

Penerapan Algoritma Naïve Bayes Terhadap Sentimen Ulasan Produk Skincare Pada E-Commerce Shopee

Divana Wahyu Putri*, Moch Arief Soeleman

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ^{1,*}111202214456@mhs.dinus.ac.id, ²m.arief.soeleman@dsn.dinus.ac.id

Email Penulis Korespondensi: 111202214456@mhs.dinus.ac.id

Submitted: 14/01/2026; Accepted: 05/03/2026; Published: 06/03/2026

Abstrak—Pesatnya pertumbuhan industri kecantikan menghasilkan volume ulasan konsumen yang besar, sehingga diperlukan sistem pengolahan otomatis untuk memahami sentimen publik. Penelitian ini bertujuan untuk mengimplementasikan analisis sentimen pada ulasan produk *skincare* menggunakan algoritma *Multinomial Naïve Bayes*. Proses pelabelan dilakukan dengan mengonversi skor *rating* menjadi kategori sentimen, di mana *rating* 4 dan 5 menjadi label positif, serta *rating* 1 dan 2 menjadi label negatif, sementara *rating* 3 dieksklusi untuk menghindari ambiguitas data. Tahap representasi fitur menggunakan *TF-IDF* dengan pendekatan *N-gram* (unigram dan bigram) yang menghasilkan 10.000 fitur dari dataset sebanyak 8.646 ulasan. Berdasarkan hasil pengujian pada 1.730 data uji, model mencapai nilai akurasi sebesar 70,06%. Hasil evaluasi *Confusion Matrix* menunjukkan bahwa model memiliki performa sangat kuat pada kelas positif dengan nilai *recall* mencapai 1.00. Namun, model mengalami kesulitan dalam mengklasifikasikan kelas negatif dan netral yang memiliki nilai *recall* mendekati 0.00. Hal ini disebabkan oleh ketidakseimbangan distribusi data (*imbalanced data*), di mana jumlah ulasan positif jauh lebih mendominasi dibandingkan kelas lainnya. Meskipun demikian, penggunaan *Multinomial Naïve Bayes* terbukti efisien dalam menangani fitur tekstual berbasis frekuensi dalam jumlah besar. Nilai rata-rata tertimbang (*weighted average*) pada *F1-score* sebesar 0.58 menunjukkan bahwa optimasi dataset diperlukan untuk meningkatkan kemampuan model dalam mengenali sentimen minoritas secara lebih akurat.

Kata Kunci: Analisis Sentimen; Skincare; Naïve Bayes; TF-IDF; N-gram; Confusion Matrix

Abstract—The rapid growth of the beauty industry has generated a large volume of consumer reviews, necessitating an automated processing system to understand public sentiment. This study aims to implement sentiment analysis on skincare product reviews using the **Multinomial Naïve Bayes** algorithm. The labeling process was conducted by converting star ratings into sentiment categories: ratings 4 and 5 were labeled as positive, ratings 1 and 2 as negative, while rating 3 was excluded to avoid data ambiguity. The feature representation stage utilized **TF-IDF** with an **N-gram** approach (unigram and bigram), generating 10,000 features from a dataset of 8,646 reviews. Based on the testing results of 1,730 test data, the model achieved an **accuracy of 70%**. The **Confusion Matrix** evaluation revealed that the model performed exceptionally well in the positive class, reaching a recall of 1.00. However, the model struggled to classify negative and neutral classes, with recall values approaching 0.00. This was caused by **imbalanced data** distribution, where positive reviews significantly dominated the dataset. Nevertheless, Multinomial Naïve Bayes proved efficient in handling large-scale frequency-based textual features. A weighted average F1-score of 0.58 suggests that dataset optimization is required to improve the model's ability to accurately recognize minority sentiments.

Keywords: Sentiment Analysis; Skincare; Naïve Bayes; TF-IDF; N-Gram; Confusion Matrix.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi di era digital telah membawa perubahan fundamental dalam perilaku konsumsi masyarakat Indonesia. Salah satu sektor yang mengalami transformasi paling signifikan adalah perdagangan elektronik atau *e-commerce*. Platform besar seperti Shopee telah menjadi pusat transaksi berbagai produk, di mana kategori *skincare* atau perawatan kulit konsisten menempati posisi teratas dalam volume penjualan. Di tengah persaingan pasar yang ketat, ulasan pelanggan (*customer reviews*) menjadi instrumen vital yang menjembatani antara ekspektasi konsumen dan realitas kualitas produk. Data ulasan ini bukan sekadar kumpulan teks, melainkan representasi dari pengalaman pengguna yang mengandung sentimen positif, negatif, maupun netral.

Masalah utama yang muncul adalah volume data ulasan yang sangat masif dan terus bertambah setiap detiknya. Mengolah ribuan ulasan secara manual untuk mengetahui tren kepuasan pelanggan adalah hal yang mustahil dilakukan secara efisien. Oleh karena itu, diperlukan sebuah pendekatan otomatis berbasis sains data untuk mengekstraksi makna dari teks-teks tersebut. Analisis sentimen hadir sebagai solusi melalui pemanfaatan teknik *text mining* dan *Natural Language Processing* (NLP). Sebagaimana dijelaskan dalam penelitian sebelumnya, analisis sentimen merupakan bentuk representasi dari pemrosesan teks yang bertujuan untuk mengidentifikasi opini publik terhadap suatu fenomena atau produk tertentu [1]. Pemanfaatan analisis ini memungkinkan perusahaan untuk mendeteksi secara dini masalah pada produk *skincare* mereka sebelum berdampak lebih luas pada citra merek.

State of the Art dalam penelitian analisis sentimen saat ini menunjukkan pergeseran dari sekadar pengenalan kata kunci menuju pemahaman konteks yang lebih dalam. Penelitian terkini di bidang *e-commerce* mulai mengintegrasikan fitur-fitur linguistik yang spesifik pada bahasa lokal, mengingat ulasan pelanggan sering kali menggunakan dialek atau bahasa prokem (slang). Meskipun penggunaan model berbasis *Deep Learning* seperti Transformer sedang populer, penggunaan algoritma berbasis probabilitas seperti Naïve Bayes tetap menjadi standar acuan (*benchmark*) yang sangat relevan karena efisiensi sumber dayanya pada teks pendek. Kebaruan penelitian ini terletak pada optimalisasi prapemrosesan data yang disesuaikan dengan karakteristik unik kosakata industri kecantikan di Indonesia, yang menggabungkan istilah medis/kimia dengan bahasa kasual pengguna media sosial.

Dalam implementasi analisis sentimen, kualitas data yang diolah sangat menentukan hasil akhir. Data ulasan pada platform *e-commerce* sering kali bersifat tidak terstruktur, mengandung kata-kata tidak baku, ejaan yang salah, hingga penggunaan bahasa gaul (*slang*). Untuk mengatasi hambatan tersebut, tahapan *preprocessing* data menjadi krusial. Proses ini meliputi *cleansing*, *case folding*, *tokenizing*, *normalisasi*, *stopword removal*, hingga *stemming* [2]. Ketajaman prapemrosesan ini terbukti mampu meningkatkan akurasi model dalam membedakan berbagai sentimen yang kompleks. Studi terkait menunjukkan bahwa tanpa *preprocessing* yang baik, model klasifikasi akan kesulitan mengenali pola kata yang memiliki bobot emosional [3].

Pemilihan algoritma klasifikasi menjadi inti dari penelitian ini. Algoritma **Naïve Bayes Classifier (NBC)** telah lama dikenal sebagai salah satu algoritma terbaik dalam pengolahan teks karena kesederhanaan teorinya yang berbasis pada teorema probabilitas Bayes. NBC bekerja dengan asumsi bahwa setiap fitur (kata) bersifat independen satu sama lain, yang dalam banyak kasus teks terbukti sangat efektif dan efisien dari segi waktu komputasi [4]. Keunggulan NBC terletak pada kemampuannya memberikan hasil yang kompetitif meskipun menggunakan dataset yang tidak terlalu besar, serta performanya yang stabil dalam menangani data teks dengan fitur yang sangat banyak [5]. Hal ini sangat relevan dengan ulasan *skincare* di Shopee yang memiliki variasi kosakata yang sangat luas namun sering kali memiliki pola yang berulang.

Meskipun NBC memiliki banyak keunggulan, beberapa penelitian sering membandingkannya dengan metode lain seperti *Support Vector Machine (SVM)* untuk melihat model mana yang lebih tangguh. SVM dikenal memiliki keunggulan dalam menangani dataset yang kompleks dan non-linear, namun sering kali membutuhkan daya komputasi yang lebih besar dibandingkan NBC [6]. Selain itu, perbandingan juga dilakukan dengan metode *Deep Learning* seperti *Long Short-Term Memory (LSTM)*. Meskipun LSTM mampu menangani konteks kalimat yang panjang, NBC tetap dianggap lebih praktis untuk implementasi cepat pada aplikasi yang membutuhkan kecepatan respons tinggi [7].

Penelitian ini memberikan kontribusi signifikan dalam tiga aspek utama. Pertama, secara **metodologis**, penelitian ini menawarkan urutan prapemrosesan data yang spesifik untuk menangani "noise" pada ulasan *e-commerce* Indonesia, sehingga meningkatkan reliabilitas hasil klasifikasi. Kedua, secara **praktis**, sistem yang dikembangkan memberikan kemudahan bagi konsumen untuk mendapatkan ringkasan kualitas produk secara objektif tanpa harus membaca ribuan ulasan secara manual. Ketiga, bagi **pelaku usaha**, penelitian ini berkontribusi sebagai alat intelijen pasar yang mampu memetakan kekuatan dan kelemahan produk berdasarkan parameter teknis kecantikan yang paling sering dikeluhkan atau dipuji oleh pelanggan. Keempat, bagi akademis, Memperkaya literatur mengenai efektivitas Naïve Bayes pada domain spesifik industri kecantikan, yang memiliki karakteristik terminologi teknis (seperti *niacinamide*, *breakout*, *moisturizing*) yang tercampur dengan bahasa sehari-hari.

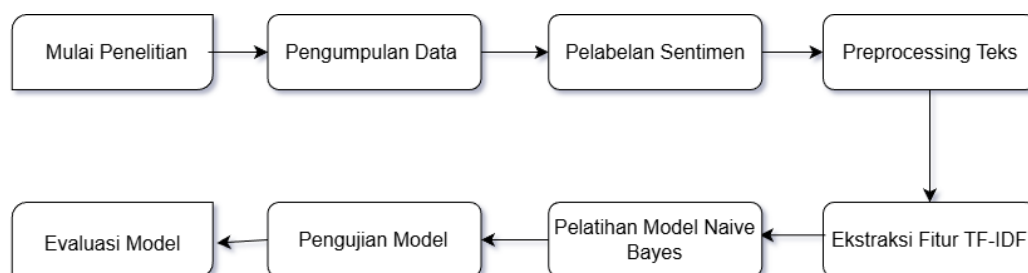
Relevansi penggunaan Naïve Bayes dalam konteks ulasan produk *skincare* juga didasarkan pada karakteristik data yang bersifat opini singkat. Dalam berbagai studi kasus sosial dan hukum, NBC terbukti mampu menangkap sentimen publik dengan tingkat akurasi yang memadai untuk digunakan sebagai dasar pengambilan keputusan [8]. Bahkan dalam dataset yang memiliki kemiripan antara kelas positif dan negatif, NBC mampu memberikan batasan probabilitas yang jelas [9]. Oleh karena itu, penerapan NBC pada ulasan produk *skincare* di Shopee diharapkan dapat memberikan gambaran akurat mengenai persepsi masyarakat terhadap produk kecantikan yang beredar di pasar lokal.

Penelitian ini tidak hanya berfokus pada hasil akhir akurasi, tetapi juga ingin membedah bagaimana algoritma Naïve Bayes merespons ulasan-ulasan yang mengandung unsur teknis kecantikan (seperti kandungan bahan kimia, efek samping, dan harga). Dengan mengintegrasikan teknik prapemrosesan data yang kuat dan algoritma NBC, penelitian ini bertujuan untuk membangun sebuah sistem klasifikasi sentimen yang dapat membantu konsumen dalam memilih produk berdasarkan ulasan objektif, sekaligus membantu pelaku usaha *skincare* di Shopee untuk melakukan evaluasi berkelanjutan terhadap produk yang mereka pasarkan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian ini disusun secara sistematis untuk memproses data ulasan *skincare* dari Female Daily yang tidak terstruktur menjadi hasil klasifikasi sentimen yang akurat. Alur kerja dalam penelitian ini secara visual dapat dilihat pada Gambar 1.



Gambar 1. Tahap Penelitian

Gambar 1 Diagram ini menunjukkan siklus pengolahan data yang dimulai dari pengumpulan dataset (00. Review.csv & 00. InfoProduct.csv), dilanjutkan ke tahap *preprocessing* (*Cleansing, Case Folding, Normalization, Filtering, Stemming*), ekstraksi fitur menggunakan TF-IDF, proses klasifikasi dengan Naïve Bayes, hingga tahap evaluasi menggunakan *Confusion Matrix*. Proses penelitian diawali dengan pengumpulan dataset dan studi literatur untuk memahami sentimen ulasan produk kecantikan di platform digital. Penggunaan analisis klasifikasi sangat penting untuk memetakan persepsi konsumen guna memberikan kontribusi nyata dalam pengambilan keputusan berbasis data yang lebih tepat sasaran [10]. Selanjutnya, dilakukan tahap *preprocessing* untuk menyusun ulang data agar menjadi lebih ringkas, efisien, dan sesuai dengan keperluan tanpa mengubah informasi intinya [11]. Tahapan ini membersihkan teks dari gangguan (*noise*) seperti format yang tidak standar agar siap diproses oleh algoritma klasifikasi. Setelah teks bersih, data diubah menjadi bentuk numerik melalui TF-IDF dan diklasifikasikan menggunakan algoritma Naïve Bayes. Tahap akhir penelitian adalah evaluasi menggunakan *Confusion Matrix* untuk mengukur performa akurasi, presisi, dan *recall* dari model yang dihasilkan.

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan gabungan dari tiga sumber dataset ulasan produk *skincare* yang berasal dari platform Female Daily. Dataset pertama dan kedua memberikan informasi detail mengenai ulasan individu dan spesifikasi produk secara mendalam, sedangkan dataset ketiga memberikan cakupan data yang lebih luas mengenai berbagai jenis produk *sunscreen* dan pembersih wajah di pasar Indonesia. Struktur dari ketiga dataset tersebut dijelaskan pada Tabel 1 sampai Tabel 3 berikut:

Tabel 1. Atribut pada Dataset Informasi Produk

Nama Atribut	Tipe Data	Deskripsi
Category	String	Kategori jenis produk (Night Cream, Toner, dll)
Merk	String	Nama brand produsen skincare
Product	String	Nama spesifik produk
Price	String	Harga produk dalam Rupiah
OverallRating	Numerik	Rata-rata rating produk secara keseluruhan
Reviewer	Numerik	Jumlah total pengguna yang memberi ulasan

Tabel 2. Atribut pada Dataset Ulasan Pengguna

Nama Atribut	Tipe Data	Deskripsi
Product	String	Nama produk yang diulas (kunci relasi)
UserName	String	Nama akun pengulas
SkinCond_Age	String	Kondisi kulit dan rentang usia pengulas
Recommend	String	Status rekomendasi produk
Review	Teks	Konten teks opini/sentimen mentah
Rating	Numerik	Skor kepuasan individu (skala 1-5)

Tabel 3. Atribut pada Dataset Pendukung

Nama Atribut	Tipe Data	Deskripsi
Type	String	Jenis kategori produk (Sunscreen, Cleanser, dll)
Name	String	Nama lengkap produk
Brand	String	Nama merek produk
Rating	Numerik	Penilaian produk
Total Reviewers	Numerik	Jumlah pengulas dalam format ribuan
Link	String	Tautan menuju halaman ulasan asli

Ketiga dataset ini diintegrasikan untuk memperkaya fitur informasi dalam proses analisis. Atribut Review dari dataset kedua akan diproses melalui tahap prapemrosesan teks, sementara atribut Rating dari seluruh dataset akan digunakan sebagai parameter utama dalam penentuan label sentimen positif dan negatif.

2.3 Pelabelan Sentimen

Pelabelan sentimen bertujuan mentransformasi data mentah menjadi kategori terstruktur agar dapat dipahami oleh mesin. Dalam penelitian ini, pelabelan otomatis dilakukan dengan mengonversi atribut Rating (skala 1-5) menjadi label biner untuk mempermudah algoritma Naïve Bayes dalam mempelajari karakteristik kata yang merepresentasikan kepuasan atau keluhan konsumen [12]. Ulasan dengan rating 4 dan 5 diklasifikasikan sebagai sentimen Positif karena menunjukkan pengalaman pengguna yang memuaskan, seperti efektivitas produk. Sebaliknya, rating 1 dan 2 dikategorikan sebagai sentimen Negatif karena sering memuat keluhan fisik spesifik seperti iritasi atau *breakout* [13]. Pemisahan tegas antara kedua kutub sentimen ini memungkinkan model untuk melakukan klasifikasi secara lebih akurat berdasarkan pembobotan kata dalam ulasan [14]. Aturan pelabelan ini dirinci pada Tabel 4.

Tabel 4. Aturan Pelabelan Sentimen Berdasarkan Skor Rating

Skor Rating	Label Sentimen	Definisi Operasional Skincare
4 - 5	Positif	Produk cocok, efektif, dan tidak memicu efek samping.
1 - 2	Negatif	Produk memicu masalah kulit atau tidak memberikan hasil.
3	Netral	Produk memberikan hasil biasa saja (Dieksklusi).

Berdasarkan Tabel 4 menunjukkan logika konversi nilai rating menjadi label kategori. Label Positif merepresentasikan keberhasilan penggunaan produk, sedangkan label Negatif mendeteksi kegagalan produk atau reaksi alergi. Rating 3 diklasifikasikan sebagai data netral dan dikeluarkan dari proses pelatihan model guna menghindari ambiguitas data yang dapat menurunkan tingkat akurasi dan presisi algoritma Naïve Bayes.

2.4 Pre-processing Data

Tahap *preprocessing* merupakan proses krusial dalam analisis sentimen untuk membersihkan data ulasan mentah dari gangguan (*noise*) agar siap diproses oleh algoritma Naïve Bayes. Mengingat ulasan pada portal Female Daily sering menggunakan bahasa informal, singkatan, dan simbol emosi, prapemrosesan yang konsisten sangat diperlukan untuk meningkatkan kualitas fitur teks. Langkah ini bertujuan untuk menyeragamkan struktur data sehingga mesin dapat mengenali makna sentimen secara lebih akurat [15].

Langkah-langkah prapemrosesan yang dilakukan dalam penelitian ini adalah sebagai berikut:

- Case Folding:** Tahap pertama adalah mengubah seluruh karakter huruf dalam ulasan menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk menghindari sistem menganggap kata yang sama sebagai dua fitur berbeda hanya karena perbedaan huruf kapital.
- Cleansing:** Proses pembersihan data dari karakter yang tidak diperlukan, seperti angka, tanda baca, simbol khusus, serta emoji. Tahap ini memastikan bahwa hanya karakter alfabet yang akan diproses lebih lanjut.
- Normalization (Normalisasi):** Mengubah kata-kata tidak baku, istilah gaul (*slang*), atau singkatan yang sering muncul dalam dataset (seperti "bgttt", "gacocok", "recom") menjadi kata baku sesuai Ejaan Yang Disempurnakan (EYD). Kualitas hasil dari tahapan ini sangat menentukan performa klasifikasi karena teks yang bersih akan mengurangi dimensi fitur yang tidak relevan dalam perhitungan probabilitas [16].
- Filtering (Stopword Removal):** Melakukan eliminasi terhadap kata-kata umum yang sering muncul namun tidak memiliki nilai informasi atau makna sentimen yang kuat, seperti kata sandang, kata hubung, dan kata ganti (contoh: "yang", "dan", "saya").
- Stemming:** Tahap akhir di mana setiap kata dikembalikan ke bentuk dasarnya dengan menghilangkan imbuhan (awalan, sisipan, atau akhiran). Dalam penelitian ini, digunakan algoritma Sastrawi untuk memproses teks bahasa Indonesia, sehingga kata-kata seperti "melembabkan" dan "kelembaban" akan disatukan menjadi kata dasar "lembab". Dengan standarisasi kata melalui seluruh rangkaian ini, algoritma dapat lebih mudah memetakan hubungan antara kata kunci tertentu dengan label sentimen yang telah ditentukan secara konsisten [17].

2.5 Ekstraksi Fitur TF-IDF

Tahap ekstraksi fitur dilakukan untuk mengubah teks ulasan yang sudah melalui proses preprocessing menjadi representasi numerik yang dapat diproses oleh algoritma Naïve Bayes. Pada penelitian ini digunakan metode TF-IDF (Term Frequency – Inverse Document Frequency) karena mampu memberikan bobot pada kata berdasarkan tingkat kepentingannya dalam sebuah dokumen. *Term Frequency (TF)* menghitung seberapa sering sebuah kata muncul dalam satu ulasan, sedangkan *Inverse Document Frequency (IDF)* memberikan bobot lebih besar pada kata yang jarang muncul diseluruh dataset. Kombinasi keduanya menghasilkan TF-IDF yang merepresentasikan tingkat relevansi setiap kata terhadap sentimen dalam ulasan. Hasil ekstraksi fitur berupa matrik numerik TF-IDF inilah yang digunakan sebagai input pada tahap pelatihan model Naïve Bayes Berikutnya

2.6 Naïve Bayes

Algoritma Naïve Bayes merupakan metode klasifikasi yang berdasar pada Teorema Bayes dengan asumsi independensi yang kuat antar fitur. Dalam analisis sentimen ulasan *skincare*, algoritma ini bekerja dengan menghitung probabilitas kemunculan kata-kata tertentu untuk menentukan apakah ulasan tersebut termasuk dalam kategori positif atau negatif. Naïve Bayes dipilih karena kemampuannya dalam menangani dataset teks yang besar dengan proses komputasi yang efisien dan tingkat akurasi yang tinggi. Secara matematis, Teorema Bayes dinyatakan dalam persamaan berikut:

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

Pada rumus (1) tersebut, menjelaskan bahwa $P(C|X)$ adalah probabilitas posterior, yaitu peluang sebuah ulasan masuk ke dalam kelas tertentu (positif atau negatif) setelah fitur kata-kata di dalamnya diketahui. $P(X|C)$ merupakan *likelihood*, yang merepresentasikan seberapa sering sebuah kata muncul dalam kategori sentimen tertentu pada data latih. Sementara itu, $P(C)$ adalah probabilitas *prior* atau peluang awal dari masing-masing kelas sentimen sebelum ulasan diperiksa, dan $P(X)$ adalah probabilitas dari fitur atau kata-kata yang muncul sebagai nilai penormal. Naïve

Bayes dipilih karena kemampuannya dalam menangani dataset teks yang besar dengan proses komputasi yang efisien namun tetap mampu memberikan tingkat akurasi yang tinggi [18].

2.7 Pengujian Model

Pengujian model merupakan tahap akhir untuk mengukur sejauh mana algoritma Naïve Bayes mampu mengklasifikasikan ulasan *skincare* dengan benar. Dalam penelitian ini, pengujian dilakukan menggunakan *Confusion Matrix*, sebuah alat evaluasi yang membandingkan hasil prediksi sistem dengan label aktual pada data uji. Metrik utama yang digunakan meliputi akurasi untuk melihat kedekatan hasil prediksi dengan nilai sebenarnya, presisi untuk mengukur tingkat ketepatan antara informasi yang diminta dengan jawaban sistem, serta *recall* untuk mengukur keberhasilan sistem dalam menemukan kembali sebuah informasi [19]. Proses pengujian ini dilakukan dengan membagi dataset menjadi data latih (*training*) dan data uji (*testing*). Evaluasi performa melalui *Confusion Matrix* sangat penting karena memungkinkan peneliti untuk mengidentifikasi kesalahan klasifikasi, baik berupa *False Positive* maupun *False Negative*. Dengan menggunakan parameter evaluasi yang komprehensif seperti *accuracy*, *precision*, dan *recall*, kualitas model yang dibangun dapat divalidasi secara objektif untuk memastikan bahwa hasil analisis sentimen memiliki reliabilitas yang tinggi dalam memetakan persepsi konsumen [17].

2.8 Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma Naïve Bayes dalam mengklasifikasikan data ulasan ke dalam kategori yang tepat. Instrumen utama yang digunakan adalah *Confusion Matrix*, yang menyajikan perbandingan antara hasil prediksi sistem dengan data aktual di lapangan [20]. Deskripsi dari setiap kategori dalam matriks evaluasi tersebut dapat dilihat pada Tabel 6.

Tabel 5. Deskripsi Komponen Confusion Matrix

Komponen	Nama	Deskripsi Operasional (Sentimen <i>Skincare</i>)
TP	True Positive	Kondisi di mana ulasan aktual adalah positif dan sistem berhasil memprediksinya sebagai kelas positif dengan benar.
TN	True Negative	Kondisi di mana ulasan aktual adalah negatif (keluhan) dan sistem berhasil memprediksinya sebagai kelas negatif dengan benar.
FP	False Positive	Kondisi di mana ulasan aktual adalah negatif namun sistem salah memprediksinya sebagai kelas positif (<i>Type I Error</i>).
FN	False Negative	Kondisi di mana ulasan aktual adalah negatif namun sistem salah memprediksinya sebagai kelas positif (<i>Type I Error</i>).

Setelah mendapatkan nilai dari *Confusion Matrix*, dilakukan perhitungan metrik evaluasi untuk mengukur kinerja model klasifikasi secara kuantitatif. Penjelasan mengenai metrik-metrik tersebut adalah sebagai berikut:

a. Accuracy (Akurasi):

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

Merupakan metrik paling mendasar yang digunakan untuk mengukur sejauh mana model mampu mengklasifikasikan seluruh data ulasan (baik positif maupun negatif) dengan benar. Akurasi dihitung dengan membagi total prediksi yang benar (TP + TN) dengan keseluruhan jumlah dataset yang diuji. Tinggi rendahnya nilai akurasi menunjukkan kedekatan hasil prediksi sistem dengan data aktual secara global [20].

b. Precision (Presisi):

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Menunjukkan tingkat ketepatan antara data yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Dalam analisis ini, presisi mengukur seberapa banyak ulasan yang diprediksi positif memang benar-benar merupakan ulasan positif yang sesungguhnya. Nilai presisi yang tinggi meminimalkan terjadinya kesalahan klasifikasi di mana ulasan negatif dianggap positif (*False Positive*) [21].

c. Recall (Sensitivitas):

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Mengukur keberhasilan sistem dalam menemukan kembali seluruh informasi relevan yang ada dalam dataset. *Recall* menghitung perbandingan ulasan positif yang berhasil diprediksi benar oleh sistem terhadap total keseluruhan data yang aslinya berlabel positif. Metrik ini sangat krusial untuk memastikan tidak banyak ulasan positif yang terlewat atau salah terdeteksi sebagai sentimen negatif (*False Negative*) [19].

d. F1-Score:

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{5}$$

Merupakan metrik perbandingan rata-rata harmonik antara presisi dan *recall*. *F1-Score* digunakan untuk memberikan gambaran performa model yang lebih seimbang, terutama jika terdapat ketidakseimbangan jumlah antara data ulasan positif dan negatif dalam dataset. Semakin mendekati nilai 1, maka semakin baik pula keseimbangan ketepatan dan jangkauan klasifikasi yang dihasilkan oleh algoritma Naïve Bayes [20].

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk mengklasifikasi sentimen terhadap ulasan produk *Skincare* pada platform Shopee menggunakan algoritma Naïve Bayes. Hasil setiap tahapan penelitian ini.

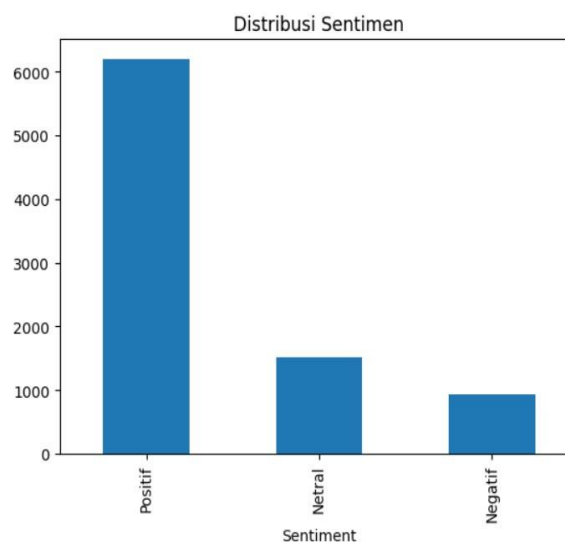
3.1 Pelabelan Sentimen

Setelah melakukan pengumpulan data, langkah berikutnya adalah proses pelabelan sentimen. Data yang sudah bersih, dari 16.941 data menjadi 8.646 data, selanjutnya diberi label sentimen, yaitu positif, negatif, dan netral. Pada penelitian ini, proses pelabelan awal dilakukan berdasarkan sistem rating yang diberikan pengguna pada setiap ulasan. Pendekatan berbasis rating ini digunakan karena sebagian besar ulasan pada platform e-commerce tidak menyediakan label sentiment secara eksplisit. Secara umum, ketentuan pelabelan sentiment berdasarkan rating, yaitu rating 4-5 dikategorikan sebagai positif, rating 3 dikategorikan sebagai netral dan rating 1-2 dikategorikan sebagai negatif.

Tabel 7. Pelabelan sentiment rating

Ulasan Produk	Rating	Rating
Di aku beneran kerja, melebabkan bgt. Tapi gak lebay gak bikin jerawat, malah jerawat batu gamau tumbuh lagi sampe sekarang udah 1 bulan pakai dan baru habis setengah krnb pakai dikit aja udah ngelock lembab diwajah. Aku pake pagi dan malam.	5	Positif
Enak sih dimukaku, yg tadinya kering jdi lembab gt. Pagi2 bangun kenyal gt mukanya, tp 3 hari pemakaian, muncul jerawat gede2 aku off 2 hari, ilang jerawatnya, kupake lagi muncul lagi. Awalnya mikir jerawatnya muncul mungkin hormone, gooling ada jg yang kyk gini. Sye.	3	Netral
Tekstur dari produk ini tuh kayak jelly yang diminum tapi meyerap dengan baik, sayangnya dia nggak cocok dikulit aku yang sensitive. Agak sedikit bikin kulit panas dan membuat kulitku jadi kemerahan.	2	Negatif

Setelah pelabelan otomatis berdasarkan rating dilakukan, penelitian ini juga melakukan pengecekan terhadap beberapa sampel ulasan untuk memastikan kesesuaian konteks teks dengan label sentiment yang diberikan. Pendekatan ini penting karena terdapat kemungkinan bahwa ulasan berisi makna campuran atau tidak sepenuhnya sesuai dengan rating. Dengan hasil distribusi sentimen awal, kategori positif terdapat 6.196 ulasan, kategori netral terdapat 1.520 ulasan dan kategori negatif terdapat 930 ulasan.



Gambar 2. Distribusi Sentimen

Berdasarkan Gambar 2, menunjukkan alur transformasi skor *rating* menjadi label sentimen terstruktur. Ulasan dengan rating 4 dan 5 dikonversi menjadi label Positif yang menandakan kepuasan pengguna, sedangkan rating 1 dan 2 menjadi label Negatif untuk mendeteksi keluhan. Rating 3 diklasifikasikan sebagai Netral dan dikeluarkan dari dataset agar tidak terjadi ambiguitas data. Proses ini menghasilkan label biner yang memudahkan algoritma Naïve Bayes dalam mengenali karakteristik kata yang membedakan kepuasan dan keluhan secara akurat.

3.3 Ekstraksi Fitur TF-IDF

Setelah melalui tahap preprocessing, teks ulasan perlu diubah ke dalam numerik agar dapat diproses oleh algoritma machine learning. Pada penelitian ini digunakan metode *term requecyinverse document frequency* (TF-IDF), yaitu teknik pembobotan kata yang umum digunakan dalam analisis teks dan klasifikasi dokumen. TF-IDF memberikan bobot yang lebih tinggi pada kata yang sering muncul dalam satu dokumen tetapi jarang muncul pada dokumen lain, sehingga dianggap memiliki nilai informasi yang lebih baik. Dalam implementasinya, pembobotan TF-IDF dilakukan menggunakan fungsi `TfidfVectorizer` dari *scikit-learn* dengan parameter `max_features = 10.000` untuk membatasi jumlah fitur dan `ngram_range = (1,2)` untuk menghasilkan fitur berupa unigram dan bigram, penggunaan bigram bertujuan menangkap konteks frasa seperti ”tidak cocok”, ”sangat bagus”, “ atau ”cepat meresap” yang sering berperan penting dalam analisis sentimen.

```
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer(max_features=10000, ngram_range=(1,2))
X = tfidf.fit_transform(reviews_df['clean_review'])
y = reviews_df['Sentiment']

print("TF-IDF Matrix:", X.shape)
```

TF-IDF Matrix: (8646, 10000)

Gambar 3. Ekstraksi Fitur TF-IDF

Berdasarkan Gambar 3, hasil ekstraksi fitur menunjukkan bahwa TF-IDF menghasilkan matrik (8.646 x 10.000), yang berarti terdapat 8.646 ulasan dan 10.000 fitur berupa kata atau pasangan kata yang dipilih berdasarkan tingkat kepentingannya dalam korpus.

3.5 Pelatihan Model Naïve Bayes

Proses pelatihan model dimulai setelah data melalui tahap representasi fitur menggunakan TFIDF dan pembagian data latih serta data uji. Berdasarkan hasil pemisahan data, diperoleh 6.916 data latih dan 1.730 data uji, masing-masing memiliki 10.000 fitur TF-IDF. Jumlah fitur yang besar menunjukkan bahwa representasi n-gram (unigram dan bigram) mampu menangkap banyak variasi kata yang relevan untuk membedakan sentimen dalam ulasan skincare. Pada tahap pelatihan, algoritma Multinomial Naïve Bayes digunakan karena algoritma ini dirancang untuk menangani data tekstual berbasis frekuensi dan terbukti efisien untuk data set berukuran besar seperti ulasan teks.

Train: (6916, 10000) Test: (1730, 10000)

Gambar 4. Pelatihan Model Naive Bayes

Berdasarkan Gambar 4, menunjukkan proses pelatihan model menggunakan algoritma Multinomial Naïve Bayes dengan 6.916 data latih dan 10.000 fitur TF-IDF. Penggunaan *n-gram* (unigram dan bigram) memungkinkan model menangkap variasi kata yang luas dalam ulasan *skincare*. Tahap ini bertujuan agar algoritma dapat mempelajari pola frekuensi kata dari data latih untuk kemudian digunakan dalam memprediksi sentimen pada data uji secara efisien.

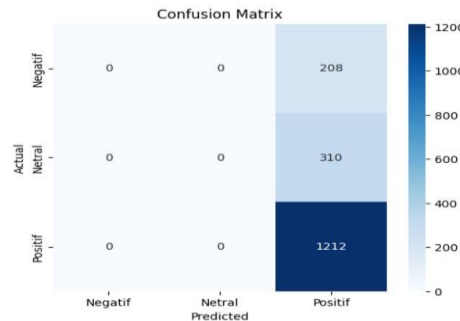
3.6 Pengujian Model

Pengujian model dilakukan untuk mengevaluasi kinerja algoritma Naïve Bayes Multinomial dalam mengklasifikasi sentimen ulasan skincare ke dalam tiga kelas, yaitu positif, negatif, dan netral. Dataset yang digunakan berjumlah 1.730 data, yang dibagi menjadi 80% data latih dan 20% data uji. Sebelum proses klasifikasi, data teks direpresentasikan ke dalam bentuk numerik menggunakan metode *countVectorizer*. Evaluasi performa model dilakukan menggunakan metrik *precision*, *recall*, *F1-score* dan *accuracy* untuk memberikan gambaran menyeluruh mengenai kemampuan model dalam mengenali setiap kelas sentimen, khususnya pada kondisi distribusi data yang tidak seimbang, hasil pengujian model Naïve Bayes pada data ujian.

Tabel 13. Hasil Pengujian Model

Kelas	Precision	Recall	F1-Score	Support
Negatif	1.00	0.00	0.01	208
Netral	0.00	0.00	0.00	310
Positif	0.70	1.00	0.82	1212
Accuracy	---	---	0.70	1730
Accuracy	0.57	0.33	0.28	1730
Weighted avg	0.61	0.70	0.58	1730

Berdasarkan Tabel 13, Model memiliki tingkat Akurasi 0,70 (70%), yang berarti sebagian besar data berhasil diprediksi dengan benar. Namun, performa model sangat tidak seimbang; model sangat kuat di Kelas Positif (Recall 1,00), tetapi gagal mengenali Kelas Negatif dan Netral (Recall mendekati 0,00). Hal ini menunjukkan adanya kecenderungan model untuk menebak semua data sebagai positif karena jumlah data positif yang jauh lebih mendominasi (*support* 1212 data) dibandingkan kelas lainnya.



Gambar 5. Hasil Pengujian Model

Berdasarkan Gambar 5, model Naïve Bayes menunjukkan performa yang tidak merata pada setiap kelas sentimen. Kelas positif memiliki performa terbaik dengan nilai *recall* sebesar 1.00 yang menandakan seluruh data positif berhasil diklasifikasi. Meskipun nilai *precision* sebesar 0.70 menunjukkan masih adanya kesalahan klasifikasi dari kelas lain. Sebaliknya kelas negative dan netral memiliki performa yang rendah. Kelas negative memiliki *precision* tinggi tetapi *recall* 0.00, sedangkan kelas netral tidak berhasil diprediksi sama sekali. Hal ini menunjukkan adanya bias model terhadap kelas mayoritas akibat ketidakseimbangan distribusi data. Oleh karena itu diperlukan penanganan ketidakseimbangan kelas agar performa klasifikasi menjadi lebih seimbang.

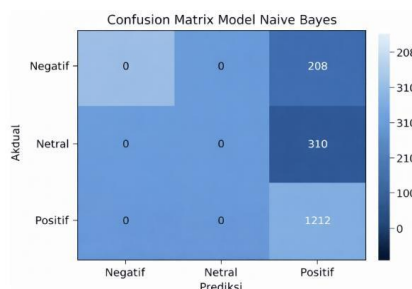
3.7 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja algoritma *Naïve Bayes Multinomial* dalam mengklasifikasi sentimen ulasan skincare ke dalam tiga kelas, yaitu positif, negatif, atau netral. Evaluasi dilakukan menggunakan metrik *precision*, *recall*, *F1-Score*, dan akurasi, serta divisualisasikan menggunakan *confusion matrix*. Berdasarkan hasil evaluasi pada tabel 8 model memperoleh akurasi sebesar 70.06% hasil ini menunjukkan bahwa sebagian besar data uji berhasil di klasifikasi dengan benar. Namun *classification report* memperlihatkan bahwa performa model tidak merata pada setiap kelas sentimen. Kelas positif memiliki performa terbaik dengan nilai *recall* sebesar 1.00, yang menandakan seluruh data positif berhasil dikenal oleh model. Ringkasan hasil evaluasi model ditampilkan pada diagram visualisasi matrik konfusi disajikan pada gambar dan tabel berikut.

Tabel 14. Hasil Akurasi Model Naive Bayes

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.00	0.00	0.00	208
Netral	0.00	0.00	0.00	310
Positif	0.70	1.00	0.82	1212
Macro Avg	0.23	0.33	0.27	1730
Weighted avg	0.49	0.70	0.58	1730

Berdasarkan Tabel 14, Accuracy: 0.70 (70,06%) dan Total support : 1730 akan tetapi nilai *precision* sebesar 0.70 menunjukkan bahwa masih terdapat data dari kelas lain yang salah diklasifikasi sebagai positif. Hal ini mengindikasikan bahwa model cenderung bias terhadap kelas positif. Sebaliknya kelas negatif dan netral menunjukkan performa yang rendah. Kelas negative memiliki nilai *precision* tinggi namun *recall* sebesar 0.00 sedangkan kelas netral tidak berhasil diprediksi sama sekali. Kondisi ini diperjelas oleh gambar 4 yang menunjukkan bahwa seluruh data negatif dan netral diprediksi sebagai kelas positif pada matrik konfusi.



Gambar 6. Hasil Evaluasi Confusion Matrix

Berdasarkan Gambar 6, nilai macro average F1-score sebesar 0.28 menunjukkan bahwa performa model pada kelas minoritas masih sangat rendah, meskipun nilai weighted average F1-score sebesar 0.58 dan akurasi sebesar 70.06% terlihat cukup tinggi. Hal ini menandakan bahwa akurasi belum sepenuhnya mencerminkan kualitas model pada dataset yang tidak seimbang. Secara keseluruhan, model Naïve Bayes efektif dalam mengklasifikasi sentimen positif, namun belum optimal dalam mengenali sentiment negatif dan netral, sehingga diperlukan strategi lanjut seperti pegganaan ketidakseimbangan kelas, peningkatan representasi fitur, dan penggunaan metrik evaluasi yang lebih sesuai agar performa model menjadi lebih seimbang.

4. KESIMPULAN

Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam klasifikasi sentimen ulasan produk skincare pada platform Shopee ke dalam tiga kategori sentiment, yaitu positif, negatif, dan netral. Berdasarkan hasil pengujian yang telah dilakukan, model Naïve Bayes mampu mencapai tingkat akurasi sebesar 70,06%, yang menunjukkan bahwa algoritma ini cukup efektif sebagai pendekatan awal dalam analisis sentimen teks ulasan skincare. Hasil evaluasi model menunjukkan bahwa performa terbaik dicapai pada kelas sentimen positif dengan nilai F1-score sebesar 0,82, sementara performa pada kelas sentimen negatif dan netral masih relative lebih rendah. Hal ini mengindikasikan adanya ketidakseimbangan kemampuan model dalam klasifikasi setiap kelas sentimen, yang dipengaruhi oleh distribusi data yang tidak seimbang serta keterbatasan representasi fitur teks yang digunakan. Meskipun demikian, hasil penelitian ini membuktikan bahwa algoritma Naïve Bayes memiliki potensi dalam mengidentifikasi pola sentimen positif pada ulasan produk skincare secara cukup akurat. Namun, model yang dibangun belum sepenuhnya optimal dalam menangani data dengan distribusi kelas yang tidak merata. Oleh karena itu, penelitian ini menyimpulkan bahwa penerapan Naïve Bayes dapat dijadikan sebagai dasar awal dalam analisis sentimen ulasan skincare, tetapi masih diperlukan pengembangan lebih lanjut. Pengembangan tersebut dapat berupa penangan ketidakseimbangan kelas, peningkatan teknik praproses data serta penggunaan metode representative fitur yang lebih kaya agar hasil klasifikasi menjadi lebih seimbang, akurat, dan representative sesuai dengan tujuan penelitian yang telah ditetapkan.

REFERENCES

- [1] A. Rahman *et al.*, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 11, no. 1, 2025, doi: 10.26418/jp.v11i1.72891.
- [2] A. Annur Rohman, G. Alfa Trisnapradika, and K. Kunci, "Perbandingan Algoritma NBC, SVM, Logistic Regression untuk Analisis Sentimen Terhadap Wacana KaburAjaDulu di Media Sosial X," *Technology and Science (BITS)*, vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7261.
- [3] A. Ramadhani, I. Permana, M. Afdal, and M. Fronita, "Analisis Sentimen Tanggapan Publik di Twitter Terkait Program Kerja Makan Siang Gratis Prabowo–Gibran Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, Dec. 2024, doi: 10.47065/bits.v6i3.6188.
- [4] E. Arya Pranata, F. Budiman, and D. Kurniawan, "Analisis Sentimen Ulasan Mobile JKN pada Playstore dengan Perbandingan Akurasi Algoritma Naïve Bayes dan SVM," *Technology and Science (BITS)*, vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7334.
- [5] R. F. P. Pratama and W. Maharani, "Comparative Analysis of Naive Bayes and SVM for Improved Emotion Classification on Social Media," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 11–20, Apr. 2025, doi: 10.29408/edumatic.v9i1.29087.
- [6] Y. Christian, T. Wibowo, and M. Lyawati, "Sentiment Analysis by Using Naïve Bayes Classification and Support Vector Machine, Study Case Sea Bank," *Sinkron*, vol. 9, no. 1, pp. 258–275, Jan. 2024, doi: 10.33395/sinkron.v9i1.13141.
- [7] A. V. Kuncoro, F. Budiman, and D. Kurniawan, "Analisis Sentimen Pengguna X terhadap Kasus Korupsi Gula Tom Lembong Menggunakan Naïve Bayes, SVM, dan Random Forest," *Technology and Science (BITS)*, vol. 7, no. 3, 2025, doi: 10.47065/bits.v7i3.8577.
- [8] N. V. R. Jhosefhin, S. Srianu, and I. Kurniawan, "Analisis Sentimen Crawling Data dari Sosial Media X tentang Gaza Menggunakan Metode SVM dan Decision Tree" *Jurnal Indonesia Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 6, no 1, 2025, doi:10.35870/jimik.v6i1.1225
- [9] M. F. A. Shidiq, D. Alita, and L. Ratu, "Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data dari Media Sosial X Pendekatan Naive Bayes dan SVM," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025, doi: 10.47080/simika.v8i1.3533.
- [10] M. N. Romadhoni and N. A. S. Winarsih, "Kinerja Naive Bayes dan SVM pada Data Survei Tidak Seimbang: Studi Klasifikasi Kepuasan Masyarakat," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 2, pp. 382–391, Aug. 2025, doi: 10.29408/edumatic.v9i2.30185.
- [11] Z. A. Mukharyahya, Y. P. Astuti, and O. N. Cahyani, "Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 119–128, Apr. 2025, doi: 10.29408/edumatic.v9i1.29512.
- [12] Y. Aprianti, A. Lia Hananto, S. Shofiah Hilabi, S. Informasi, and U. Buana Perjuangan Karawang, "Klasifikasi Sentimen Komentar Pengguna pada Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes," vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1023.



- [13] B. R. A. Putri, Fitrianti, "Penerapan Dan Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Dalam Analisis Sentimen Terhadap Kepuasan Pengguna Aplikasi Flo," *Jurnal Ilmiah Informatika Global*, vol. 16, no. 2, 2025, doi: 10.36982/jig.v16i2.4285.
- [14] B. Z. Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, 2022, doi: 10.30871/jaic.v6i1.3654.
- [15] A. Syafi'i, M. Afdal, E. Saputra, and R. Novita, "Analisis Sentimen Ulasan Pengguna Aplikasi Penjualan Pulsa Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 3, pp. 1300–1308, Jul. 2024, doi: 10.32493/jtsi.v7i3.41364.
- [16] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, "Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor," *METIK JURNAL*, vol. 6, no. 2, pp. 97–103, Dec. 2022, doi: 10.47002/metik.v6i2.372.
- [17] I. Kurniawan *et al.*, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *Jurnal Teknik Informatika dan Sistem Informasi (JUTISI)*, vol. 9, no. 1, 2023, doi: 10.35957/jutisi.v9i1.4121.
- [18] K. Nurfebria and S. Sriani, "Sentiment Analysis of Skincare Products Using the Naive Bayes Method," *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1663–1676, Sep. 2024, doi: 10.51519/journalisi.v6i3.817.
- [19] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*, Depok, Indonesia: UG Penerbit Gunadarma, 2023. ISBN: 978-602-0764-57-3.
- [20] D. Atikah, A. Hananto, Tukino, and E. Novalia, "Analisis Opini Pengguna Aplikasi Shopee Dengan Naïve Bayes Classifier," *JIKA (Jurnal Informatika) Universitas Muhammadiyah Tangerang*, vol. 9, no. 3, pp. 363-366, 2025, doi: 10.31000/jika.v9i3.13456.
- [21] Pramesti Melinea Berlianti and Erwin Yudi Hidayat, "Implementasi Naïve Bayes Classifier untuk Sentimen Produk Kecantikan Berdasarkan Ulasan Female Daily," *The Indonesian Journal of Computer Science*, vol. 13, no. 6, Dec. 2024, doi: 10.33022/ijcs.v13i6.4499.