pembelian. Website ini memungkinkan pengguna untuk mencari produk tertentu dan menampilkan hasil analisis sentimen berdasarkan komentar atau ulasan pengguna.

Fitur utama dari website ini meliputi: kolom input pencarian produk, output berupa jumlah komentar **positif** dan **negatif**, serta visualisasi dalam bentuk diagram batang untuk memperjelas proporsi sentimen. Selain itu, sistem memberikan indikator **rekomendasi** secara otomatis berdasarkan dominasi sentimen positif terhadap suatu produk. Dengan demikian, konsumen dapat dengan cepat memahami persepsi umum terhadap produk tersebut sebelum melakukan pembelian.

Fungsi teknis dari sistem ini adalah mengintegrasikan hasil klasifikasi sentimen dari model machine learning ke dalam tampilan antarmuka web secara real-time, sehingga mendukung proses pengambilan keputusan konsumen secara lebih informatif dan efisien.



Gambar 7. Tampilan Website

#### H. Analisis Kritis

Model *Naïve Bayes* yang digunakan memiliki kelebihan dalam efisiensi komputasi dan kemudahan implementasi, namun kelemahannya terletak pada asumsi independensi antar fitur yang sering kali tidak terpenuhi dalam data teks nyata. Selain itu, ketidakseimbangan kelas (jumlah sentimen positif jauh lebih banyak daripada negatif) menyebabkan penurunan performa terutama dalam hal recall pada kelas negatif. Hal ini berpotensi menyebabkan bias model dalam memprediksi produk sebagai layak direkomendasikan meskipun tidak sepenuhnya didukung oleh distribusi ulasan negatif.

#### I. Keterbatasan Penelitian

Beberapa keterbatasan yang dapat diidentifikasi dalam penelitian ini antara lain:

1. Labeling dilakukan secara otomatis menggunakan kamus sentimen, bukan manual atau rating-based, sehingga rentan terhadap kesalahan interpretasi makna konteks.
2. Dataset terbatas pada satu brand (MSI Official Store) dan hanya tiga platform (Tokopedia, Shopee, Blibli), sehingga generalitas model terbatas.
3. Model belum mengakomodasi sentimen netral dan belum memanfaatkan fitur-fitur lanjutan seperti N-gram atau word embedding.

## IV. SIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi produk berbasis sentimen menggunakan algoritma *Naïve Bayes*, sesuai dengan tujuan penelitian. Penelitian ini menunjukkan bahwa implementasi algoritma *Naïve Bayes* mampu memberikan klasifikasi yang cukup efektif terhadap sentimen konsumen. Hasil pengumpulan data sebanyak 1.287 baris data. Hasil data kemudian dilakukan preprocessing data dan menghasilkan data bersih sebanyak 921 ulasan yang terbagi menjadi 696 ulasan positif (76%) dan 225 ulasan negatif (24%). Penggunaan pendekatan probabilistik dalam algoritma *Naïve Bayes* memungkinkan model untuk mempelajari distribusi kata kunci yang umum muncul dalam sentimen positif dan negatif. Dengan membagi data ke dalam dua bagian yaitu 80% data pelatihan (736 data) dan 20% data pengujian (185 data), model yang dibangun mampu menghasilkan akurasi sebesar 71% pada data testing dan 82% pada data training. Nilai precision untuk klasifikasi negatif mencapai 80%, sedangkan precision untuk klasifikasi positif mencapai 71%, yang menunjukkan bahwa algoritma ini cukup andal dalam mengenali kecenderungan sentimen berdasarkan distribusi kata yang muncul dalam teks ulasan.

Temuan utama penelitian ini adalah keberhasilan algoritma *Naïve Bayes* dalam membangun sistem rekomendasi produk berbasis sentimen dalam bentuk aplikasi website. Sistem rekomendasi ini membantu pengguna dalam membuat keputusan pembelian yang lebih informatif berdasarkan analisis sentimen konsumen. Meskipun menunjukkan efektivitas, penelitian ini memiliki beberapa keterbatasan. Pelabelan data dilakukan secara otomatis menggunakan kamus sentimen, sehingga rentan terhadap kesalahan interpretasi makna kontekstual. Data sentimen konsumen yang digunakan dalam penelitian ini diperoleh dari website e-commerce resmi MSI Official Store pada beberapa platform e-commerce seperti Tokopedia, Shopee, dan Blibli. Untuk penelitian mendatang penting untuk memperkaya variasi dataset dengan menambahkan ulasan dari brand dan platform e-commerce yang berbeda. Selain itu, model dapat ditingkatkan performanya dengan mengintegrasikan klasifikasi sentimen netral dan mengeksplorasi penggunaan fitur-fitur yang lebih kompleks seperti N-gram atau word embedding.

**DAFTAR PUSTAKA**

- [1] M. Andika, S. Masithoh, Y. N. Kholiq, D. A. Nisa, and N. Rohmah, "Efektivitas Marketplace Shopee sebagai Marketplace Belanja Online yang Paling Disukai Mahasiswa," *J. Educ. Technol.*, vol. 1, no. 1, pp. 24–29, 2021.
- [2] M. Saputra *et al.*, "Jurnal Tamwil: Jurnal Ekonomi Islam DAMPAK E-COMMERCE TERHADAP PENDAPATAN PEDAGANG SANDANG DI PASAR BATUSANGKAR," *101 / J. Tamwil*, vol. 9, no. 2, 2023, [Online]. Available: <http://ecampus.iainbatangsangkar.ac.id/ojs/index.php/tamwil/index>
- [3] Y. L. . Rehatalanit, "Peran E-Commerce Dalam Pengembangan Bisnis," *J. Teknol. Ind.*, vol. 5, no. 0, pp. 62–69, 2021, [Online]. Available: <https://journal.universitassuryadarma.ac.id/index.php/jti/article/view/764>
- [4] S. Utami, H. Artikel, and K. Kunci, "Belanja online menjadi trend remaja selama pandemi covid-19 A B S T R A K," vol. 13, no. 1, pp. 730–738, 2020, [Online]. Available: <https://jurnal.iain-bone.ac.id>
- [5] Badan Pusat Statistik, "Statistik eCommerce 2022/2023," <https://www.bps.go.id/publication/2023/09/18/f3b02f2b6706e104ea9d5b74/statistik-ecommerce-2022-2023.html>, Sep. 18, 2023.
- [6] D. Strategi, B. Arnold, R. S. Dewi, I. S. Wijaya, and D. F. Sugiyanti, "Dampak E-Commerce Terhadap Perilaku Konsumen," vol. 1, no. 2, pp. 56–66, 2022.
- [7] T. Herawati and M. Iqbal Fasa, "Perilaku Konsumen dalam Berbelanja Online dimasa Pandemi Covid-19." [Online]. Available: <http://journal.scimadly.com/index.php/iefj>
- [8] K. Nurnasikha, S. Farisa, and C. Haviana, "Klasifikasi Bidang Ilmu Publikasi Ilmiah Terindeks SINTA Menggunakan Metode *Naïve Bayes*."
- [9] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [10] R. Rizaldi and R. Aryanti, "Analisis Sentimen Pengguna Terhadap Aplikasi Indodana Di Google Play Store Menggunakan Metode *Naive Bayes Classifier*," *J. Informatics Manag. Inf. Technol.*, vol. 3, no. 4, pp. 98–105, 2024, doi: 10.47065/jimat.v4i3.400.
- [11] E. Wulandari, A. Turnip, and Y. Deliana, "Evaluation of Environment Friendly Consumer Behavior with PLS-SEM Method and *Naive Bayes*." [Online]. Available: <https://ssrn.com/abstract=4981139>
- [12] I. N. Kusuma *et al.*, "ANALISIS SENTIMEN PADA PENGGUNA APLIKASI DANA MENGGUNAKAN ALGORITMA NAIVE BAYES," 2024. [Online]. Available: <https://play.google.com/store/apps/details?id=id.dana>.
- [13] A. Sunjaya, N. Wijaya, N. P. Wong, and S. Winardi, "Implementation of Sentiment Analysis of Shopee E-Commerce Reviews using *Naive Bayes*, N-Gram, and Information Gain," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381959.
- [14] Y. A. Singgalen, "KLIK: Kajian Ilmiah Informatika dan Komputer Social Network Analysis and Sentiment Classification of Robotic Restaurant Content using *Naive Bayes Classifier*," *Media Online*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1710.
- [15] A. Basir, T. Yudha, P. Muhamad, A. Abdillah, and A. Faizin, "Jurnal Informatika : Jurnal pengembangan IT Analysis of Electronic Wallet User Sentiment on Twitter ( x ) Social Media Using the *Naive Bayes Classifier* Algorithm," vol. 10, no. 1, pp. 79–85, 2025, doi: 10.30591/jpit.v10i1.8180.
- [16] A. Suryadi, W. A. Syb'an, N. Alfa'inna, and E. H. Hermaliani, "Implementasi Web Scraping dan Sentiment Analysis Terhadap Berita Menggunakan Machine Learning," *Swabumi*, vol. 11, no. 1, pp. 28–34, 2023, doi: 10.31294/swabumi.v11i1.15145.
- [17] S. Lestanti and S. N. Budiman, "Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan *Naive Bayes*," vol. 5, no. December, pp. 114–126, 2024.
- [18] A. Rosadi *et al.*, "Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan *Naive Bayes Classifier* Twitter Text Mining," vol. 20, pp. 39–52, 2021.
- [19] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi *Naive Bayes* dalam menganalisis sentimen data ulasan PLN Mobile," *Petir*, vol. 15, no. 2, pp. 264–275, 2022, doi: 10.33322/petir.v15i2.1733.
- [20] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *Ilk. J. Ilm.*, vol. 13, no. 2, pp. 101–109, 2021, doi: 10.33096/ilkom.v13i2.721.101-109.
- [21] U. Islam, N. Sumatera, and U. Medan, "Sentiment Analysis Of Shopee Application User Reviews Using *Naive Bayes Classifier*," vol. 14, no. 03, pp. 317–324, 2024, doi: 10.54209/infosains.v14i03.